ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)

UNIVERSITY OF ZAGREB
FACULTY OF ORGANIZATION AND INFORMATICS
VARAŽDIN

Central European CONFERENCE on Information and Intelligent Systems

Proceedings

34th International Scientific Conference

CECIIS

Organized by:
Faculty of Organization and Informatics
University of Zagreb

September 20-22/2023
Dubrovnik, Croatia

ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)

UDC 004:007

34th International Scientific Conference

Central European Conference on Information and Intelligent Systems

CECIIS

Proceedings

September 20th - 22nd, 2023
Dubrovnik, Croatia

Organized by:
Faculty of Organization and Informatics
University of Zagreb

Publisher
Faculty of Organization and Informatics, University of Zagreb
Pavlinska 2, HR-42000 Varaždin, Croatia

Contact
E-mail: ceciis@foi.unizg.hr
Phone: +385 42 390 890
Fax: +385 42 213 413
Web: ceciis.foi.hr
Fb: facebook.com/ceciis.foi

For Publisher
Nina Begičević Ređep, Faculty of Organization and Informatics, University of Zagreb, Croatia

Editors
Neven Vrček, University of Zagreb Faculty of Organization and Informatics, Croatia
Luis de Marcos Ortega, University of Alcalá, Spain
Petra Grd, University of Zagreb Faculty of Organization and Informatics, Croatia

Editorial Board
Victoria Appatova, University of Cincinnati, USA
Igor Balaban, University of Zagreb, Croatia
Nina Begičević Ređep, University of Zagreb, Croatia
Marko Bohanec, Jožef Stefan Institute, Slovenia
Leo Budin, University of Zagreb, Croatia
Tonči Carić, University of Zagreb, Croatia
Carlos Carrascosa, Universitat Politècnica de València, Spain
Anna Czarczyńska, Kozminski University, Poland
Luis de Marcos Ortega, University of Alcalá, Spain
Blaženka Divjak, University of Zagreb, Croatia
Jasminka Dobša, University of Zagreb, Croatia
Juraj Ďuďák, Slovak University of Technology in Bratislava, Slovakia
Matjaž Gams, Jožef Stefan Institute, Slovenia
Gabriel Gaspar, Slovak University of Technology in Bratislava, Slovakia
Gordan Gledec, University of Zagreb, Croatia
Andrina Granić, University of Split, Croatia
Marjan Heričko, University of Maribor, Slovenia
Dieter Hertweck, Herman Hollerith Research Centre, Germany
Edouard Ivanjko, University of Zagreb, Croatia
Dimitris Karagiannis, University of Vienna, Austria
Valentina Kirinić, University of Zagreb, Croatia
Marina Klačmer Čalopa, University of Zagreb, Croatia
Michal Kohani, University of Zilina, Slovakia
Mario Konecki, University of Zagreb, Croatia
Melita Draganić, University of Zagreb, Croatia
Marjan Krašna, University of Maribor, Slovenia
Alen Lovrenčić, University of Zagreb, Croatia
Sandra Lovrenčić, University of Zagreb, Croatia
Ivan Luković, University of Belgrade, Serbia
Ivan Magdalenić, University of Zagreb, Croatia
Aleksandar Marković, University of Belgrade, Serbia
Marta Mas Machuca, UIC Barcelona, Spain
Dunja Mladenić, Jožef Stefan Institute, Slovenia
Oliver Moravčik, Slovak University of Technology, Slovakia
Enrique Mu, Carlow University, USA
Karlo Nyarko, University of Osijek, Croatia
Jan Paralič, Technical University of Košice, Slovakia
Elisabeth Pergler, University of Applied sciences, Austria
Wolf Rauch, University of Graz, Austria
Sonja Ristić, University of Novi Sad, Serbia
Markus Schatten, University of Zagreb, Croatia
Zlatko Stapić, University of Zagreb, Croatia
William Steingartner, Technical University of Košice, Slovakia
Nebojša Stojčić, University of Dubrovnik, Croatia
Vjeran Strahonja, University of Zagreb, Croatia
Diana Šimić, University of Zagreb, Croatia
Katarina Tomičić-Pupek, University of Zagreb, Croatia
Jan Tretmans, Radboud University Nijmegen, The Netherlands
Wim Van Petegem, KU Lueven, Belgium
Violeta Vidaček Hainš, University of Zagreb, Croatia
Sirje Virkus, Tallinn University, Estonia
Mladen Vouk, North Carolina State University, USA
Ksenija Vuković, University of Zagreb, Croatia

Publishing Board
Igor Tomičić, Faculty of Organization and Informatics, University of Zagreb, Croatia
Bernarda Kos, Faculty of Organization and Informatics, University of Zagreb, Croatia

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical photocopying, recording or otherwise without the prior permission of the copyright owner.
CECIIS Proceedings have been indexed in the following databases: ProQuest (SciTech Premium Collection, ProQuest Technology Collection and ProQuest Advanced Technologies & Aerospace database) and Web of Science.

ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)

Creative Commons Licence
Central European Conference on Information and Intelligent Systems Proceedings is licensed under a Creative Commons Attribution-NonCommercialNoDerivatives 4.0 International License.
http://creativecommons.org/licenses/by-nc-nd/4.0/

Contents

Index of Authors	VII
List of Reviewers	IX
Preface	XI

Computer Games

Markus Schatten, Marinela Schatten, Ljiljana Inkret-Martinčević **Development of Gamified Cognitive Agents for Primary Schools - A Case Study**	3
Bogdan Okreša Đurić, Tomislav Peharda, Pasqual Martí **Towards a Gamified System to Influence Behaviour of Users in the Context of Smart Mobility**	9
Tomislav Peharda, Bogdan Okreša Đurić, Igor Tomičić **Towards Application of Programming Language for Communication Flows Specification in Multi-agent Systems on Real-World Use Cases**	17

Data and Knowledge Bases

Božidar Kovačić, Vanja Slavuj, Miodrag Sretenović **Data Visualisation System for E-learning: Overview of System Design**	25

Digital Human

Tihomir Katulić, Anamarija Musa, Darja Lončar **Understanding some of the open data challenges to data protection in the developing European legal framework**	35
Marko Bohanec **Inter- and Intra-Personal Differences, and Consistency of Decision Rules, in Multi-Criteria Modelling Method DEX: A Preliminary Study**	43
Kaja Prislan Mihelič, Igor Bernik, Simon Vrhovec, Anže Mihelič **Perceived Threat of Cyber Attacks and its Role in the Adoption of Tablet Computers by Older Adults**	49
Lucija Gosak, Jeroen Ooge, Nino Fijačko, Janja Kamenšek, Primož Kocbek, Nika Debeljak, Katrien Verbert , Gregor Štiglic **Self-Care Oriented Smartphone Apps for Type 2 Diabetes: A Comparative Analysis**	57
Katarina Tomičić-Pupek, Martina Tomičić Furjan, Igor Pihir **Digital Human Resource Management - Designing a digital platform for educational institutions' HR process**	67
Marija Đukić, Milica Škembarević, Olga Jejić, Ana Pajić Simović **Toward the utilization of chatbots in the banking sector**	79
Darko Babić, Mario Fiolić, Dario Babić **Inspection of the road surface condition on black spots – Croatia case study**	89

Digital Transformation and Corporate Culture

Iva Gregurec, Miroslav Mandić, Maja Zeljak
Examining the Ability of Users in the Croatian Market to Distinguish Organic and Sponsored Search Results 101

Rok Grgec, Leo Mršić, Mladen Konecki
Automated Analysis of Short Digital Messages with the Potential of Application in Business 111

Tea Mijač
Student-Centered Digital Services: Assessing User Experience in Higher Education 129

Snježana Križanić, Neven Vrček
Business process modeling and digital twins: a literature review 145

Nikolina Posarić, Damjan Ćorić
Analysis of Organizational Culture Using Cameron and Quinn's Typology: A Case Study of an ICT Organization 153

Nikolina Žajdela Hrustek, Renata Mekovec, Valentina Kirinić
E-Schools project - Framework for Digital Competencies of School Principals: Competencies in the Area Digital infrastructure 161

Education and Learning Analytics

Till Winkler, Rony G. Flatscher
Cognitive Load in Programming Education: Easing the Burden on Beginners with REXX 171

Bojan Radišić, Ivan Dunđer, Sanja Seljan
Data Analysis of the Motivation and Factors for a Shorter Duration of Study 179

Melita Milić , Blaženka Divjak
Innovative Learning Design and Teachers' Competencies: Prerequisites for Meaningful Use of Digital Material in Schools 189

Marko Matus, Goran Hajdin, Igor Balaban
Joint Creative Classroom as a Response to Industry 4.0 Demands 199

Ivana Ružić, Igor Balaban
The Use of Social Robots as Teaching Assistants in Education: Literature Review 205

Blaženka Divjak, Josipa Bađari, Darko Grabar, Damir Horvat, Bart Rienties, Petra Vondra
Enhancing Learning Design through User Experience Research: Insights from a Survey in Four European Countries 213

Josipa Lipovac, Dorotea Potoč, Tomislav Jagušt
How Programming Helps Early-Age Learners Tackle Complex Concepts 223

Ivan Kekez, Diana Šimić
Learning Analytics Frameworks: A Review of Challenges and Practices in Higher Education 231

ICT Entrepreneurship and Innovation

Daria Jergan, Ivana Đunđek Kokotec, Marina Klačmer Čalopa
Inflationary Expectations of Consumers in the Republic of Croatia 241

Intelligent Energy

Stefanie Hatzl, Katharina Moitzi
Identification of technology adoption paths for smart technologies 251

Intelligent Systems

Dennis Rall, Thomas Fraunholz, Bernhard Bauer
AI-Democratization: From Data-first to Human-first AI 261

Martina Ptiček, Jasminka Dobša
Annotation and Machine Identification of Metaphors in Croatian Newspaper Articles 269

Victor Mitrana, Andrei Paun, Mihaela Paun, Jose Ramon Sanchez Couso
Some Remarks on the Formal Operations Inspired by the Gene Assembly in Ciliates 287

Juraj Ďuďák, Gabriel Gašpar, Jakub Perička, Eva Vangelová, Daniel Gurín
Design of system for collecting and analyzing of human gait parameters 295

Filip Babić, Petra Grd, Igor Tomičić, Ena Barčić
Deep Transfer Learning for Detecting Face Images with Moustache 301

Vlatka Davidović, Sanda Martinčić Ipšić
Towards Sequence-to-Sequence Neural Model for Croatian Abstractive Summarization 309

Marko Horvat, Ljudevit Jelečević, Gordan Gledec
Comparative Analysis of YOLOv5 and YOLOv6 Models Performance for Object Classification on Open Infrastructure: Insights and Recommendations 317

Gaurish Thakkar, Marko Tadić, German Rigau
ComFac4LT – Computing Facilities for Language Technologies 325

Gabriel Gašpar, Juraj Dudak, Ivan Sladek, Ingrid Zuziakova, Tomas Meravy
Utilization of Open-Source in Upgrading Technological Devices for Industry 4.0 readiness 341

Ivana Kovačević, Mihaela Orić, Ivana Hartmann Tolić, Emmanuel Karlo Nyarko
Modelling the Fermentation Process in Winemaking using Temperature and Specific Gravity 347

Amila Akagić, Emir Buza, Marko Horvat
Mapping RGB-to-NIR with Pix2Pix Image-to-Image Translation for Fire Detection Applications 353

Ena Barčić, Petra Grd
Optimizing Point Cloud Processing for Facial Reconstruction Using Open-Source Tools: a Preliminary Study 361

Marko Horvat, Klara Krmpotić, Andrija Krtalić, Amila Akagić
Bridging Blockchain Technology and Humanitarian Demining: A Novel Concept for Decentralized Storage of Landmine and UXO Locations 369

Ivan Magdalenić, Elvis Popović, Neven Vrček, Boris Tomaš, Lovro Posarić
Localization of intruder UAV operator using Unmanned Aerial Vehicles 377

Dijana Oreški, Dunja Višnjić, Nikola Kadoić
Comparison of Machine Learning Algorithms on Noisy Data 383

Miran Pobar
YOLOv7 Model for Small Object Handling in Maritime Images 391

Karel Križnar, Jože Rožanec, Blaž Fortuna, Dunja Mladenić
Explainable Artificial Intelligence Meets Active Learning: A Novel GradCAM-based Active Learning Strategy 399

Vedran Orešković, Ana Meštrović, Slobodan Beliga
Towards Computational Content Analysis of Crises-Related News in Electronic Media 407

Quality in ICT

Miro Zdilar
Towards Automated Detection of Qualitative Spreadsheet Errors in Multi-user Environments 419

Strategic Planning and Decision Making

Hannia Gonzalez-Urango, Enrique Mu, Anna Florek-Paszkowska, Milagros Pereyra-Rojas
Validation and use of a framework to assess challenges to virtual education in the context of emergency remote teaching: Peru and Spain 427

Sam Sabah, Imran Hossain, David Weiss, Alexander Tillmann
Towards Increasing Active Citizen Involvement in Urban Planning through Mixed Reality Technologies 435

Mehtap Dursun, Rana Duygu Alkurt
Evaluating the Net Zero Policy Performances of European Countries Through DEA 441

Nikola Zornić, Aleksandar Marković, Nebojša Avramović
A Bibliometric Review of Agent-Based Models Application in Economics 447

Systems for Smart Mobility

Markus Schatten, Neven Vrček
A Method for Intruder UAV Pilot Localization based on Neural Networks using Intercepting Drone Constellations 457

Trends in Software Engineering

Marija Kuštelega, Renata Mekovec
A Systematic Analysis of Requirements Elicitation Problems and Challenges 465

Donata Szombathelyi, Tomislav Jakopec, Josipa Selthofer
The impact of using ChatGPT on employees of Croatian IT Companies 473

Ahmed Shareef, Boris Tomaš, Neven Vrček
Semantic Interoperability of Digital Twins in Smart Cities 481

Nadica Hrgarek Lechner
Using Card Sorting to Classify Elements of the Security Risk Management Program for Medical Devices 489

Dijana Peras, Zlatko Stapić, Mislav Matijević
AI techniques and tools in Agile Software Development: Preliminary research 501

Aránzazu San Juan Llano, Thi Mai Trang Nguyen, Daniel Rodríguez, Matthieu Chardy
POEMA: a Personal Cloud for Inclusive Education 509

Luis de Marcos, Daniel Rodríguez García, José-María Gutiérrez-Martínez, Adrián Domínguez-Díaz, Sergio Caro Álvaro
Recommendation and Prediction in a Microservice Web Application for Cyber Ranges 517

William Steingartner, Valerie Novitzká
Semantic methods in our research and teaching 525

Publication Ethics And Malpractice Statement	535
Author Guidelines	539
CECIIS 2023 Sponsors	553
CECIIS 2023 Program and Organizing Committees' Members, Research, Professional Tracks' Chairs and Student Poster and Presentation Section Chair	559

Index of Authors

Akagić, Amila 353, 369
Alkurt, Rana Duygu 441
Avramović, Nebojša 447
Babić, Dario 89
Babić, Darko 89
Babić, Filip 301
Bađari, Josipa 213
Balaban, Igor 199, 205
Barčić, Ena 301, 361
Bauer, Bernhard 261
Beliga, Slobodan 407
Bernik, Igor 49
Bohanec, Marko 43
Buza, Emir 353
Caro Álvaro, Sergio 517
Chardy, Matthieu 509
Ćorić, Damjan 153
Davidović, Vlatka 309
de Marcos, Luis 517
Debeljak, Nika 57
Divjak, Blaženka 189, 213
Dobša, Jasminka 269
Domínguez-Díaz, Adrián 517
Ďuďák, Juraj 295, 341
Đunđek Kokotec, Ivana 241
Dunđer, Ivan 179
Dursun, Mehtap 441
Đukić, Marija 79
Fijačko, Nino 57
Fiolić, Mario 89
Flatscher, Rony G. 171
Florek-Paszkowska, Anna 427
Fortuna, Blaž 399
Fraunholz, Thomas 261
Gašpar, Gabriel 295, 341
Gledec, Gordan 317
Gonzalez-Urango, Hannia 427
Gosak, Lucija 57
Grabar, Darko 213
Grd, Petra 301, 361
Gregurec, Iva 101
Grgec, Rok 111
Gurín, Daniel 295
Gutiérrez-Martínez, José-María 517
Hajdin, Goran 199
Hartmann Tolić, Ivana 347
Hatzl, Stefanie 251
Horvat, Damir 213
Horvat, Marko 317, 353
Hossain, Imran 435
Hrgarek Lechner, Nadica 489
Inkret-Martinčević, Ljiljana 3
Jagušt, Tomislav 223
Jakopec, Tomislav 473
Jejić, Olga 79
Jelečević, Ljudevit 317
Jergan, Daria 241
Kadoić, Nikola 383
Kamenšek, Janja 57
Katulić, Tihomir 35
Kekez, Ivan 231
Kirinić, Valentina 161
Klačmer Čalopa, Marina 241
Kocbek, Primož 57
Konecki, Mladen 111
Kovačević, Ivana 347
Kovačić, Božidar 25
Križanić, Snježana 145
Križnar, Karel 399
Krmpotić, Klara 369
Krtalić, Andrija 369
Kuštelega, Marija 465
Lipovac, Josipa 223
Lončar, Darja 35
Magdalenić, Ivan 377
Mandić, Miroslav 101
Marković, Aleksandar 447
Martí, Pasqual 9
Martinčić Ipšić, Sanda 309
Matijević, Mislav 501
Matus, Marko 199
Mekovec, Renata 161, 465
Meravy, Tomas 341
Meštrović, Ana 407
Mihelič, Anže 49
Mijač, Tea 129
Milić , Melita 189
Mitrana, Victor 287
Mladenić, Dunja 399
Moitzi, Katharina 251
Mršić, Leo 111
Mu, Enrique 427
Musa, Anamarija 35
Nguyen, Thi Mai Trang 509
Novitzká, Valerie 525
Nyarko, Emmanuel Karlo 347
Okreša Đurić, Bogdan 9, 17
Ooge, Jeroen 57
Oreški, Dijana 383
Orešković, Vedran 407
Orić, Mihaela 347
Pajić Simović, Ana 79
Paun, Andrei 287
Paun, Mihaela 287
Peharda, Tomislav 9, 17
Peras, Dijana 501
Pereyra-Rojas, Milagros 427
Perička, Jakub 295
Pihir, Igor 67
Pobar, Miran 391
Popović, Elvis 377
Posarić, Lovro 377
Posarić, Nikolina 153
Potoč, Dorotea 223
Prislan Mihelič, Kaja 49
Ptiček, Martina 269
Radišić, Bojan 179
Rall, Dennis 261
Rienties, Bart 213
Rigau, German 325
Rodríguez, Daniel 509, 517
Rožanec, Jože 399
Ružić, Ivana 205
Sabah, Sam 435
San Juan Llano, Aránzazu 509
Sanchez Couso, Jose Ramon 287
Schatten, Marinela 3
Schatten, Markus 3, 457
Selthofer, Josipa 473
Seljan, Sanja 179
Shareef, Ahmed 481
Sladek, Ivan 341
Slavuj, Vanja 25
Sretenović, Miodrag 25
Stapić, Zlatko 501
Steingartner, William 525
Szombathelyi, Donata 473
Šimić, Diana 231
Škembarević, Milica 79
Štiglic, Gregor 57
Tadić, Marko 325
Thakkar, Gaurish 325
Tillmann, Alexander 435
Tomaš, Boris 377, 481
Tomičić Furjan, Martina 67
Tomičić, Igor 17, 301
Tomičić-Pupek, Katarina 67
Vangelová, Eva 295
Verbert, Katrien 57
Višnjic, Dunja 383
Vondra, Petra 213
Vrček, Neven 145, 377
Vrhovec, Simon 49
Weiss, David 435
Winkler, Till 171
Zdilar, Miro 419
Zeljak, Maja 101
Zornić, Nikola 447
Zuziakova, Ingrid 341
Žajdela Hrustek, Nikolina 161

List of Reviewers

Babič, František, Slovakia
Babić, Snježana, Croatia
Balaban, Igor, Croatia
Barać, Dušan, Serbia
Beliga, Slobodan, Croatia
Bernik, Igor, Slovenia
Brdnik, Saša, Slovenia
Cagáňová, Dagmar, Slovakia
Carić, Tonči, Croatia
Carrascosa, Carlos, Spain
Chodarev, Sergej, Slovakia
Ciucan, Liviu, Romania
Cvetkoska, Violeta, North Macedonia
Delibašić, Boris, Serbia
de-Marcos, Luis, Spain
Divjak, Blaženka, Croatia
Dobrović, Željko, Croatia
Dudak, Juraj, Slovakia
Dudáš, Adam, Slovakia
Đunđek Kokotec, Ivana , Croatia
Dunđer, Ivan, Croatia
Đorđević Milutinović, Lena S., Serbia
Fabac, Robert, Croatia
Ferencik, Norbert, Slovakia
Flatscher, Rony G., Austria
Galinec, Darko, Croatia
Gašpar, Gabriel, Slovakia
Gnip, Peter, Slovakia
Grd, Petra, Croatia
Gregurić, Martin, Croatia
Guardia Ortiz, Lourdes, Spain
Hajdin, Goran, Croatia
Horvat, Marko, Croatia
Huber, Marco , Germany
Huraj, Ladislav, Slovakia
Ivanjko, Edouard, Croatia
Ivković, Nikola, Croatia
Jagušt, Tomislav, Croatia
Jednak, Sandra, Serbia
Jenčik, Marian, Slovakia
Kadoić, Nikola, Croatia
Knežević, Goran, Croatia
Kovšca, Vladimir, Croatia
Krašna, Marjan, Slovenia
Kubanek, Mariusz, Poland
Kučina Softić, Sandra, Croatia
Kvet, Marek, Slovakia
Kvet, Michal, Slovakia
Lang, Ján, Slovakia
Lovrenčić, Alen, Croatia
Lovrenčić, Sandra, Croatia
Luković, Ivan, Serbia
Ljubešić, Nikola, Slovenia
Magdalenić, Ivan, Croatia
Maina, Marcelo, Spain
Marton, Peter, Slovakia
Mekovec, Renata, Croatia
Meštrović, Ana, Croatia
Mijač, Marko, Croatia
Mijač, Tea, Croatia
Mršić, Leo, Croatia
Mu, Enrique, USA
Muštra, Mario, Croatia
Nyarko, Emmanuel Karlo , Croatia
Oreški, Dijana, Croatia
Oreški, Predrag, Croatia
Paralič, Jan, Slovakia
Pataki, Norbert, Hungary
Pergler, Elisabeth, Austria
Pihir, Igor, Croatia
Pištej, Rasťo, Slovakia
Pobar, Miran, Croatia
Prunera, Jaume Jordan, Spain
Rabuzin, Kornelije, Croatia
Radaković, Davorka, Serbia
Radošević, Danijel, Croatia
Radovanović, Sandro, Serbia
Rienties, Bart, United Kingdom
Ristić, Sonja, Serbia
Rodríguez García, Daniel, Spain
Rožanec, Jože, Slovenia
Schreiner, Wolfgang, Austria
Seljan, Sanja, Croatia
Sittar, Abdul , Slovenia
Stapić, Zlatko, Croatia
Starčević, Antonio, Croatia
Steingartner, William, Slovakia
Sturm, Jan, Czech Republic
Szymoniak, Sabina, Poland
Škorput, Pero, Croatia
Šumak, Boštjan , Slovenia
Tillmann, Alexander, Germany
Tomičić, Igor, Croatia
Tomičić-Pupek, Katarina, Croatia
Umek, Lan, Slovenia
Van Petegem, Wim, Belgium
Vidaček Hainš, Violeta, Croatia
Vidović, Ivan, Croatia
Virkus, Sirje, Estonia
Vrček, Neven, Croatia
Winkler, Till, Austria
Zornić, Nikola, Serbia
Žajdela Hrustek, Nikolina, Croatia
Žugec, Bojan, Croatia

Preface

We are pleased to present to you the Proceedings of the Central European Conference on Information and Intelligent Systems 2023. We trust that this compilation will prove to be both valuable and inspiring to you.

Contained within these Proceedings are the full text conference papers from the Central European Conference on Information and Intelligent Systems (CECIIS), an annual international event. This conference is organized by the Faculty of Organization and Informatics Varaždin, Croatia, and hosted by the University of Dubrovnik in Dubrovnik. The event took place from September 20th to 22nd, 2023. CECIIS has successfully convened for over three decades, with the aim to be the interface between scholars and professionals within the field of information and intelligent systems and related domains and the official language of the Conference is English.

Under the thematic focus of "Digital Human," CECIIS 2023 was designed to facilitate scholarly exploration and discourse on the transformative opportunities and challenges of digital technologies. The term digital human is included in a wide area of topics related to interaction of human beings in an evolving digital environment. Society is changing due to digital technologies while we, as humans, live part of our lives on social networks, metaverse and interact with artificial intelligence. ChatGPT is an important milestone in the development of AI but also in the evolution of digital humans. We can expect more advanced modalities of human augmentation and further immersion in the digital environment. At the most advanced level, we will have a digital human as a computer-generated representation of a human being. This can take many forms, from a simple avatar used in a video game to a highly realistic, photorealistic rendering used in films or virtual reality. The technology will create more advanced digital representations that are going to be increasingly indistinguishable from real humans. One of the key advantages of digital humans is that they can be created and customized quickly and easily. This allows a wide range of applications, from entertainment and gaming to training simulations and even medical and psychological research. However, the use of digital humans also raises some ethical questions. Further, the new digital dimension redefines our personal and social interactions impacting on our understanding of ourselves as humans, and on the relations with others and with the milieu. We are still unsure about the impact of these advanced technologies on education, labour market, economy, human rights and many other areas. In search of answers, we have to act within large-scale interdisciplinary collaborations and try to find the best approach to understand the future that has already started.

It is also pertinent to emphasize that the convergence of Program and Organizing Committee members, academics, and experts from diverse fields allows us to create a cross-disciplinary approach of the conference. This format propels discussions towards topics poised to emerge prominently in the near future, while concurrently facilitating valuable networking among individuals with varied domains of expertise, their collective insights contribute substantial value to the fabric of CECIIS.

The Proceedings of CECIIS 2023 encompass 59 meticulously reviewed scientific papers. These papers were subject to rigorous peer review by no fewer than two independent evaluators and subsequently endorsed by the international Program Committee. The papers span an array of categories including Computer Games, Data and Knowledge Bases, Digital Human, Digital Transformation and Corporate Culture, Education and Learning Analytics, ICT Entrepreneurship and Innovation, Intelligent Energy, Intelligent Systems, Quality in ICT, Strategic Planning and Decision Making, Systems for Smart Mobility, and Trends in Software Engineering. The selection of these papers hinged upon their relevance, innovativeness, applicability to the information and intelligent systems domain, and overall scholarly merit.

The review process itself has been conducted with the dedication of 104 reviewers, of which 48 were from Croatia and 56 from abroad. Collectively, they conducted 148 reviews, culminating in the acceptance of 59 papers. Authored by 161 individuals representing 16 nations, these papers were subsequently presented at the conference. Following the conference, a subset of chosen papers will undergo further review, potentially leading to extended publication within the Journal of Organizational and Information Sciences (JIOS), the international scientific publication published by the Faculty of Organization and Informatics.

A pivotal hallmark of this year's conference lies in the distinguished invited keynote speakers, individuals whose intellectual contributions have shaped their respective spheres of scientific and professional inquiry. CECIIS 2023 was honoured to host the following keynote speakers:

Professor Lode Lauwaert, Ph.D., who addressed the topic of "Ethics of risk in the context of AI". Professor Lauwaert is professor of technology at KU Leuven (Belgium), where he also holds the Chair Ethics and AI. His current research is on AI from the perspective of both environmental ethics and cross-cultural ethics.

Damir Bogdan, CEO of uptownBasel Infinity presented the topic "How Technologies of tomorrow will impact our Industries today - an overview about Quantum Computing". Damir Bogdan is heading a Center of Excellence for Quantum-computing and AI, QuantumBasel. He is also CEO of Actvide AG, a consulting company specialized on transformation within the digital age, leading C-Levels into the new future. He is engaged at Plug and Play, Silicon Valley's largest innovation platform. Damir Bogdan is multiple Board-Member within the Health Care, High-Tech and Industry sector. He is Senior Advisor at IWI University of St. Gallen and member of several Swiss startup committees and juries. He is a well-booked keynote speaker and panellist on digital transformation, cultural change, disruption, and startups.

In tandem with the scholarly sessions, the conference showcased two sections dedicated to poster presentations. Notably, this encompassed 6 scientific research posters and 20 student-generated posters. A collaborative effort involving 50 authors, these posters contributed to the intellectual tapestry of the conference.
A particularly insightful component of the program was the panel discussion titled "Science and AI", aptly moderated by Professor Neven Vrček, Ph.D., from the Faculty of Organization and Informatics University of Zagreb, Croatia. Esteemed panellists converged to share their perspectives on application of artificial intelligence in scientific context. Topics encompassed novel AI-driven technologies, associated opportunities, benefits, challenges, and their applications in scientific context.

We extend our sincere thanks to all contributors who enriched the Conference program and these Proceedings. This encompasses our esteemed keynote speakers, authors, reviewers, session chairs, as well as the Program and Organizing Committee members whose dedication was instrumental. The formidable support rendered by the Faculty of Organization and Informatics underpins the unequivocal success of the conference.

Lastly, profound appreciation is extended to our esteemed academic and industrial partners and sponsors whose unwavering support made CECIIS 2023 a resounding achievement. Collaboration with the Faculty of Transport and Traffic Sciences of the University of Zagreb, Faculty of Electrical Engineering and Computing of the University of Zagreb, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, and the University of Dubrovnik was central to the conference's realization. Financial support from sponsors including InfoDom d.o.o., SICK Mobilisis d.o.o., Oracle d.o.o., Factory X d.o.o. and Automobil Lončar was invaluable.

In conclusion, it is our sincerest hope that the participants of CECIIS 2023 found their experience enriching and productive, and that they find relevant and interesting papers in the following pages. The enduring connections forged with participants over the years reinforce our commitment to continuous improvement. Anticipation is already building for the forthcoming Central European Conference on Information and Intelligent Systems slated for the autumn of 2024.

September, 2023

On the behalf of the Program and Organizing Committees

Neven Vrček, Luis de Marcos Ortega and Petra Grd
Editors

Computer Games

Markus Schatten, Marinela Schatten, Ljiljana Inkret-Martinčević
Development of Gamified Cognitive Agents for Primary Schools - A Case Study

Bogdan Okreša Đurić, Tomislav Peharda, Pasqual Martí
Towards a Gamified System to Influence Behaviour of Users in the Context of Smart Mobility

Tomislav Peharda, Bogdan Okreša Đurić, Igor Tomičić
Towards Application of Programming Language for Communication Flows Specification in Multi-agent Systems on Real-World Use Cases

Development of Gamified Cognitive Agents for Primary Schools - A Case Study

Markus Schatten
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
markus.schatten@foi.unizg.hr

Marinela Schatten
Construction, science and mining school, Varaždin
Hallerova aleja 3, 42000 Varaždin, Croatia
marinela.schatten@skole.hr

Ljiljana Inkret-Martinčević
2nd Primary School, Varaždin
Augusta Cesarca 10, 42000 Varaždin, Croatia
ljiljana.inkret-martincevic2@skole.hr

Abstract. *In this paper we provide a first case study on the development and implementation of cognitive agents for primary school students' supports using a design science approach. Cognitive agents are artificial intelligence systems that are able to interact with humans using various technologies including but not limited to text to speech (TTS), speech to text (STT), natural language processing (NLP), belief-desire-intention (BDI), as well as motion capture (MoCap), lip sync and animation in order to provide for an immersive user experience (UX). We have developed four cognitive agents for a Croatian school especially for second, third and seventh grade pupils in collaboration with their teachers. The cognitive agents development process has additionally been gamified in sense that students have participated in it through a competition to create the face of the agent as well as to provide a name for it.*

Keywords. gamification, cognitive agents, primary school, artificial intelligence, croatian language, design science

1 Introduction

The advent of artificial intelligence (AI) has opened up new horizons in various domains, including education. Cognitive agents have gained major popularity through various systems by major software development vendors like Google (*Assistant*), Microsoft (*Cortana*), Apple (*Siri*) and Amazon (*Alexa*) as well as most recently various chatbots based on generative AI like *ChatGPT* (OpenAI) and *Bard* (Google). their application domains include Internet of things (IoT) and fog computing (Foukalas, 2020), education (Baylor, 1999), home service robots (Van Dang et al., 2017), mental health therapy (Suganuma et al., 2018), cognitive radio (Mitola, 2002) and many more. These are systems that can interact with humans using a variety of technologies, including but not limited to Text-to-Speech (TTS), Speech-to-Text (STT), Natural Language Processing (NLP), Belief-Desire-Intention (BDI) architecture, knowledge base (KB), as well as motion capture (MoCap), lip sync and animation, providing an immersive User Experience (UX) (Lee, 2010).

In our previous work, we have developed a cognitive agent architecture called Beautiful ARtificial Intelligence Cognitive Agent (B.A.R.I.C.A.) that is based on a number of AI techniques and technologies (Schatten, Okreša Đurić, and Peharda, 2021). We have also proposed a framework that combines cognitive agents, gamification practices, and telemedicine (Schatten, Okreša Đurić, and Protrka, 2021). Furthermore, we have developed the B.A.R.I.C.A. infrastructure, which allows for the implementation of open source cognitive agents that can communicate using the Croatian language (Schatten, Đurić, et al., 2022).

However, to the best of our knowledge, there has been no study that introduces gamified cognitive agents to primary school students. Therefore, this paper presents a first case study based on a design science approach on the development and implementation of cognitive agents for primary school students' supports, specifically for second, third, and seventh grade pupils in a Croatian school. The cognitive agents development process has additionally been gamified in the sense that students have participated in it through a competition to create the face of the agent as well as to provide a name for it.

The development of gamified cognitive agents is part of the project anon. The project aims to establish the first STEAM (where STEAM stands for *Science, Technology, Engineering, Arts, and Mathematics*) primary school in anon. During the project, the anon-primary-school currently undergoes a digital transformation through the STEAM concept, in-

cluding equipping four classrooms (including an astronomic observatory), comprehensive teacher education, and the development of curriculum and activities for students. The project is led by the `anon-school`, with partners from the `anon-faculty` and `anon-company` from the Kingdom of Norway and features the introduction of advanced technologies into the school's curricula including but not limited to augmented reality systems, robotics, asteroid tracing systems, game development, 3D printing as well as cognitive agents.[1]

The rest of the paper is organized as follows: Section 2 provides a review of the related work. Section 3 describes the methodology used in the development and implementation of the cognitive agents. Section 4 discusses our finding in four experimental usage sessions conducted in the 2nd Primary School in Varaždin. Finally, Section 5 concludes the paper and suggests future work directions.

2 Related Work

The use of cognitive agents in education has been explored in several studies. For instance, in previous work (Schatten, Okreša Đurić, and Peharda, 2021) we have developed a cognitive agent for university students that is able to communicate using spoken Croatian language and provides personalized support information including but not limited to schedule, office hours, campus information and similar. The agent does however not include learning materials. Another study by Sikström et al. (Sikström et al., 2022) found that cognitive agents can significantly enhance the learning experience in a virtual learning environment by providing personalized feedback and guidance.

Chatbots have also been utilized in the educational sector. A study by Kuhail et al. (Kuhail et al., 2023) found that chatbots can provide instant responses to students' queries, enhancing their learning experience in higher education. In the context of primary education, several studies have explored the use of chatbots. For example, a study by Putjorn (Putjorn, 2022) found that chatbots can enhance the English language skills of primary school students. Similarly, a study by Ghazali et al. (Ghazali and Saad, 2022) found that chatbots can significantly enhance the mathematics learning experience of primary school students. Lastly, a study by Othman (Othman, 2023) discussed the use of AI mobile application chatbots for teaching English as a Foreign Language (EFL) to primary school students in Saudi Arabia.

Despite these advancements, none of the above studies focus on the combination of gamification and cognitive agents. Moreover, none of these studies focus on learning materials provided in the Croatian language. This paper aims to fill this gap by presenting a case study on the development and implementation of gamified cognitive agents for primary school students in Croatia.

[1]More details abot the project are available here: https://steam.hr

3 Development Process

To implement gamified cognitive agents for primary school students we have adopted a design science approach. Design science is a research methodology in the field of information systems and technology that aims to create novel artifacts in the form of models, methods, and systems. These artifacts are designed to support people in developing, using, and maintaining IT solutions Johannesson and Perjons, 2014. The work of design science is applied to various fields of human practice, not limited to information systems and technology.

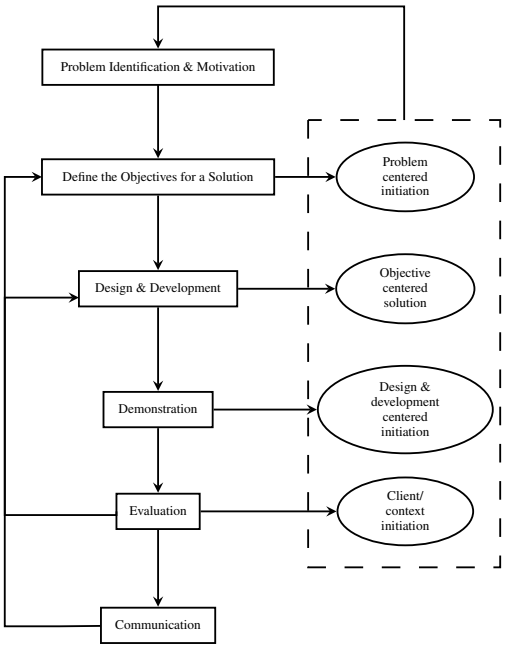


Figure 1: Design science methodology (Azasoo and Boateng, 2015) adapted from (Peffers et al., 2007)

Design science methodology (see figure 1) typically involves a number of main steps which are not strictly sequential and may be iterated over multiple times in a design science research project. Bellow is a description of these steps and how we have applied them in the development of our gamified cognitive agents:

1. **Problem Identification and Motivation:** This step involves identifying a problem that needs to be solved and establishing its relevance. It is crucial to understand the problem thoroughly and to articulate the motivation for solving it.

   In our case the problem to be addressed is defined in the `anon-project` objectives to which the development of cognitive agents contribute.

It involves enhancing the STEAM competencies of students across all primary educational levels. This shall be achieved through the integration of advanced technology and innovative teaching methods, including the organization of gamified workshops. These workshops will utilize specially developed cognitive agents and an augmented reality systems tailored for various STEM-related subjects. Additionally, the project aims to empower students with special educational needs by developing and piloting four new and improved extracurricular educational programs. These programs, equipped with specially designed learning materials for potentially gifted students or those with special educational needs, will be developed by teachers under the mentorship of partner experts in relevant fields. By the end of the project, these programs are expected to become an integral part of the school curriculum.

2. **Define the Objectives for a Solution:** Based on the problem identified, the objectives of the solution are defined. These objectives should be achievable and should directly address the problem.

   The primary objectives for the development of our gamified cognitive agents are outlined as follows:

   (a) *Ease of Educational Processes:* The foremost objective is to simplify the educational processes for primary school students, with a particular emphasis on students with special educational needs. The gamified cognitive agents are designed to facilitate learning and make it more engaging for these students.

   (b) *Ease of Use:* The cognitive agents are developed with a focus on user-friendliness. The interface and functionalities are designed to be intuitive and easy to navigate, reducing the learning curve for students and teachers alike.

   (c) *Cross-Platform Compatibility:* To ensure accessibility and ease of use from home, the cognitive agents should be implemented as web applications and are compatible with both mobile and desktop platforms. This flexibility allows students to engage with the learning material at their convenience, regardless of the device they are using.

   (d) *Immersiveness and Student Involvement:* To enhance the immersive learning experience, students are involved in the development process of the cognitive agents. They are given the opportunity to design the cognitive agents' faces and name them, fostering a sense of ownership and engagement in the learning process.

   These objectives guide the design and development process of the gamified cognitive agents, ensuring that the solution is tailored to meet the specific needs of the primary school students, particularly those with special educational needs.

3. **Design and Development:** In this step, an artifact is designed and developed to meet the defined objectives. The artifact could be a model, a method, or an instantiation.

   The design of the cognitive agents has been iterative in the sense that each agent prototype was developed in full and then tested and evaluated in a special class session with students and their teachers which were both involved in the development of the cognitive agent. The feedback from each session was then used for the development of the next cognitive agent prototype (with retrofitting all useful functionality into the previous instances).

   The development process had the following phases:

   (a) *Preparation phase* Firstly all teachers that were going to participate in the development process participated in a two day workshop on the use of AI in education with a special accent on generative AI, chatbots and cognitive agents.

   (b) *Teacher and student participation* Afterwards teachers were asked to prepare learning materials in form of questions the cognitive agents should be able to answer. Additionally, they were asked to organize a student competition and group activity with their students to draw a face picture for the cognitive agent which will be animated by the development team as well as to choose a name for the agent. The drawn face pictures and chosen names are shown on figure 2.

   (c) *Agent implementation* All agents' backends were developed in Python using the chatterbot module for an implementation of a regression based natural language processing (NLP) model and chatbot similar to the implementation provided in (Schatten, Okreša Đurić, and Peharda, 2021; Šokec, 2019) as well as Flask for the implementation of a web application. The frontends were developed using JavaScript especially using the SpeechRecognition application programming interface (API) that allowed us to capture speech and convert it to text. The face images were animated using CrazyTalk and the speech was generated using various online speech-to-text services. The agents were additionally tested by a teacher before

Figure 2: Cognitive agent's faces and chosen names (created by students of the 2nd Primary School Varaždin)

demonstration, and feedback was included into the implementation.

4. **Demonstration:** The artifact is then demonstrated to solve one or more instances of the problem. This could involve using the artifact in an experiment, a simulation, a case study, or other appropriate activity.

   For each cognitive agent a special class session was organized in which the students could test its abilities. The sessions were held by the teacher who organized the competition (usually the headroom teacher of the group), the teacher which tested it as well as one or two members from the development team. The session was organized in form of a role-playing game (RPG) in which each student assumed the role of the teacher and could ask the agent questions about the agreed topic and provide feedback about how well the answer was. After each session students were invited to provide feedback to the development form (in oral form for 2nd and 3rd grade students, in written form for 7th grade students). Additionally, after the session a discussion involving the teachers and the development team members was held in order to clarify additional shortcomings and potential improvements.

5. **Evaluation:** The artifact is evaluated to determine how well it solves the problem. This could involve observing the use of the artifact, conducting experiments, or performing analysis.

   In our case, the provided feedback was thoroughly analyzed and various comments and suggestions were included into the implementation. Some example suggestions include: (1) the possibility to write down or select the text for a given question since some (especially younger) students with pronunciation problems weren't recognized well by the speech recognition system, (2) include a button for starting the speech recognition since, due to background noise, sometimes the question was not recognized in full (i.e. cut in part) or connected to other parts of speech, (3) for topics that include mathematical formula or geometric shapes (especially mathematics and physics) to include visualization of the formula or shape being talked about).

6. **Communication:** The results are then communicated to the relevant audiences. This could involve publishing research papers, presenting at conferences, or producing technical reports.

   Apart from this paper, all the developed gamified cognitive agents have been published and are available for free use and testing on line[2]. Additionally, results have been communicated to relevant media outlets.

4 Discussion

The experimental sessions held during the demonstration phase of the study provided valuable insights into the effectiveness and potential of the gamified cognitive agents. Each cognitive agent was tested in a special class session, organized in the form of a role-playing game, where students assumed the role of the teacher and interacted with the agent. The feedback collected from these sessions, both oral and written, was instrumental in identifying areas of improvement and gauging the overall success of the initiative.

The most significant finding from these sessions was the increased level of student engagement. Teachers reported that student participation was much higher than usual, particularly among lower-grade students. This suggests that the gamified cognitive agents were successful in making the learning process more engaging and interactive. Furthermore, students expressed a high level of satisfaction and enthusiasm about being included in the development process and were eager to provide feedback.

In terms of students with special educational needs, the sessions had varying impacts. Students with learning disabilities were notably more engaged than usual,

[2]The agents are available on the following URLs https://dragon.foi.hr/micko/ https://dragon.foi.hr/zamrljan/ https://dragon.foi.hr/iva/ https://dragon.foi.hr/drslavica/

indicating that the gamified cognitive agents could potentially serve as an effective tool for this group. However, there was no significant increase in engagement among gifted students, although they were keen on providing feedback and suggestions. This suggests that while the cognitive agents may not necessarily enhance engagement among gifted students, they could still serve as a platform for these students to express their ideas and contribute to the learning process.

An interesting side effect was observed among students with pronunciation difficulties. These students made considerable efforts to pronounce each word correctly to ensure that the speech-to-text system could recognize it. This aligns with the findings of (Mitra et al., 2003), who reported that automated instructional approaches could improve English pronunciation. This suggests that the gamified cognitive agents could potentially serve as a tool for improving pronunciation and language skills among students.

5 Conclusion and Future Research

This paper presented a case study on the development and implementation of gamified cognitive agents for primary school students in Croatia. The initiative was part of the "Full STEAM Ahead!" project, which aims to enhance the STEAM competencies of students across all primary educational levels through the integration of advanced technology and innovative teaching methods. The cognitive agents were developed using a design science approach and were tested in special class sessions where students interacted with the agents and provided feedback.

The most significant finding from these sessions was the increased level of student engagement, particularly among lower-grade students and those with learning disabilities. This suggests that the gamified cognitive agents were successful in making the learning process more engaging and interactive. Moreover, the students demonstrated a strong sense of satisfaction and excitement due to their involvement in the development process, and they were proactive in offering their feedback. An intriguing observation was made with students who had difficulties with pronunciation. These students put in substantial effort to articulate each word accurately so that the speech-to-text system could correctly interpret it. This implies that the gamified cognitive agents could be used as an effective tool for enhancing pronunciation and language skills among students.

Looking ahead, our future research is aimed towards creating the next five iterations of gamified cognitive agents as part of the "Full STEAM Ahead!" project. These agents will be developed based on the feedback and insights gained from the initial implementation. We will continue to involve students in the development process, fostering a sense of ownership and engagement in the learning process. Furthermore, we will explore additional functionalities and improvements based on the feedback received from students and teachers. This includes the possibility of integrating visual aids for topics that include mathematical formulas or geometric shapes, and improving the speech recognition system to better accommodate students with pronunciation difficulties. Through these efforts, we aim to further enhance the learning experience for primary school students and contribute to the digital transformation of education in Croatia.

Acknowledgement

This work has been fully supported by the Full STEAM Ahead! project which was funded under the call "Strengthening STEM skills in primary schools and development of regional science centers for primary education in the STEM field" (cro. Jačanje STEM vještina u osnovnim školama i razvoj regionalnih znanstvenih centara za osnovnoškolski odgoj i obrazovanje u STEM području) from the EEA Financial Mechanism and the Norwegian Financial Mechanism for the period from 2014 to 2021.

We are very grateful to all students from the 2nd Primary School Varaždin, their teachers (Natalija Lončar, prof., Martina Dugonjić, prof., Goran Kreminski, prof. and Željka Kopić, prof.) which participated in the creation of the cognitive agents. Also, we would like to thank the members of the Artificial Intelligence Laboratory from the Faculty of Organization and Informatics (Bogdan Okreša Đurić, PhD and Tomislav Peharda, mag.inf.) which helped in the organization of the special sessions.

References

Azasoo, J., & Boateng, K. (2015). A retrofit design science methodology for smart metering design in developing countries. https://doi.org/10.1109/ICCSA.2015.23

Baylor, A. (1999). Intelligent agents as cognitive tools for education. *Educational technology*, 36–40.

Foukalas, F. (2020). Cognitive iot platform for fog computing industrial applications. *Computers & Electrical Engineering*, *87*, 106770.

Ghazali, N. D. S. M., & Saad, A. F. (2022). Online student performance system integrating multidimensional data visualization and chatbot for primary school. *International Journal of Artificial Intelligence*, *9*(2), 63–73.

Johannesson, P., & Perjons, E. (2014). *An introduction to design science* (Vol. 10). Springer.

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018.

Lee, I. (2010). *Encyclopedia of e-business development and management in the global economy*. IGI Global.

Mitola, J. (2002). *Cognitive radio. an integrated agent architecture for software defined radio.* (Doctoral dissertation). Kungliga Tekniska Hogskolan (Sweden).

Mitra, S., Tooley, J., Inamdar, P., & Dixon, P. (2003). Improving english pronunciation: An automated instructional approach. *Information Technologies & International Development*, *1*(1), pp–75.

Othman, K. (2023). Towards implementing ai mobile application chatbots for efl learners at primary schools in saudi arabia. *Journal of Namibian Studies: History Politics Culture*, *33*, 271–287.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45–77.

Putjorn, P. (2022). Designing augmented learning platform with iot and chatbot-based learning technology for primary school students in northern thailand. *2022 6th International Conference on Information Technology (InCIT)*, 275–279.

Schatten, M., Đurić, B. O., Peharda, T., & Tomičić, I. (2022). A cognitive agent's infrastructure for smart mobility. *Transportation Research Procedia*, *64*, 199–204.

Schatten, M., Okreša Đurić, B., & Peharda, T. (2021). A cognitive agent for university student support. *2021 IEEE Technology & Engineering Management Conference-Europe (TEMSCON-EUR)*, 1–6.

Schatten, M., Okreša Đurić, B., & Protrka, R. (2021). Conceptual architecture of a cognitive agent for telemedicine based on gamification. *Central European Conference on Information and Intelligent Systems*, 3–10.

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education*, *188*, 104564.

Šokec, T. (2019). Modeliranje kontekstualno svjesnog agenta za razgovor na hrvatskom jeziku uz pomoć konačnog automata i strojnog učenja [Mentor: Markus Schatten].

Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018). An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: Feasibility and acceptability pilot trial. *JMIR mental health*, *5*(3), e10454.

Van Dang, C., Tran, T. T., Gil, K.-J., Shin, Y.-B., Choi, J.-W., Park, G.-S., & Kim, J.-W. (2017). Application of soar cognitive agent based on utilitarian ethics theory for home service robots. *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 155–158.

Towards a Gamified System to Influence Behaviour of Users in the Context of Smart Mobility

Bogdan Okreša Đurić, Tomislav Peharda
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{dokresa,tpeharda}@foi.unizg.hr

Pasqual Martí
Universitat Politècnica de València
Valencian Research Institute for Artificial Intelligence
Camino de Vera, s/n 46022 Valencia. Edificio 1F
pasmargi@vrain.upv.es

Abstract. *One of the main concerns of modern transportation systems is the environmental impact of their operation. Such an impact is partially tied to the use customers make of the service. This paper proposes a conceptual framework to apply gamification techniques in the context of smart mobility, aiming to influence the behaviour of the users towards a more sustainable use of resources. The framework is exemplified by developing an example for a transport-sharing system. The requirements for an implementation as well as the implications thereof are analysed and discussed. Our work reflects the potential gamification has to enhance the sustainability of smart mobility solutions.*

Keywords. mobility, smart mobility, gamification, transportation, behaviour, sustainability

1 Introduction

Gamification is a broadly applicable concept of applying game design elements and mechanisms in non-game domains (Deterding et al., 2011; Huotari and Hamari, 2017), (Schell, 2019, p. 54). Some of these concepts may refer to point scoring, competition with others, rewards, and achievements, all with the ultimate goal of making the process more enjoyable. The mentioned concepts proved to bring more participation, motivation, productivity, as well as the desire for mastery (Hamari et al., 2014)(Goethe, 2019, p. 23). Furthermore, techniques of gamification can be utilised towards the goal of achieving the desired behavioural or psychological outcomes of the users (Matallaoui et al., 2017).

Some of the domains where gamification has been frequently used are education (e.g. (Kalogiannakis et al., 2021)), employee training (Iacono et al., 2020), customer engagement, etc. An application to learn a new language might be used as an example of using gamification in education (Huang and Soman, 2013). Gamification might be used in a way that, when users get to learn new words and phrases, they receive scoring points that help them level up, a good example of this is the language learning service Duolingo (Teske, 2017). The higher the level, the better they are (theoretically) in the language they learn. On a leaderboard, they can see how they stand compared to their friends.

In the domain of employee training, gamification could be used in different ways. An example could be that employees have a certain amount of literature to learn for compliance (Armstrong and Landers, 2018). After each section, they are asked questions to verify what they learned. Once they go through all the literature and quizzes, they are rewarded with a certification that acknowledges that they completed all necessary tasks for compliance.

Customer engagement is also a domain where gamification is being frequently applied. For example, after each purchase that exceeds a predefined amount, a customer might be getting loyalty scoring points (Harwood and Garry, 2015). Once the customer earns enough scoring points, they might exchange for different benefits, such as discounts, products, or similar.

There are many examples of gamification being used in the domain of medicine, with a rich collection of use cases specifically in the context of telemedicine (Schatten, Okreša Đurić, et al., 2021). Therein, gamification techniques are usually aimed at the elderly, although using gamification as a driver of engagement of the elderly telemedicine users is still a topic to be researched (de Vette et al., 2015).

Finally, gamification in the domain of transportation is a topic that has found some use cases where it is well known, yet there is room for further research and additional use cases to be developed (Okreša Đurić, 2022; Schatten, Okreša Đurić, et al., 2022), such as those presented in this paper. Although there are many contexts in the domain of transportation where some techniques of gamification are used already, most can be categorised as either referring to public transportation, or personal vehicles (Okreša Đurić, 2022).

One example beyond the two areas named above is related to installing and using personal physical-tracking devices (Greysen et al., 2021). Such applications motivate their users to choose healthier choices, promote physical activity of users, and motivate users

to engage in other types of behaviour, such as social engagement or sustainable behaviour.

Ultimately, gamification in transportation is no stranger to sustainability and promoting users' behaviour towards behaving sustainably, or to motivate users to use sustainable mobility options (Zinke-Wehlmann and Friedrich, 2019).

Nowadays, there is a great concern for the environment, and transportation has a relevant impact on it, as it consumes a great amount of energy and, in some cases, generates direct carbon dioxide emissions. Current transportation research takes that into account by including in the optimisation functions a factor that aims to increase the service sustainability (Mihyeon Jeon and Amekudzi, 2005). The present research proposes integration of gamification techniques in a transportation service with the aim of influencing the behaviour of its users towards a more sustainable mobility. The popularisation of smartphone technology and the thrive of on-demand mobility applications provide the basis for the development of interactive and user-tailored mobility solutions. Taking advantage of this, our goal is to develop a framework to reproduce a transport system and implement concrete gamification techniques on it, and then test its effectiveness in terms of service sustainability.

The rest of the paper is structured as follows. Section 2 comments on relevant works and technologies within the topic at hand. Then, Sec. 3 describes the conceptual framework, tailoring it to a transportation application. Section 4 presents the issues that implementing the framework would involve. The implications of gamifying mobility are discussed in Sec. 5. Finally, Sec. 6 concludes the work and outlines future research.

2 Related Work

This section briefly presents previous work developed in the areas of transportation, multi-agent simulation and gamification. In addition, SimFleet is presented, a simulation tool we aim to integrate in our framework to test and evaluate gamification techniques over realistic smart mobility scenarios.

The basis of transportation is to provide displacement services to those who require it. From a general perspective, a transportation problem involves optimising the use of resources to complete the displacement in the best possible way, according to a specific optimisation function. These resources combine the vehicles that provide the service as well as all the infrastructure the vehicles employ, such as road networks or power stations. The optimisation of the transportation service usually combines two perspectives (Vansteenwegen et al., 2022). On the one hand, the transportation provider aims to reduce operation costs. Furthermore, the users wish for a fast and reliable transportation. In general, transportation solutions offer a compromise between both perspectives, ensuring the economic viability of the service as well as guaranteeing a minimum service quality.

Because of the complex nature of transportation, multi-agent simulation is adequate to reproduce mobility services in virtual environments (Horni et al., 2016). In this work we depart from and adapt SimFleet (Palanca, Terrasa, et al., 2019), an agent-based transportation simulator that focuses on providing an environment for running and testing complex coordination and negotiation services in the domain of people and goods mobility. Additionally, SimFleet provides wide applicability, which is a welcome feature for this research. Vehicle fleets, which among others could be courier companies or taxi services, require high-quality coordination, in order to preserve efficiency, on-time, and economically sustainable services. Oftentimes it is very challenging to achieve the perfect balance of all three previously listed characteristics, making different negotiation strategies very useful.

In this research, fleet refers to a larger number of vehicles that have a goal to deliver a service. The proposed simulation tool involves three types of actors: customer agent, transporter agent, and fleet-manager agent. The behaviour of each of these agents can be fully coded by the simulator user. Thanks to that, the simulator allows for the representation of many different mobility schemes with various degrees of freedom for each actor. Moreover, the predefined communication flow can also be altered by the user, giving place to a highly configurable simulator that fits the requirements of the present research.

Research described in (Okreša Đurić, 2022) investigates what domains of smart and sustainable mobility are adequate candidates for the application of gamification concepts. The explored domains therein are: eco-driving, driving safety, marketing, vehicle user interfaces training, increasing social awareness, and increasing trust in autonomous driving. Mastering vehicle user interfaces shall greatly increase overall driver safety, as it would indicate that a driver is familiar with all functionalities of the vehicle, hence they can use them in an appropriate manner when needed. The aspect of social awareness could use gamification elements in a way that the driver gets recognised and appreciated when being respectful towards other drivers. Similarly, gamification could help promote more economical and ecological transportation habits, by rewarding the driver for using more eco-friendly vehicles, or by not engaging in mobility during rush hours.

Research presented in (Schatten, Okreša Đurić, et al., 2022) explores how cognitive agents could help in the domain of smart mobility. The authors provide conceptual solutions on how cognitive agents may enhance public transport, vehicle sharing, parking experience, crossroads safety, electronic toll systems, and variable information panels. In public transport, a cognitive agent might help with purchasing tickets, provid-

ing information about possible routes, checking on the arrival time, etc. Parking experience might be extended with a cognitive agent which would navigate a driver to a free parking spot. Similarly, at crossroads, a cognitive agent could help people with sight disabilities to inform whether they shall pass the street or not.

3 Conceptual Framework

The foundational part of the system conceptually described in this paper is the system providing transportation services, such as transportation sharing, public transportation, or private transportation. SimFleet simulator is used here as a specific system building on the idea of, and allowing for the simulation of, a transportation sharing system. Other concepts which form the basis of this framework include the users, vehicles, the environment, and the crucial addition: gamification techniques.

The main contribution of the framework described in this paper, shown in Fig. 1, is introduction of gamification techniques into the context of public and private transportation, especially when transportation sharing is considered. Gamification techniques have been proven to be able to influence behaviour of its users in various domains, e.g. (Kalogiannakis et al., 2021; Martinho et al., 2020; Tomičić and Schatten, 2020). Gamification is presented here as a sort of an added value concept over a working transportation-concentrated system, providing the benefits of influencing the users' behaviour.

The constituent parts of the system can be abstracted as agents (people or simulated agents), shown as solid circles in Fig. 1, in the context comprising vehicles and the surrounding environment (shown as squares in Fig. 1). The agents can be represented as pertaining to the belief-desire-intentions (BDI) paradigm. Such agents contain certain beliefs about the world surrounding them (knowledge about the context and other agents), desires that they strive to achieve, i.e. goals, and intentions that describe how the agent is planning to fulfil the defined desires, based on their beliefs. This approach allows for greater autonomy and decision-making capabilities in artificial intelligence systems by allowing agents to reason about their actions and make decisions based on their current state and goals. BDI agents have been used successfully in various fields such as robotics (Palanca, Rincon, et al., 2023), game theory (Ambroszkiewicz and Komar, 1999), and various expert systems.

In essence, BDI agents rely on their beliefs to construct their understanding of their surroundings, i.e. of the context. These perceptions are constructed under the influence of agent's perceptors, but they may be shaped by the application of a gamified interface, i.e. application of various gamification techniques, leading to altered desires that subsequently shape the behaviour of the agent. This interplay highlights one of the benefits of incorporating gamification into intelligent systems - specifically, enabling effective task execution through adaptive desire modification. In other words, even though the context might not actually change, applied gamification techniques are expected to induce change in agents' knowledge (their belief), and subsequently modify their desires, or their intentions of reaching an already set desire, towards a more sustainable behaviour.

Gamification techniques or mechanics that are here seen as convenient, suitable, and applicable include, but are not limited to, the following, described in detail below: (a) quests, (b) scoring points, (c) scoreboard, (d) avatar customisation, (e) in-app currency, (f) roles.

Quests provide the agent with an option they can use and utilise towards fulfilling some of their desires. A single quest may consist of multiple tasks. As agents engage with quests, they receive immediate feedback on their performance, which enables the agent to update its internal representation of the status of the system, and adjust its future behaviours accordingly.

Scoring system is a gamification technique that can be understood from the agent's point of view as earning rewards for achieving certain objectives, tasks, or simply behaving in a specific way. Thereafter, the interacting individuals acquire new information regarding their abilities to perform specific actions, e.g. the value thereof, within the simulated system. This acquired knowledge updates the agent's beliefs about their own capabilities and preferences related to shared mobility practices, steering future choices and promoting the desired behaviours associated with sustainable, efficient, and ecologically sound transport options.

Scoreboards can contribute to modifying an individual's belief structure via competition and comparison mechanisms. Agents viewing their rankings relative to others' performances can stimulate cognitive dissonance and the need for self-validation, leading to changes in beliefs surrounding their abilities and values toward sustainability-oriented mobility habits.

Avatar customisation can potentially transform agent's beliefs about its roles in promoting sustainable or green mobility. Through the process of personalising virtual representations depicted on the app's digital landscape, individuals are provided the opportunity to experiment with different identities and personas, exploring alternative perspectives relating to their relationship with the wider community and natural environments.

In-app currency could provide primarily people with the means of achieving tangible or sensed real-

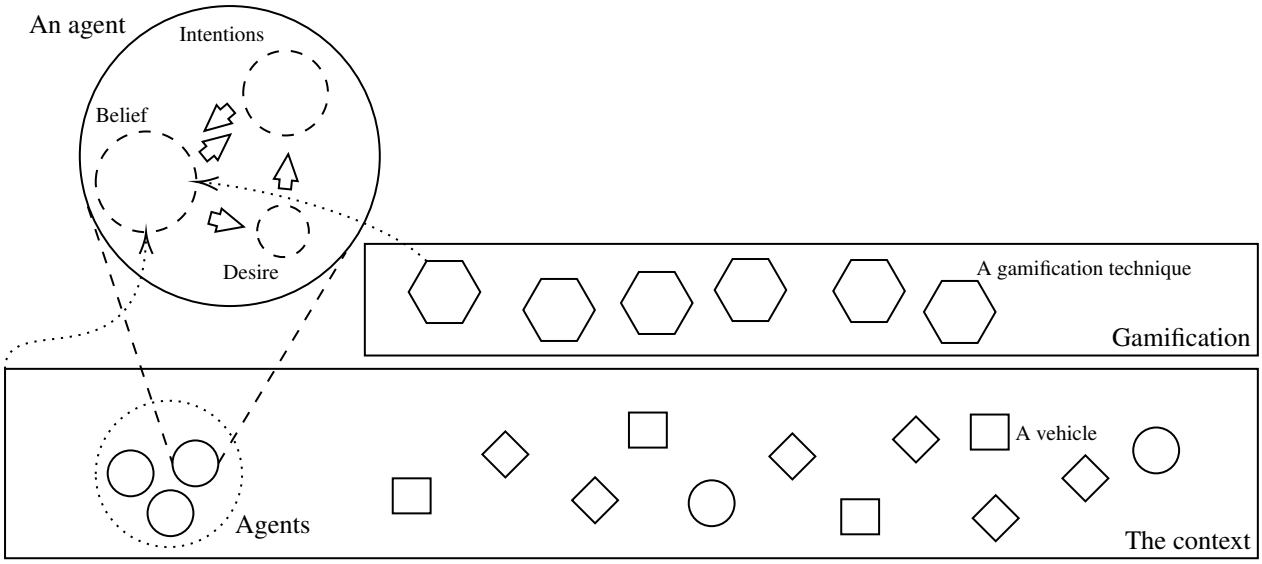


Figure 1: Visualised conceptual framework, showing the relationship of agents, the context, and gamification

life outcomes of their participation in the gamified system. Currency, in this aspect, can be considered as a specific case, or an outcome, of the scoring system. This mechanism is the most complex one to implement, as it would demand cooperation with real-life entities and organisations, which would recognise the currency and give it certain value, e.g. after accruing a certain number of points, the user is given a certain amount of currency, and can exchange a part of that currency at the local hair dresser's.

Roles can be used to further profile users based on their historical performance, i.e. on how they behaved in the past. Roles can be assigned automatically, or made available to the user, so they can activate, for example, at most one role at any given time. Some roles may grant users priority in choosing a vehicle to share, increase privacy in a shared vehicle by reducing the number of possible other travellers, or automatically choose pedestrian paths for them over e.g. cycling options. Roles may be though of as tightly integrated with avatar customisation.

The described framework is made to be general-purpose, wherefore the more detailed specific transport-sharing example is provided below.

3.1 Transport-Sharing Services

The aforementioned general-purpose framework is thought to be tailored to various transportation services. This, of course, requires the development of specific gamification techniques that fit the transportation mode. To exemplify a specification of the framework, a carsharing system has been chosen. Following, we describe the operation of such a system to later formulate the gamification proposals. Keep in mind that the term "car" may be substituted by variety of vehicles, for instance motorbikes or bicycles; the described operation can thus be applied to other types of transportation-sharing services.

In a carsharing service the transportation provider has a series of vehicles available for usage. Vehicles are located either in stations (bicycles) or parked anywhere inside a predefined area (motorbikes, cars). Users of the service are able to see the location of every available (not booked) vehicle and issue a booking to make use of one. The user is asked to leave the vehicle either in a station or correctly parked within a specified area. In general, the user pays either a monthly fee for the service or an amount per trip, computed from the total time and/or travelled distance.

The main costs of a carsharing operation come from vehicle relocation and energy consumption. Relocation is the displacement of a vehicle from its current station or location to a different one, aiming to improve the service by having the available vehicles well distributed. Energy consumption represents the cost of recharging/refueling the vehicles. These sources of costs are, in addition, closely related to the sustainability of the service, as the more relocations and energy consumption, the more impact does the service have on the environment. It is therefore in the best interest of both the transportation provider and society as a whole to influence the users of the service to make a better use of it.

3.2 Gamifying a Carsharing Service

Upon submitting desired trips through our application interface, sustainable alternatives are displayed that may prove to be slightly slower to complete, but incorporate increased pedestrian travel distances, reduced

traffic exposure, and circumvention of urban centres. Various points are assigned to these environmentally conscious choices, with the aim of engaging and motivating users in adopting greener transportation methods. Additional, if any, walking requirements resulting from these decisions are calculated, subsequently offering corresponding bonuses.

The system offers users alternative stations or areas where they can pick up or park their vehicle, thus incentivising users to accept and accomplish quests entailing redistribution journeys between regions with varying concentration ratios. The alternative location may incur a longer trip and or more walking distance to the user. They shall be rewarded accordingly.

We create a competitive atmosphere by providing unique customisation attributes based on accumulated scores; targeted challenges throughout holidays or other significant events can galvanize responsible conduct and inspire community interactions. By addressing several aspects of ride-sharing behaviour, we aspire to develop a more sustainable urban framework.

Scoreboards and customisation opportunities create a sense of community among the users and a bit of healthy competition. Therefore, a positive competitive atmosphere is created by providing unique customisation attributes based on accumulated scores; targeted challenges throughout holidays or other significant events can galvanize responsible behaviour and inspire community interactions. Score-based systems is expected to be an outcome of a list of challenges that should be obeyed if full points are expected. Finally, points may be exchanged for rewards, e.g. simple customisation options. By introducing challenges that promote certain kinds of behaviour, it might be possible to motivate users to adopt a more sustainable usage during specific time periods (e.g. holiday season).

4 On Implementing the Described Model

This section provides the authors' view on the data that would support the framework described in Sec. 3, and the simulation that would illustrate the framework's working.

4.1 Data Gathering

The gamification techniques described in Sec. 3 require various types of data as input to compute quests and tailor them to the user's usage of the transportation service. Moreover, additional user data may be shared both with the application and other users of the service, to access the customisation and scoreboard features.

The obvious prerequisites to a successful application of the described framework is a working transport-sharing system that features, at least, a demand prediction system, and a user behaviour learning system. Both of these require specific sets of data, yet both of these systems can provide output data that can be used as input for the various gamification mechanics, such as quests.

A demand prediction system is a crucial element in forecasting future areas of high demand and in motivating users to make trips that reduce energy consumption and empty vehicle movement (e.g. relocations). Based on the results of such a system, specific areas can be identified as more or less relevant, thus making them interesting in the context of devising quests, and pointing other gamification techniques towards them. On the other hand, a user behaviour learning system is identified as a fundamental input for creating user-tailored quests or rewards, based on the learnt trends of each user, their behaviour, and their interaction with the system.

4.2 Simulation

The described framework allows for the setup and execution of simulations that evaluate the effect of the proposed gamification techniques. To characterise the many actors of the transportation system, agent-based modelling (ABM) is employed. ABM allows for the definition of each actor as a rational entity. Specifically, we make use of BDI agents, designed to model human-like reasoning and decision-making processes.

The BDI agents are deployed in SimFleet. The key feature of the simulator is that it offers a flexible implementation of agent strategic behaviour. Each type of agent executes its own strategic behaviour, in which different types of users could be defined, studying the behaviour evolution according to their susceptibility to the gamification techniques. In addition, the simulations can be configured to collect all necessary data for us to evaluate and adjust the system.

As for the implementation of strategic behaviours, finite state machines (FSM) are chosen. An FSM defines a behaviour split in states. According to each specific state, the agent will be awaiting different inputs and may act in various manners. In general, an agent will begin its execution in an initial setup state and transition to a new one whenever a specific input is received. The agent's FSM may be cyclic, executing the same behaviour periodically; or one-shot, being executed until a goal is achieved. Finally, it is worth mentioning that SimFleet agents are able to execute, in parallel, more than one behaviour, thus allowing the programmer to split the processing of distinct types of input in distinct behaviours.

5 Discussion

The authors find it necessary to argue here in the direction of security and privacy, but also some other topics, such as ethics.

5.1 Gamification and User Trust

The first point of discussion is derived from the title of the paper, and is focused on influencing people and their behaviour, and the morals of such an approach. Although the science of influencing behaviour is not new, and this research is not the first to delve into the topic of influencing behaviour of people using a specific technique, it is beneficial to argue here that users should be made aware of the goals of using the system conceptually described in this paper, and that the ultimate purpose of it is to motivate a more sustainable and ecological transportation behaviour.

Building on the latter point, the system must be built, and communicated to users, in such a way that would ensure transparency and build awareness about the collected data and the possibility to deny the collection. Denying access to data would pose a substantial risk of reducing the initially promised set of features. Benefits of using the system, and allowing access to the necessary data, should therefore be communicated in a clear and user-friendly fashion. Building trust of users might be enhanced by open-sourcing the implementation.

The third point that is discussed here is related to the growing concern about explainability in the context of artificial intelligence and related models. Lack of explainability can lead to mistrust, lack of accountability, and difficulties in debugging misbehaving systems. It raises ethical concerns over transparency, fairness, and the impact of automated decision making on society. Ultimately, addressing this challenge requires concerted effort towards developing techniques and metrics to assess AI interpretability, ensuring alignment with human values and societal expectations, and fostering collaboration across disciplines to achieve explainable AI.

In order to give the users a piece of mind, and build trust, the feature of activating and deactivating specific features such as individual gamification techniques or user-tailored techniques, should be considered. Some users might be fine with participating in a scoring system, and being awarded points based on their performance, but they might not be willing to share their success with the general public. Therefore, giving them the option to deactivate specific gamification techniques, such as leaderboards in this particular example, might provide those users with a more fulfilling experience of using the provided system.

5.2 Social Awareness Potential

In addition to already described features of the selected gamification mechanisms, quest mechanics often introduce elements of social interaction and cooperation, further motivating agents to adopt more sustainable travel patterns (e.g. transportation-sharing services) and positively impacting their overall attitudes towards mobility. Implementing scoring points as part of a BDI agent design may prove advantageous in fostering the adoption of sustainable mobility habits among both individual users and society as a whole.

Presenting score as a scoreboard, with the results visible not only privately, additional effect are achieved. As participants strive to improve their standings, they may recognise the benefits of adopting efficient and low-carbon transport modes, gradually revising their belief structures and translating such modifications into actionable intentions to promote environmental friendliness across daily life activities involving mobility.

Avatar customisation is geared more towards people than software agents. The introspective journey provided by the possibility of customising their appearance towards others provides users with the ability to present themselves to other users in the desired fashion. Furthermore, it might instil a sense of environmental responsibility within individuals, ultimately contributing to revised belief frameworks around shared transportation practices and encouraging the adoption of pro-social attitudes favouring collective progress toward sustainability objectives.

5.3 Gamifying Transportation

With regards to the specific application field, the introduction of dynamic features to a transportation service must be carefully studied to avoid the distrust of its users. Traditional mobility services are generally expectable, basing their operation in restricted timetables. The more dynamic a mobility system is, the more it can be tailored to its users; however, dynamicity also incurs more demands towards the user, such as the explicit introduction of the desired trip, or the definition of desired travelling times, for instance. A balancing of dynamic and predictable features must be achieved in order to give the impression of a reliable yet adaptable service.

Finally, the introduction of gamification techniques, as commented throughout the paper, does not modify the reality but influences the user's perception and behaviour. The introduced mechanisms must take this into account to avoid negative user perceptions. On the one hand, the user should feel motivated but never forced to follow quests; similarly, they should have positive feelings towards their own score, but the fact of having a lower score with respect to other users should not bring up negative feelings on the user. Gamification is meant to complement but not dominate the mobility service, and it never should affect its quality. Too aggressive gamification features may end up frustrating the users and deteriorating the image they have of the transport service. This, in turn, would have a great negative impact on the service's economic sustainability.

6 Conclusion & Future Reasearch

The conceptual framework for a gamified system in the context of smart mobility is provided in this paper. The main goal of such a system is to influence behaviour of users towards a more sustainable and energy-efficient behaviour.

The framework described in Sec. 3 is exemplified in the subsequent sections, thus being shown in a more understanding manner on a specific case of transportation-sharing services, i.e. as an abstraction applicable to a gamified carsharing service. This example is intended to show the reader how the framework, which is designed with abstraction in mind, can be used in the context of smart sustainable mobility.

The framework is expected to be researched and developed in more detail in the future, as well as expanded and described using richer and more detailed examples that could be used for providing further arguments for and against the framework presented herein.

Acknowledgement

This work has been partially supported by the Croatian Science Foundation under the project number IP-2019-04-5824. Pasqual Martí is supported by grant ACIF/2021/259 funded by the "Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital de la Generalitat Valenciana."

References

Ambroszkiewicz, S., & Komar, J. (1999). A Model of BDI-Agent in Game-Theoretic Framework. In J.-J. C. Meyer, P.-Y. Schobbens, G. Goos, J. Hartmanis, & J. Van Leeuwen (Eds.), *Formal Models of Agents* (pp. 8–19). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-46581-2_2

Armstrong, M. B., & Landers, R. N. (2018). Gamification of employee training and development. *International Journal of Training and Development*, *22*(2), 162–169.

de Vette, F., Tabak, M., Dekker - van Weering, M., & Vollenbroek-Hutten, M. (2015). Engaging Elderly People in Telemedicine Through Gamification. *JMIR Serious Games*, *3*(2), e9. https://doi.org/10.2196/games.4561

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. https://doi.org/10.1145/2181037.2181040

Goethe, O. (2019). *Gamification Mindset*. Springer International Publishing. https://doi.org/10.1007/978-3-030-11078-9

Greysen, S. R., Changolkar, S., Small, D. S., Reale, C., Rareshide, C. A. L., Mercede, A., Snider, C. K., Greysen, H. M., Trotta, R., Halpern, S. D., & Patel, M. S. (2021). Effect of Behaviorally Designed Gamification With a Social Support Partner to Increase Mobility After Hospital Discharge: A Randomized Clinical Trial. *JAMA Network Open*, *4*(3), e210952. https://doi.org/10.1001/jamanetworkopen.2021.0952

Hamari, J., Koivisto, J., Sarsa, H., et al. (2014). Does gamification work?-a literature review of empirical studies on gamification. *HICSS*, *14*(2014), 3025–3034.

Harwood, T., & Garry, T. (2015). An investigation into gamification as a customer engagement experience environment. *Journal of Services Marketing*.

Horni, A., Nagel, K., & Axhausen, K. W. (2016). Introducing matsim. In *The multi-agent transport simulation matsim* (pp. 3–7). Ubiquity Press.

Huang, W. H.-Y., & Soman, D. (2013). Gamification of education. *Research Report Series: Behavioural Economics in Action, Rotman School of Management, University of Toronto*.

Huotari, K., & Hamari, J. (2017). A Definition for Gamification: Anchoring Gamification in the Service Marketing Literature. *Electronic Markets*, *27*(1), 21–31. https://doi.org/10.1007/s12525-015-0212-z

Iacono, S., Vallarino, M., & Vercelli, G. V. (2020). Gamification in Corporate Training to Enhance Engagement: An Approach. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(17), 69. https://doi.org/10.3991/ijet.v15i17.14207

Kalogiannakis, M., Papadakis, S., & Zourmpakis, A.-I. (2021). Gamification in Science Education. A Systematic Review of the Literature. *Education Sciences*, *11*(1), 22. https://doi.org/10.3390/educsci11010022

Martinho, D., Carneiro, J., Corchado, J. M., & Marreiros, G. (2020). A Systematic Review of Gamification Techniques Applied to Elderly Care. *Artificial Intelligence Review*, *53*(7), 4863–4901. https://doi.org/10.1007/s10462-020-09809-6

Matallaoui, A., Hanner, N., & Zarnekow, R. (2017). Introduction to Gamification: Foundation and Underlying Theories. In S. Stieglitz, C. Lattemann, S. Robra-Bissantz, R. Zarnekow, & T. Brockmann (Eds.), *Gamification: Using Game Elements in Serious Contexts* (pp. 3–18). Springer International Publishing. https://doi.org/10.1007/978-3-319-45557-0

Mihyeon Jeon, C., & Amekudzi, A. (2005). Addressing sustainability in transportation systems: Definitions, indicators, and metrics. *Journal of infrastructure systems*, *11*(1), 31–50.

Okreša Đurić, B. (2022). Gamification in Smart and Sustainable Mobility - A Review and Roadmap. In N. Vrček, L. Guàrdia, & P. Grd (Eds.), *Proceedings of the Central European Conference on Information and Intelligent Systems* (pp. 9–16). Faculty of Organization and Informatics.

Palanca, J., Rincon, J. A., Carrascosa, C., Julian, V. J., & Terrasa, A. (2023). Flexible Agent Architecture: Mixing Reactive and Deliberative Behaviors in SPADE. *Electronics*, *12*(3), 659. https://doi.org/10.3390/electronics12030659

Palanca, J., Terrasa, A., Carrascosa, C., & Julián, V. (2019). Simfleet: A new transport fleet simulator based on mas. *Highlights of Practical Applications of Survivable Agents and Multi-Agent Systems. The PAAMS Collection: International Workshops of PAAMS 2019, Ávila, Spain, June 26–28, 2019, Proceedings 17*, 257–264.

Schatten, M., Okreša Đurić, B., Peharda, T., & Tomičić, I. (2022). A Cognitive Agent's Infrastructure for Smart Mobility. In M. Petrović, I. Dovbischuk, & A. L. Cunha (Eds.), *Transportation Research Procedia* (pp. 199–204). Faculty of Transport and Traffic Sciences.

Schatten, M., Okreša Đurić, B., & Protrka, R. (2021). Conceptual Architecture of a Cognitive Agent for Telemedicine based on Gamification. In N. Vrček, E. Pergler, & P. Grd (Eds.), *Proceedings of the Central European Conference on Information and Intelligent Systems* (pp. 3–10). Faculty of Organization and Informatics.

Schell, J. (2019). *The Art of Game Design: A Book of Lenses* (3rd ed.). Taylor & Francis Group.

Teske, K. (2017). Duolingo. *CALICO Journal*, *34*(3), 393–401.

Tomičić, I., & Schatten, M. (2020). Conceptual Network Analysis of Gamification Practices in Primary and Secondary Education. *MIPRO 2020 43rd International Convention Proceedings*, 1335–1341.

Vansteenwegen, P., Melis, L., Aktaş, D., Montenegro, B. D. G., Vieira, F. S., & Sörensen, K. (2022). A survey on demand-responsive public bus systems. *Transportation Research Part C: Emerging Technologies*, *137*, 103573.

Zinke-Wehlmann, C., & Friedrich, J. (2019). Commute Green! The Potential of Enterprise Social Networks for Ecological Mobility Concepts. In L. M. Camarinha-Matos, H. Afsarmanesh, & D. Antonelli (Eds.), *Collaborative Networks and Digital Transformation* (pp. 128–139). Springer International Publishing. https://doi.org/10.1007/978-3-030-28464-0_12

Towards Application of Programming Language for Communication Flows Specification in Multi-agent Systems on Real-World Use Cases

Tomislav Peharda, Bogdan Okreša Đurić, Igor Tomičić
Artificial Intelligence Laboratory
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{tomislav.peharda, dokresa, igor.tomicic}@foi.unizg.hr

Abstract. *In this work-in-progress article we focus on the application of programming language for communication flows specification in multi-agent systems on real-world use cases. Agents orchestration in multi-agent systems architecture may be very troublesome, mainly due to agents being independent units that may be implemented differently. The proposed programming language for communication flows specification attempts to overcome this challenge by providing explicit communication flows definitions, that agents' communications component relies on, which enables enhanced orchestration capabilities. Use cases that are covered in this paper are in the domains of streaming audio, video, and sensors data, and provide a couple of examples of how the proposed programming language may be used.*

Keywords. multiagent systems, communication flows specification, agent orchestration

1 Introduction

It is often the case within computer science that there is a type of task that is being executed repeatedly to either deliver a resource or simply do a job that shall impact the requesting party (directly or indirectly). A software term that is used to describe such a program is agent. The default definition of an agent (Russell and Norvig, 2022, p. 54) sets it as anything that is located in an environment, and can perceive it, and act upon it. One of the main characteristics of an agent, though, is its autonomy in performing a task, which means it acts as a standalone unit that is adequately designed to achieve the goal without human directions or intervention.

If a task is too complex for a single agent to complete it, the task is broken down into multiple sub-tasks, in which case each sub-task is being executed by a separate agent. In this scenario, agents complete their sub-tasks and communicate the results to other agents. This way, all agents collaborate to achieve the common goal. A cluster of interconnected agents is called a multiagent system (MAS).

Since agents are autonomous units, from a technical standpoint, each agent might be designed and implemented differently. When speaking of agent coordination and orchestration in MAS architecture, this may bring up several challenges, one of which is inconsistency in communication flows (Schatten, Tomičić, et al., 2020). For example, agent A is designed and implemented to communicate with agent B, however, agent B is not designed to communicate with agent A. In this case, agent orchestration will not be successful, as agents won't participate in the communication and will not be able to achieve the higher goal.

The herein proposed solution to this problem is a new programming language based on process calculus (Parrow, 2001, p. 481) for communication flows specification. The language shall implement concepts and mechanisms that enhance agents orchestration capabilities in MAS architecture.

The goal of this paper is to provide examples of how programming language may be utilized in the use cases of data streaming. Significance is also put on what possibilities would an engineer be provided with by using the language, and how it compares to the traditional approaches to agent design and implementation.

The rest of the paper is structured as follows. Section 2 provides information on related work, basic concepts necessary herein, and similar research. Section 3 describes the programming language proposed as a result of this research, followed by Sec. 4 where the programming language use is exemplified in a real-world use case. Finally, Sec. 5 provides a short discussion, conclusion, and some ideas that provide opportunities for future work.

2 Related Work

Regarding message transport, Corkill (Corkill, 1988) suggests a broadcasting concept in which each agent

communicates with all agents in a MAS. In this case, the expectation is that each agent shall implement a filtering mechanism on the receiving side to discard messages that are not of its interest. Although flexible, this approach does not consider performance implications so much.

Similarly, Goldman, and Ziberstein in (Goldman and Zilberstein, 2003) build on top of the previously described idea with the addition of taking into account that external agents are not part of the observed MAS, and likely do not have a filtering mechanism in place. To address this, the authors propose that there is a middleware agent in the MAS to which all agents send their messages to, and the middleware agent is responsible for routing the received messages to the appropriate parties.

Berna-Koes, Nourbakhsh, and Sycara in their research (Berna-Koes et al., 2004) focus on transporting non-textual messages such as video, audio, or other media types, which are not adequately supported by most MAS frameworks as they primarily rely on textual messages. Their research proposes the introduction of backchannels, enabling agents to exchange messages of these media types.

On the other hand, service orchestration providers are also relevant to the subject of this research. Some of these providers are Kubernetes, and Docker Swarm. Basically, what these providers enable is infrastructure for the coordination of a larger number of services at once. However, the infrastructure support these providers enable is on a high level and refers to enabling services to share resources, storage, network, and enable communication (Powell, 2023). Essentially, it is about characteristics that all services share, regardless of their purpose or type. Since communication (protocol, language, etc.) differs from one service type to another, it is considered a low-level component, thus is something that these providers do not take into account.

In the research presented in (Schatten, Tomičić, et al., 2020), the authors describe a conceptual model for the orchestration platform for hybrid artificial intelligence methods that is based on a MAS architecture. The idea is that the platform enables orchestration of various game engines via application programming interfaces (APIs). Essentially, APIs are used for external users to start up game engines that they wish to use, while these game engines are being treated as agents.

Holonic systems (Rodriguez et al., 2011) in MAS architecture refer to an entity that may be observed as a standalone autonomous agent or as a piece of a larger system. As a standalone agent, holon has built-in business logic by which it acts independently and correspondingly communicates with other external agents. In the context of a larger system, it is collaborating with other internal peer agents to achieve a common goal. Such hierarchy enables the construction of a more complex organizational structure which has proved to be helpful when a complex task needs to be broken down into multiple sub-tasks.

Process calculus (Parrow, 2001) stands for a family of approaches for modeling concurrent systems with a focus on describing interaction, communication, and synchronization between agents. The main features that all approaches within process calculus share are message-passing between the independent processes that represent interactions, the use of a small number of primitives, operators to describe processes, and specific algebraic rules for process operators. Reduction rules are an essential aspect of process calculus as they help achieve the explicit description of communication that takes into account parallel composition, sequentialization, and transformation of some input into an output. Process calculus provides a formal backdrop for modeling complex distributed systems and shall be used as a theoretical foundation for the research at hand. π-calculus, which is part of the Process calculus family is used as a formal backdrop for the construction of the programming language.

3 Programming Language for Communication Flows Specification

To solve the problem of agents orchestration in MAS, the proposal is to use a programming language for communication flows specification that adds support for consistent flows specification across all agents. The high-level idea behind the programming language is that by creating communication flows specification, which essentially describes what pairs of agents engage in communication, each agent is provided with a list of communication flows it needs to implement.

The programming language, which is still being developed, is based on process calculus. Process calculus is used as a foundation for this programming language, as it embodies characteristics for communication specification, which takes into account parallelism, sequentization, and input-to-output transformation, which fit very well with the communication flows specification.

With the programming language comes a declarative engine, which is capable of processing communication flows specification. Once communication flows specification is parsed, each agent is provided with a list of communication flows it shall implement, which ensures proper agent orchestration capabilities. Ultimately, this solves the problem of communication flows inconsistency between agents, where for example, agent A might be designed to communicate with agent B, but agent B is not designed to communicate with agent A. By adding safeguard in the agent codebase, it can be easy to identify if an agent is not implementing some of the communication flows, which may especially come handy if communication flows specification gets updated in the future.

Another valuable feature of the communication

flows specification is that it can serve as a source of truth for describing all communication flows in an observed MAS. Without communication flows specification, to get an idea of all communication flows in MAS, one should look into every agent codebase individually to determine its communication flows.

Three aspects of communication that the programming language is set to enhance are (a) specifying pairs of agents that communicate, (b) communication protocol, and (c) expected input and output. These aspects provide great grounds for building more advanced communication channels. Specifying pairs of agents enables a description of one-way communication between two pairs of agents. Each agent might have multiple communication flows defined where in some cases it might be in the role of a receiving party, while in others it may be in the role of the sender party. By using different symbols for indication of communication between agents, it is possible to specify what type of protocols shall be used for the communication. Currently, transmission control protocol (TCP) and user datagram protocol (UDP) protocols are supported. Lastly, defining expected input can be useful to announce that a particular agent is only capable of processing messages received in a particular format, which is helpful from the perspective of data integrity and validation. Similarly, defining expected output makes sure that other agents are acknowledged in what format should the incoming message be.

Types of agents that are supported by the programming language are: basic agent, channel agent, and holon agent. The basic agent may only send and receive messages from the channel agent and the holon agent. The channel agent serves as a proxy, with capabilities to filter and transform input to output. The holon agent is conceptually designed to serve as a gateway to agents outside the observed MAS. The holon agent is anticipated for communication with agents outside the observed MAS.

4 Application of Programming Language to Real-world Use Cases

In this section, we provide a description of how to the application of the programming language could be used to specify communication flows in a MAS.

4.1 Communicating Sensors

The use case is as follows: there are three basic agents, agent B, C, and D, which are sensors capturing temperature value every 5 minutes. These three sensors are located within the same area, however, within 5 kilometers radius. The idea is that by looking at the average value of the captured temperatures from these 3 agents, we can gain insight into the temperature of the observed area.

The three agents are solely responsible for capturing temperature. Data they collect they send to the agent A which is a channel. In the programming language, each basic agent interface utilizes the reduction rule to specify the sender and the receiver in the format sender → receiver. If the basic agent specification, keyword self is used for implication of the currently observed agent. If the keyword self is used as a sender in the communication flow specification, that implies that the currently observed agent shall send a message to the receiving counter-party. As previously said, within the programming language, channels are used to optionally transform and proxy data to the interested parties. Each of the sensors agents send a message in the JSON format `{"temperature": <numeric value ↪ >, "location": <name of location>}`. That being said, the input of the channel is also defined to take an input in this exact format. This is defined by the reduction rule in the format input → output which indicates the transformation. If a message of a different format is received, it will be dropped. Whenever the channel receives a message in the adequate format, it parses the message and extracts out only the value, which also indicates it drops the location attribute. The channel sends out the transformed message.

In this use case, there is a basic agent D which is listening to any messages emitted by the channel A. Agent D is responsible for creating an average value of temperature values it receives in the window of 5 minutes and sharing the reckoned value to the audience. For example, this could be achieved by sending an email, posting a message on a messaging application, or similar.

The snippet provided below is a description of the communication flows specification in the described multi-agent systems architecture-based use case in the proposed programming language. Figure 1 provides a visual representation of the communication flows in the described MAS.

```
channel A:
    json({"temperature": ?temp, "location
        ↪ ": ?loc}) --> ?temp

agent B:
    self -> A

agent C:
    self -> A

agent D:
    self -> A

agent E:
    A -> self
```

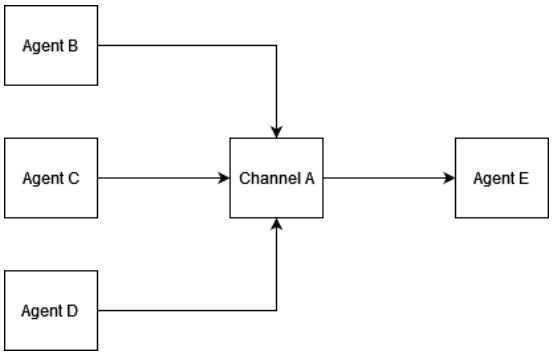


Figure 1: Communications flows

4.2 Streaming Game Platform

The second use case presented here uses a streamed holographic 3D game engine described in more detail in (Schatten, Okreša Đurić, et al., 2022). The game engine described therein makes it possible for players to stream games over a 3D holographic gaming console. The game engine, as well as the implementation of the gaming console, is implemented using a MAS, since the required level of synchronisation and complexity fits tightly into the observable natural application domains of systems comprising agents.

The above referenced streamed holographic 3D game engine's prototype architecture, shown in Fig. 2, features at least three types of agents: (a) a game agent, (b) a game streaming agent, (c) a videoconferencing agent. The game agent is running the game, and making sure the camera transformer is used correctly, so the game is shown as intended on the holographic 3D gaming console. The game streaming agent makes sure that the game is streamed towards the player, i.e. that the streaming channel is established and maintained towards the player, and that the player is provided with the satisfying gaming experience. The videoconferencing agent implements a group videoconferencing option that provides the player with the possibility of sharing a link towards the public or a select number of individuals, who could watch the stream, join in a group discussion on the player's performance, and watch the player's camera stream alongside the game stream.

Based on the above description, every single player's experience can be described as being the result of a process realised as a group effort of a combination of multiple agents, i.e. a multiagent system. As such, it can be described using the programming language described earlier in this paper, in terms of communication flows specification. In describing this example in the context of the programming language for communication flows specification, only a single communication flow is considered: a signal is sent from the user interface towards the game streaming agent, forwarded towards the game agent. Furthermore, a specific example of interacting with the game is observed in more detail, in

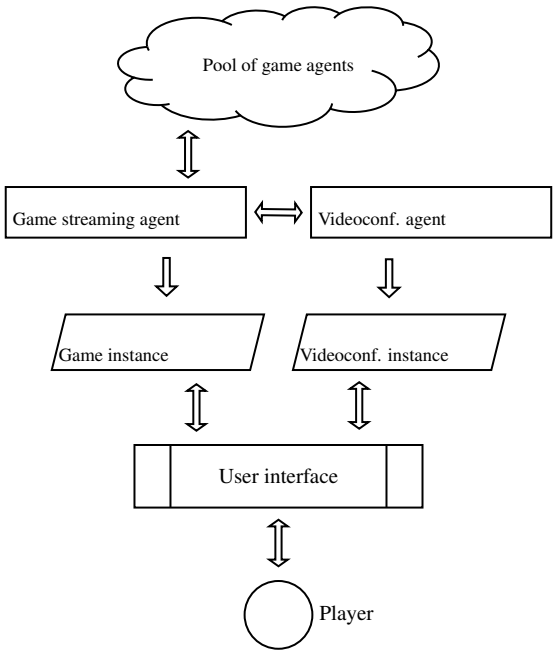


Figure 2: Architecture prototype of the streamed holographic 3D game engine (Schatten, Okreša Đurić, et al., 2022)

order to keep the example in the constraints appropriate for this paper.

Therefore, three agents can be observed: the user interface (agent UI), the game streaming agent (agent S), and the game agent (agent G). User interface can be modelled as an agent for the purposes of modelling the system using the proposed programming language, even though the implementation may not follow this specification, and the user interface is not implemented as a standalone agent similar to the other two, but as a more abstract form of a reactive agent that is not autonomous.

Agent UI sends player-activated controls from a controller to agent S. The game streaming agent (agent S) takes input from the user interface, and translates it into the correct command that is expected by the game agent. Usually, the game agent is expecting one of the commands in a predefined set of inputs that are used for controlling the player or other in-game elements, written in a configuration file, referenced as `config` in the listing below. Translation is performed using a list of expected input values and the appropriate output values, such as is shown here, where the event from the controller is listed as the key, and the appropriate action is listed as the value of the key-value pair:

```
{
    "controls":
        {
            "UP": "up",
            "DOWN": "down",
            "LEFT": "left",
            "RIGHT": "right",
```

```
            "SELECT": "esc",
            "START": "enter",
            "A": "space",
            "B": "enter"
        }
}
```

The communication flow described above can be modelled using two channels (UIS and SG), and three agents (UI, S, G). The first channel is used by agent UI to send user input to agent S (channel UIS), and the second channel is used to send the appropriate command from agent S to agent G (channel SG). Translation of the received user input can be performed either as a part of an agent, or as a part of a channel. Since a channel is implemented as a specific form of an agent, translation action can be given to the channel, as is shown for the purposes of this example. The channel translates the input into the command based on both of the received values, and the applicable configuration file. The listing below can be used to describe the example:

```
channel UIS:
    json({"key": ?key, "player": ?player})
        ↪ --> json({"key": ?key})

channel SG:
    json({"game": ?game, "key": ?key}) -->
        ↪ config[?game][?key]

agent UI:
    self -> UIS

agent S:
    UIS -> self & self -> SG

agent G:
    SG -> self
```

5 Conclusion and Future work

The proposed programming language, and the corresponding declarative engine are taking a holistic approach to solving the problem of agent orchestration and coordination in the multi-agent systems architecture. The programming language's main focus is put on enhancing communication flows specification, which in the context of MAS implies to agent orchestration. Some of the identified flaws of the traditional approaches to the communications flows specification primarily deal with communication flows inconsistency across different agents, which leads to challenges in the orchestration.

Programming language based on π-calculus is designed to solve this problem by introducing communication flows specification, which shall provide description of communication flows of the entire MAS in a single specification flow. This also gives visibility into all communication flows present in MAS, which would otherwise be impossible, as one would need to look into each agent separately to determine communication flows. The declarative engine, which is essentially in the role of a parser extracts out valuable information for each agent separately, so that each agent might be provided with the list of communication flows it needs to implement.

The above is one of the benefits of the implementation approach using the proposed programming language, as opposed to the traditional approaches to agent design and implementation. Therein, there is a high possibility for a higher degree of code redundancy and duplication, especially with regard to the communication component, as each agent is required to implement it individually. Additionally, there is a possibility for inconsistency in communication flows between agents.

By being provided with per-agent-parsed list of communication flows to implement, each agent is in a good spot to achieve proper orchestration within the MAS it resides in. By implementing communication mechanism for all the required communication flows, all agent communication shall be consistent, as there should be no discrepancies where one agent might be designed to send message to another one, while the other is not designed for receiving the message.

Using the described example in the domain of sensors capturing temperature, we have shown how the utilization of different types of agents supported by the programming language may conceptually lead to proper orchestration, and ease of communication flows specification. The example also touches on the flexibility of agents business logic, as agents may have different purpose, which implies that the programming language in no way attempts to limit the business logic, but only to enhance the communication component. This is further emphasised using the second provided example.

The future work of this research is in the direction of designing and developing an orchestration platform that utilizes the communication flows specification, and starts up agents based on it. The platform intends to take the current work a step further, as with the programming language solely, agents are provided with the communication flows that are needed to be implemented. The idea around the orchestration platform is that whole agent communication would be covered by the platform itself by implementing the communication layer around an agent. This essentially means that with the communication flows specification being in the place, the only needed part for agent to work properly would be to specify what business logic should an agent do.

Acknowledgement

This work has been fully supported by the Croatian Science Foundation under the project number IP-2019-04-5824.

References

Berna-Koes, M., Nourbakhsh, I., & Sycara, K. (2004). Communication efficiency in multi-agent systems. *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, *3*, 2129–2134.

Corkill, D. D. (1988). *Design alternatives for parallel and distributed blackboard systems*. University of Massachusetts at Amherst. Computer; Information Science [COINS].

Goldman, C. V., & Zilberstein, S. (2003). Optimizing information exchange in cooperative multi-agent systems. *Proceedings of the second international joint conference on Autonomous agents and multi-agent systems*, 137–144.

Parrow, J. (2001). An Introduction to the π-Calculus. In J. A. Bergstra, A. Ponse, & S. A. Smolka (Eds.), *Handbook of Process Algebra* (1st ed., pp. 479–542). Elsevier.

Powell, R. (2023). *Docker Swarm vs Kubernetes: How to choose a container orchestration tool*. CircleCI. Retrieved June 28, 2023, from https://circleci.com/blog/docker-swarm-vs-kubernetes/

Rodriguez, S., Hilaire, V., Gaud, N., Galland, S., & Koukam, A. (2011). Holonic multi-agent systems. In *Self-organising software* (pp. 251–279). Springer.

Russell, S. J., & Norvig, P. (Eds.). (2022). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson Education Limited.

Schatten, M., Okreša Đurić, B., & Peharda, T. (2022). Towards a Streamed Holographic 3D Game Engine. In N. Vrček, L. Guàrdia, & P. Grd (Eds.), *Proceedings of the Central European Conference on Information and Intelligent Systems* (pp. 17–22). Faculty of Organization and Informatics.

Schatten, M., Tomičić, I., & Okreša Đurić, B. (2020). Orchestration Platforms for Hybrid Artificial Intelligence in Computer Games – A Conceptual Model. In V. Strahonja, W. Steingartner, & V. Kirinić (Eds.), *Central European Conference on Information and Intelligent Systems* (pp. 3–8). Faculty of Organization and Informatics, University of Zagreb.

Data and Knowledge Bases

Božidar Kovačić, Vanja Slavuj, Miodrag Sretenović
Data Visualisation System for E-learning: Overview of System Design

Data Visualisation System for E-learning: Overview of System Design

Božidar Kovačić, Vanja Slavuj
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, 51000 Rijeka, Croatia
{bkovacic, vslavuj}@uniri.hr

Miodrag Sretenović
MI&DA Ltd.,
Šulekova 29, 47000 Karlovac, Croatia
m.sretenovic@mida.hr

Abstract. *One of the ways in which distance teaching and learning practice and results may be improved is to collect relevant data from various student activities performed during a course and analyse them so as to be able to draw relevant conclusions. Such data usually includes, but is not limited to, the data obtained from evaluation activities aimed at evaluating the set learning outcomes. The collected data is subjected to the defined learning analytics procedures by both the experts (e.g., educational decision-makers) and non-experts (e.g., course instructors) in learning analytics, so a visual representation of data may help in increasing the transparency and comprehensibility of the results. Designated learning data visualisation systems play a vital role in that regard and may offer a number of customised graphical displays of data.*
The paper at hand argues in favour of learning data visualisation helping to design better and more successful distance learning setups. It aims to describe a newly developed system for learning data visualisation which uses the data collected from a learning management system and allows it to be automatically prepared, analysed, and visualised for educational decision-making. The emphasis in the paper is put on the design and implementation of the system (all of its components), its back-end elements, and database design.

Keywords. data visualisation, database design, distance learning, e-learning, educational decision-making

1 Introduction

Possibly the main motivation for developing systems that provide visualisations of learning data is to use the data to create a better and more in-depth understanding of the learning process (Sedrakyan et al., 2019). Thus, visualising data related to different types of tests and graded learning activities within an e-learning system offers an opportunity to more easily perceive and better comprehend large amounts of data (Lowe & Matthee, 2020), to quickly and efficiently interpret the data (Donohoe & Costello, 2020), to make informed decisions about the learning and teaching process (Mavrikis et al., 2019), and to identify and monitor learning trends revealed by the data. This, in turn, makes a solid basis for increasing the quality and success of knowledge acquisition (Lang et al., 2022).

As is the case with other tools for online or distance learning, learning management systems (LMSs) gather data regarding the results of graded learning activities, allowing export of data in a number of standardised data formats (Black et al., 2008). Using this type of data, teaching and learning experts are able to evaluate students' learning achievement in a more precise and informed way, as well as to improve their understanding of the educational setup by:

1. visualising data on graded activities, and
2. analysing the learning success of individual students, groups of students, or entire student populations for different graded activities.

The goal of this paper is to present the overall architecture of the system for learning data visualisation that we designed and developed to extend the possibilities of an LMS used at our university, and to report on the progress of its development. A previous paper (Kovačić, 2022) focused solely on the implementation of the data import module (including the underlying database design) which uses exported LMS files with data on graded learning activities. This paper extends the system description and presents how the system is designed in order to prepare customisable visualisations for different activity types, emphasising the implementation of other modules necessary for data visualisation.

This paper is structured as follows. Section 2 gives an overview of background work and describes the current state of development of the system for learning data visualisation. Section 3 elaborates on the design of the data visualisation system, specially emphasising the extension and implementation of its database to accommodate novel functionalities for data preparation, import, and analysis. Section 4 concludes

the paper and charts future developmental paths for the system.

2 Background Work

In this section we provide an overview of related work on systems for data visualisation and describe the system for learning data visualisation developed at our institution as it currently stands.

2.1 Literature Overview

The review of relevant scientific sources reveals that learning data visualisation, in the context of distance or online education, is a thriving area of research. It attracts practitioners who are concerned with the implementation of data visualisation systems (e.g., learning dashboards) on the one side, and educators who employ them as part of their teaching setups on the other. This area or research often relies on the broader fields of learning analytics and educational data mining by using their techniques to prepare the data for visualisation itself (Romero & Ventura, 2019). Researchers have stressed the importance of data visualisation literacy in both personal and professional life as crucial for making sound decisions (Börner et al., 2019).

Relevant literature also reveals a number of examples of implemented learning data visualisation systems. For example, Santoso and colleagues (Santoso et al., 2018) describe the development of their comprehensive learning dashboard intended for lecturers, implemented as a plugin to the well-known Moodle LMS. Their work focused on developing a tool called SCELE that allowed course instructors to collect and visualise data so as to facilitate their observation and analysis of learning data. Data visualisation itself was carried out using various JS and CSS libraries.

Bodily and colleagues (Bodily at al., 2018) developed a dashboarding system that supports students during online learning. It consists of two separate dashboards for content and skill recommendation (recommendations made based on the gathered data about gaps in student knowledge and skills) that are changing in real-time.

Naranjo and colleagues (Naranjo et al., 2019) presented CloudTrail-Tracker, an educational dashboarding tool for course instructors that provides a visualisation of resource usage (for the selected time period) by students in a massive online open course, as well as specific activities carried out by a particular student. The system may also be used by the students (overview of learning progress) and system administrators (overview of resource consumption).

Han and colleagues (Han et al., 2020) developed a dashboard to support collaborative argumentation in face-to-face contexts. It consists of two dashboards: one intended for students and one for course instructors. The used collaboration system (Trello) recorded learning behaviours of the students (i.e., made activity logs) that were then transferred into a designated database using an API. Following data analysis, students and teachers received colour-coded feedback on the activities they were engaged in.

Gutiérrez and colleagues (Gutiérrez et al., 2020) report on LADA: a learning analytics dashboard intended for academic advisors in their decision-making, based on predictive analysis. The tool was implemented in the context of higher education to be used when needing to make difficult academic choices about students' studies.

2.2 Current State of the Developed System for Data Visualisation

The developed system for learning data visualisation stores data regarding course instructors, courses, students, and testing/evaluation activities. The number and type of evaluation activities may vary across courses, but this variety is addressed by the system. The type depends on the way the course is carried out (e.g., evaluation during the semester in combination with a final exam, evaluation during the semester without a final exam, workshop modality, etc.) as well as the evaluation strategy opted for by the course instructor.

If provided appropriate files containing the results of evaluation activities, the system offers users visualisations of such data (e.g., course instructors, the management, educational decision-makers, etc.) without the need for a system administrator to change the settings or make any further adjustments to the file and its contents. Such an automated data import functionality allows users to define their own structure of records in the table designated for importing data from the selected (exported) file. During this automated process, a set of metadata is generated on the imported results about evaluations so as the system is able to link the data on evaluation activities from the file with the tables for storing data on the evaluation results in the visualisation system. The automated import functionality is carried out using an original algorithm for automatic data import, which was made possible by extending the relational data model so as to accommodate new metadata (Kovačić, 2022). The structure of the system is given in Fig. 1.

The developed system for data visualisation enables the user to visualise the results of evaluation activities in a course that have been defined in advance (e.g., different types of activities, specific grading methods, or unique requirements for passing an activity) based on the imported data.

The data visualisation system consists of the following modules (Kovačić, 2022):

- Interface – an interactive communication channel between the system and the users.

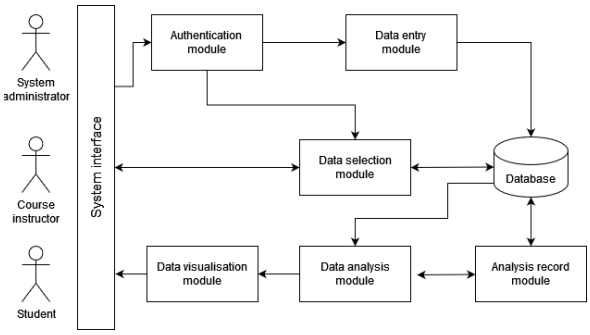


Figure 1. Structure of the system for data visualisation

- Authentication module – after entering a combination of username and password, the system authenticates the user, defines their role within the system (i.e., system administrator, course instructor or student), and authorizes the user to work with certain parts of the system.
- Data entry module – the system administrator enters data on courses, study groups, institutions, course instructors, and students[1] into the system, while the course instructor imports the data regarding the results of evaluation activities using an import file.
- Data selection module – course instructors and students can select a particular course and specific data regarding evaluation activities in order to create a customised visualisation (applying data-selection criteria).
- Data analysis module – provides a statistical analysis of the selected data regarding evaluation activities and prepares them for visualisation.
- Analysis record module – stores data on all data visualisations performed by the system in order to allow for a comparison between the current visualisation and previously generated ones.
- Data visualisation module – generates various types of diagrams to be presented as part of data visualisation.

After a successful login using their credentials, the user is authenticated and authorised to work with specific functionalities of the system. In the preparation step preceding user's interaction with the system, the administrator is responsible for entering or updating data on system users and courses so as to allow the users to work with only those courses assigned to them. Once data import regarding evaluation activities for a particular course has been completed, the course instructor selects specific data points that need to be included in the visualisation of data and analyses the results presented by the system. Records kept by the system allow the course instructor to compare and monitor students' learning progress and make appropriate teaching and learning interventions in the course. On the other hand, students are shown the results of their own learning endeavours, which may be compared to the (Kovačić, 2022) results of their fellow students, thus instigating comparison, encouraging competitiveness, and increasing motivation to perform better.

[1] Entering data on students is optional as the system allows for automatic data entry using the import file (at a later stage).

3 System Design and Implementation

In this section we provide an in-depth description of how the system for data visualisation is designed and implemented based on the previously provided description of its functionalities, with a special emphasis on the underlying database.

3.1 Technological Basis of the System for Data Visualisation

The technology stack used to implement the system for learning data visualisation includes the following elements: (1) MySQL relational database management system (RDMS), (2) Apache HTTP server, (3) Phalcon framework, and (4) Chart.js system for data visualisation.

The MySQL RDMS, Oracle's free and open-source software, is used to implement data storage for the data visualisation system. It is characterised by high reliability and ease of use, providing an environment for integrated development, design and administration of databases.

The Apache HTTP server was used to develop and manage an open-source HTTP server that hosts the application. The associated Apache project aims at providing a secure and extensible HTTP server in compliance with the modern HTTP standards (Apache HTTP Server Project, 2023).

The Phalcon framework was used in order to enter the data into the system, store it, select and prepare it for visualisation, and update the archive of performed analyses. Phalcon is an open-source full-stack PHP framework optimised so as to provide the lowest overhead in applications that use the MVC (model, view, controller) architectural pattern (Phalcon framework documentation, 2023). This particular architectural pattern separates the application into three main logical components in which a model represents data the system is working with as well as the rules governing its use, a view represents the user interface of the application, and a controller enables the flow of data between the models and associated views (The MVC architecture, 2023).

In the process of learning data visualisation, graphical representations of data are implemented using the Chart.js JavaScript library. This is a simple and flexible library that provides numerous chart types, plugins and customisation options, many of them available from and maintained by the community (Chart.js documentation, 2023).

3.2 System Implementation

3.2.1 Authentication Module

The main task of the authentication module is to store data regarding the users of the system, and to link user data to data about the courses (and other forms of organised educational units for which evaluation activities are carried out). The module enables the following activities:

- verifying user's identity by validating the username and password provided by the user,
- linking users to courses for which data visualisation is performed,
- linking courses with study groups and institutions.

Relational schema implementing the authentication module of the learning data visualisation system is given below, and the accompanying entity relationship diagram (ERD) is provided in Fig. 2.

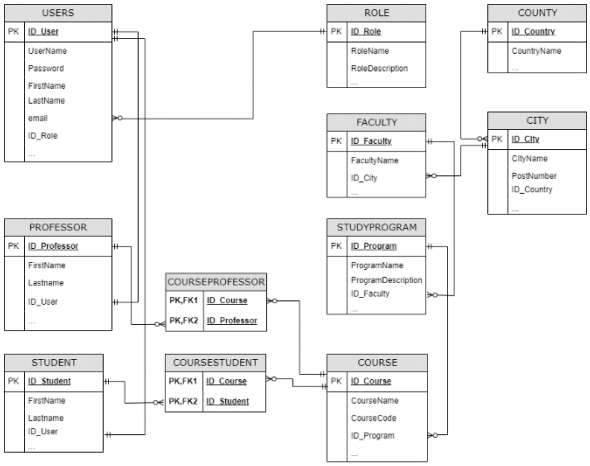


Figure 2. ERD of the system's authentication module

USERS(ID_User, UserName, Password, FirstName, LastName, email, ID_Role)
ROLE(ID_Role, RoleName, RoleDescription)
COURSE(ID_Course, CourseName, CourseCode, ID_Program)
STUDYPROGRAM(ID_Program, ProgramName, ProgramDescription, ID_Faculty)
FACULTY(ID_Faculty, FacultyName, ID_City)
CITY(ID_City, CityName, PostNumber, ID_Country)
COUNTRY(ID_Country, CountryName)
PROFESOR(ID_Professor, FirstName, LastName, ID_User)
STUDENT(ID_Student, FirstName, LastName, ID_User)
COURSEPROFESSOR(ID_Course, ID_Professor)
COURSESTUDENT(ID_Course, ID_Student)

Verification of user's identity is achieved using the USERS relation which contains the username, password, first and last names of the user, and the assigned user role in the system. System administrator enters data regarding course instructors into the USERS relation, which are then transferred into the PROFESSOR relation (course instructors are not allowed to change the data on their own). Finally, the administrator enters the data regarding the course assigned to a course instructor (relation COURSEPROFESSOR). There are two ways of entering data regarding the students:

1. the system administrator enters the data directly into the USERS relation, which are then transferred into the STUDENT relation and linked with the courses in the COURSESTUDENT relation.
2. using the automated process of importing data from the import file, during which data is entered automatically into the STUDENT and COURSESTUDENT relations. If the students are allowed access to the system for learning data visualization, their data is added to the USERS relation and the system generates and send them a password (which is to be changed upon their first access to the system).

The ROLE relation contains data identifying the roles system users may be assigned and the accompanying description of each role. The COURSE relation stores data such as course names, course identifiers, as well as relationships with the study programme (relations STUDYPROGRAM, FACULTY, CITY, COUNTRY).

3.2.2 Data Entry Module

The data entry module is responsible for storing the data regarding the results of evaluation activities within a course and linking the imported data with data on courses, course instructors, and students. The module enables the following activities:

- selecting courses and files for data import,
- linking data from the import file to the relation for storing data regarding the results of evaluation in the system for data visualisation,
- reviewing imported data.

Relational schema implementing the data entry module of the learning data visualisation system is given below.

COURSE(ID_Course, CourseName, CourseDescription)
PROFESSOR(ID_Professor, FirstName, LastName, …)

COURSEPROFESSOR(ID_Course, ID_Professor)
STUDENT(ID_Student, FirstName, LastName, …)
COURSESTUDENT(ID_Course, ID_Student)
GROUP(ID_Group, GroupName, GroupDescription, …)
STUDENTGROUP (ID_Course, ID_Student, ID_Group)
STRUCTURE (ID_Structure, StructureName, StructureDescription)
CATEGORY (ID_Category, CategoryName, CategoryDescription, ID_Structure)
COURSEDATA(ID_Results, ResultName, ResultDescription, ID_Course)
COURSESTRUCTURE(ID_Course, ID_Structure)
COURSEDATASTRUCTURE(ID_Course, ID_Structure)
COURSERESULTDATA (ID_CRD, ID_Course, ID_Profesor, ID_Result, ID_Structure, ID_Category, ID_Group, ColumnNumber, ColumnName, FileColumnName)
RESULTS (ID_Result, ID_Student, Column1, Column2, …., ColumnN)

Course instructors select one of the courses assigned to them and add new tuples into the relation for storing imported data (relation COURSEDATA). Only then are they able to specify the import file (its location on the user's computer). Data entry is facilitated by the automatic process which consists of two steps:

1. The course instructor selects the way the course is carried out (i.e., modality) and an arbitrary number of categories (activity which includes multiple data points from the import file, e.g., evaluation activity which is carried out in several groups of students). This is implemented by relations STRUCTURE, COURSEDATASTRUCTURE, and CATEGORY. Then the data header from the import file is read (table header regarding the results of evaluation activities). The course instructor then selects the data to be imported (table columns from the import file), decides on the modality in which the course is carried out, and an arbitrary category for the selected data point. After each selection is made, data is stored in the COURSERESULTDATA relation and a column from the RESULTS relation is chosen to store the selected data point from the import file.
2. Data is imported using the automated process, i.e. by reading the data on each student (tuple from the table in the import file) and transferring it to the RESULTS table (data point regarding the column of the RESULTS relation used to store the data from the file is defined by the COURSERESULTDATA relation).

After all the data have been entered into the system, selecting a group of students for which the analysis will be performed may be carried out in one of two ways:

1. The course instructor defines groups of students and links them with to data from the import file (during the first step of data import).
2. Using the automated process which takes each data point that has been previously included in the category and records only those students who have the same value as the selected data point. The student group is created subsequently (relations GROUP and STUDENTGROUP).

After importing the data, the course instructor may opt to review the modality in which the course is carried out, along with all the evaluation activities in the course that have been entered into the system using the import file and the overview of all the entered data (relations COURSERESULTDATA and STUDENTGROUP).

The ERD of the data entry and data analysis modules is provided in Fig. 3.

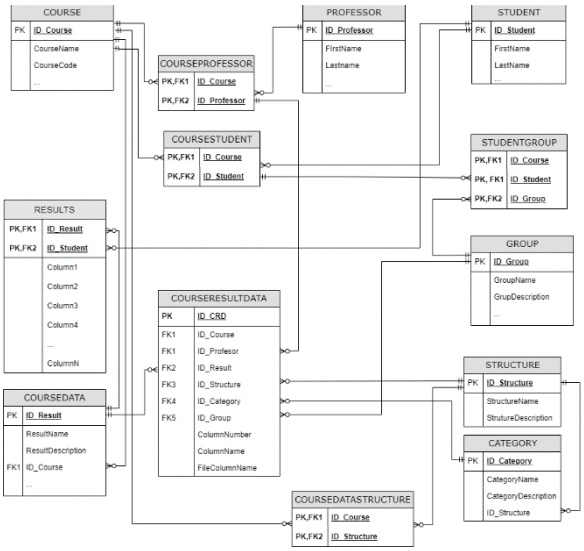


Figure 3. ERD of the system's data entry and data analysis modules

3.2.3 Data Analysis Module

The data analysis module is in charge of receiving the dataset that needs to be visualised by the system and performing statistical analyses using the data (e.g., maximum values, minimum values, averages, standard deviations, etc.). The analysed data is then rendered as a diagram of a particular type, depending on the characteristics of the data or the setting associated with a diagram type. The analysis is performed for each selected dataset that needs to be graphically displayed, and is entered into the ANALYSE relation.

System's data analysis module is based on the same relational data model as the data selection module (see Fig. 3). The main difference between the two is that the data selection module does not deal with the data from the ANALYSE relation, while the data analysis module performs statistical operations for the selected dataset and calculates statistical values in the ANALYSE relation.

3.2.4 Data Selection Module

The purpose of the data selection module is to display the data imported into the data visualisation system regarding evaluation activities, categories, and modalities in which the course is carried out.

It allows the course instructor or the student to select a (sub)set of data that needs to be visualised by the system, as well as the criteria for sorting and/or filtering this data. Relations ANALYSE, ANALYSESTRUCTURE, and ANALYSEGROUP are used to store the relevant data. In the last step, the system is able to select all the data about students from the RESULTS relation and prepare it in the format required by the diagram-generating subsystem (i.e., for data visualisation).

The ERD of the data selection module is provided in Fig. 4.

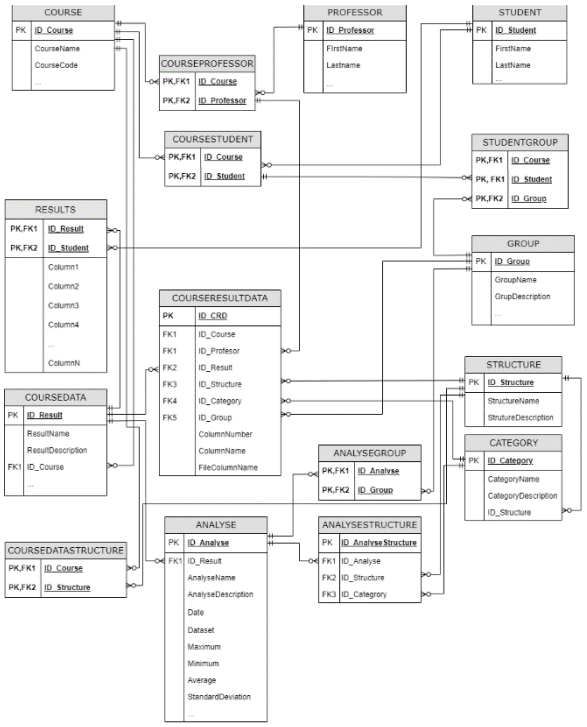


Figure 4. ERD of the system's data selection module

3.2.5 Data Visualisation Module

The data visualisation module displays diagrams in the process of data visualisation by using the data in the ANALYSEGRAPH, ANALYSEGRAPHELEMENT, GRAPHTYPE, and GRAPHELEMENT relations.

The GRAPHTYPE relation keeps records regarding the data about diagrams used for data visualisation. Records related to the selected diagram type are stored in the ANALYSEGRAPH relation. Values necessary for the diagram to be rendered on the screen (descriptive data such as diagram title, diagram description, axis names, etc.) are kept in the GRAPHELEMENT relation. Finally, additional user settings for the selected diagram type are stored in the ANALYSEGRAPHELEMENT relation.

All the data contained in the above described relations are entered into the relations only if the user wants to use them for comparison with any subsequent data visualisations. This allows the system to store both the dataset for data visualisation and configuration settings for diagram display.

3.2.6 Analysis Record Module

The purpose of the analysis record module is to store data on all previous data visualisations in order for the user to be able to compare them to the newly performed ones. This allows the users to detect changes, keep track of trends, and identify problems or difficulties in the learning and teaching setup. Furthermore, it allows the users of the system to make informed decisions concerning the observed issues and to improve the educational process as a whole.

The relational schema of the data analysis, data visualisation, and analysis record modules is given below.

ANALYSE(ID_Analyse, ID_Result, AnalyseName, AnalyseDescription, Date, Dataset, Maximum, Minimum, Average, StandardDeviation)
ANALYSESTRUCTURE(ID_AnalyseStructure, ID_Analyse, ID_Structure, ID_Category)
ANALYSEGROUP(ID_Analyse, ID_Group)
GRAPHTYPE(ID_Graph, GraphTypeName, GraphTypeDescription)
ANALYSEGRAPH(ID_Analyse, ID_Graph, JSON, Decription, SelectionCriteria)
GRAPHTYPELEMENT (ID_GraphType, ID_Element, ElementName, ElementDesc, DefaultValue)
ANALYSEGRAPHELEMENT(ID_AnalyseGraphElement, ID_Analyse, ID_GraphType, ID_Element, Value)

Both data visualisation module and analysis record module use the same relational data model. The main difference, however, is that the Data visualisation module writes data needed for current graph visualisation into the ANALYSEGRAPH relation, while the analysis record module uses previously generated data from the ANALYSEGRAPH relation to visualise data in previous data analyses runs.

The ERD of the data visualisation module is provided in Fig. 5.

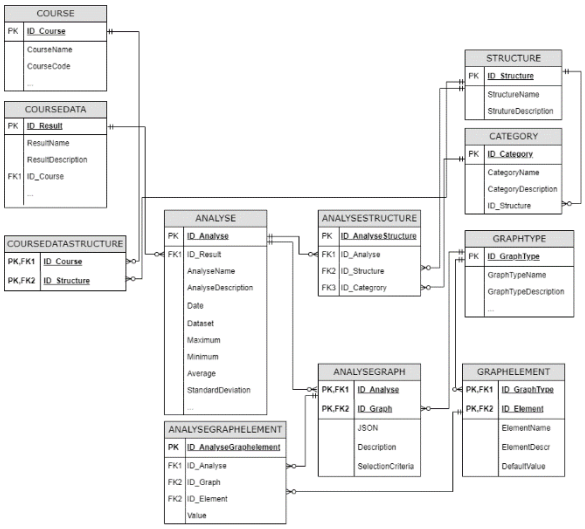
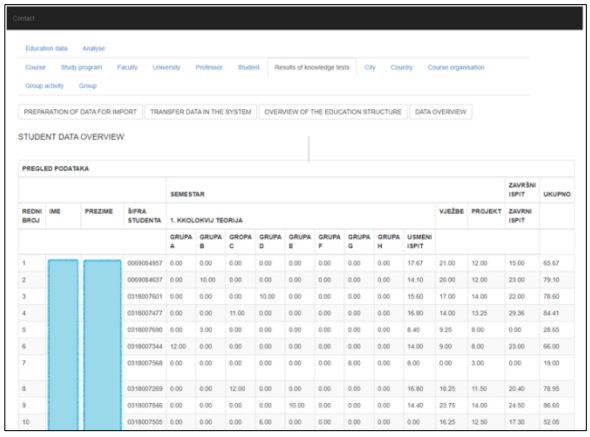
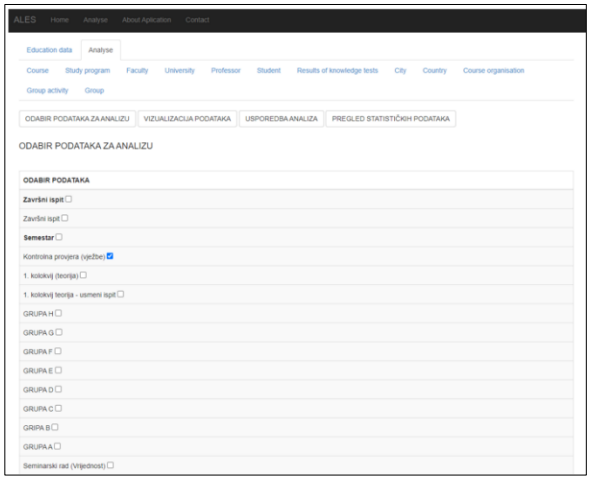


Figure 5. ERD of the system's data visualisation module

Fig. 6 offers an example of overview of student learning data (after import), data selection, and data visualisation as a series of screenshots taken directly from the described data visualisation system.





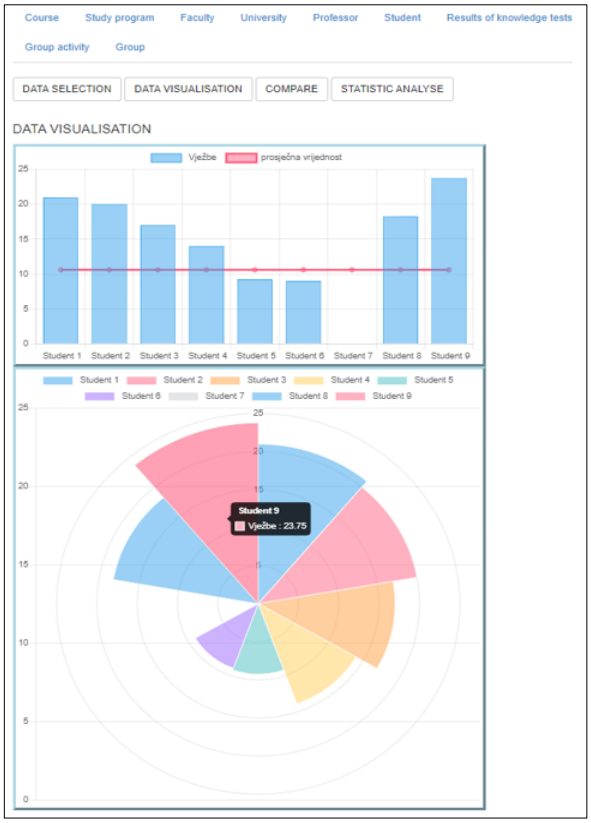


Figure 6. Screenshots from the system

4 Conclusion

This paper focused on the description of the system for learning data visualisation developed at our institution. A system that enables data analysis and its subsequent graphical representation, may aid all the stakeholders (i.e., course instructors, students, decision-makers, etc.) in making well-informed and effective decisions that improve both teaching and learning practices. The developed system allows learning analytics experts and non-experts to perform analyses on student generated data (evaluation activities) by importing the data and automatically preparing it (metadata creation) for further analysis.

Special emphasis was given to the implementation of the system's back-end and the design of its database, which enables the various functionalities of the system and its six interconnected modules.

Future work on the system will be aimed towards improving the results of analysis of data selected for visualisation, so as to enable the discovery of new trends in knowledge acquisition and improve decision-making during the teaching and learning process.

Acknowledgments

This work was supported by the University of Rijeka projects uniri-drustv-18-73 and uniri-drustv-18-224.

References

Apache HTTP Server Project. (2023). Retrieved from https://httpd.apache.org/

Black, E. W., Dawson, K., & Priem, J. (2008). Data for free: Using LMS activity logs to measure community in online courses. *The Internet and Higher Education*, 11 (2), 65–70. doi: https://doi.org/10.1016/j.iheduc.2008.03.002

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education*, 30, 572–598. doi: https://doi.org/10.1007/s12528-018-9186-0

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *PNAS*, 116(6), 1857–1864. doi: https://doi.org/10.1073/pnas.180718011

Chart.js documentation. (2023). Retrieved from https://www.chartjs.org/docs/latest/

Donohoe, D., & Costello, E. (2020). Data Visualisation Literacy in Higher Education: An Exploratory Study of Understanding of a Learning Dashboard Tool. *International Journal of Emerging Technologies in Learning (iJET)*, 15(17), 115–126. doi: https://doi.org/10.3991/ijet.v15i17.15041

Gutiérrez, F. et al. (2020). LADA: A learning analytics dashboard for academic advising. *Computers in Human Behavior*, 107, article 105826. doi: https://doi.org/10.1016/j.chb.2018.12.004

Han, J., Kim, K. H., Rhee, W., & Cho, Y. H. (2020). Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Computers & Education*, 163, article 104041. doi: https://doi.org/10.1016/j.compedu.2020.104041

Kovačić, B. (2022). Extension of e-learning data visualization system for different knowledge reviews datasets. In L. Gómez Chova, A. López Martínez, & J. Lees (Eds.), ICERI2022 Proceedings (pp. 7396–7403). IATED Academy, Valencia. doi: 10.21125/iceri.2022.1882

Lang, C. et al. (2022). What is learning analytics? In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), *Handbook of Learning Analytics – Second edition* (pp. 8-18). SoLAR, Vancouver. doi: 10.18608/hla22.001

Lowe, J., & Matthee, M. (2020). Requirements of Data Visualisation Tools to Analyse Big Data: A Structured Literature Review. *Responsible Design, Implementation and Use of Information and Communication Technology*, 2020 Mar 6; 12066, 469–480. doi: 10.1007/978-3-030-44999-5_39

Mavrikis, M., Geraniou, E., Gutierrez-Santos, S., & Poulovassilis, A. (2019). Intelligent analysis and data visualisation for teacher assistance tools: the case of exploratory learning. *British Journal of Educational Technology*, 50 (6), 2920–2942. doi: https://doi.org/10.1111/bjet.12876

Naranjo, D. M., Prieto, J. R., Moltó, G., & Calatrava, A. (2019). A Visual Dashboard to Track Learning Analytics for Educational Cloud Computing. *Sensors*, 19(13), article 2952.https://doi.org/10.3390/s19132952

Phalcon framework documentation. (2023). Retrieved from https://docs.phalcon.io/5.0/en/introduction

Romero, C., & Ventura, C. (2019). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. doi: 10.1002/widm.1355

Santoso, H. B., Batuparan, A. K., Isal, R. Y. K., & Goodridge, W. H. (2018). The development of a learning dashboard for lecturers: A case study on a student-centered e-learning environment. *Engineering Education Faculty Publications*, Paper 236.

Sedrakyan, G., Mannens, E., & Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, 50, 19–38. doi: https://doi.org/10.1016/j.jvlc.2018.11.002

The MVC architecture. (2023). Retrieved from https://docs.phalcon.io/3.4/en/mvc

Digital Human

Tihomir Katulić, Anamarija Musa, Darja Lončar
Understanding some of the open data challenges to data protection in the developing European legal framework

Marko Bohanec
Inter- and Intra-Personal Differences, and Consistency of Decision Rules, in Multi-Criteria Modelling Method DEX: A Preliminary Study

Kaja Prislan Mihelič, Igor Bernik, Simon Vrhovec, Anže Mihelič
Perceived Threat of Cyber Attacks and its Role in the Adoption of Tablet Computers by Older Adults

Lucija Gosak, Jeroen Ooge, Nino Fijačko, Janja Kamenšek, Primož Kocbek, Nika Debeljak, Katrien Verbert , Gregor Štiglic
Self-Care Oriented Smartphone Apps for Type 2 Diabetes: A Comparative Analysis

Katarina Tomičić-Pupek, Martina Tomičić Furjan, Igor Pihir
Digital Human Resource Management - Designing a digital platform for educational institutions' HR process

Marija Đukić, Milica Škembarević, Olga Jejić, Ana Pajić Simović
Toward the utilization of chatbots in the banking sector

Darko Babić, Mario Fiolić, Dario Babić
Inspection of the road surface condition on black spots – Croatia case study

Understanding some of the open data challenges to data protection in the developing European legal framework

Tihomir Katulić, Anamarija Musa
University of Zagreb
Faculty of Law
Trg Republike Hrvatske 14, Zagreb
{tihomir.katulic, anamarija.musa}@pravo.unizg.hr

Darja Lončar
Rimac Group
Zagreb
darja.loncar@rimac-automobili.com

Abstract. *According to the EU Charter of Fundamental Rights, both the right to privacy and the right to the protection of personal information are different, emancipated rights that are complementary to one another. The high level of data protection was further improved by the General Data Protection Regulation. Open data is information that may be used for commercial or non-commercial purposes and is made accessible to the public in an open and machine-readable manner. It is anticipated that open data would increase public sector openness while also fostering the (data) economy and data-driven innovations, particularly with regard to the IT services sector and SMEs. The Open Data Directive has established rules for the release of open data and the re-use of public sector information in the EU. The new European data strategy from 2020 has underlined the necessity for open data, even outside of the data owned by the public sector. According to this strategy, the EU's single market for data will be strengthened by using more open data. Consequently, the EU has adopted the new Data Governance Act as a cross-sectoral instrument that tries to increase data accessibility by regulating the reuse of protected data held by public sector, promoting the sharing of data for altruistic purposes and regulating data intermediaries as a novel approach to fostering open data economy.*[1]

Keywords. Open data, data protection, DGA, GDPR

1 Introduction

This paper contains a brief overview of the respective EU data protection and open data legislation, i.e., the notions of data protection and open data as understood in the EU law. The central part of the paper will deal with the intersection between data protection and open data by analyzing selected aspects which seem to be especially relevant for balancing between data protection and open data.

Legislative activities of the European Union (EU) have long ago established a global standard of data protection and privacy legal framework. Both the right to protection of personal data and right to privacy are distinctive, emancipated (Gonzalez Fuster, 2014) yet supplemental and parallel fundamental rights of individuals as regulated by the EU Charter of Fundamental Rights of the European Union. The Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation; GDPR) which became applicable in 2018 has additionally raised this level of data protection adopting a number of mechanisms aimed at ensuring an adequate level of compliance in relation to perceived risk to potential data breaches from the perspective of data controllers and processors (Katulić, Vojković 2016), although even then there were opinions that some of its finally agreed upon provisions, like the layered approach to data breach reporting, have somewhat blunted the efficacy in general prevention (Papakonstantinou, De Hert 2016).

The Regulation has also explicitly regulated the data subject data protection rights, including the data portability right and laying out the rules for restriction of these rights while respecting the essence of the fundamental rights and freedoms when necessary (Art. 12 to 23 of the GDPR).

2 Open Data in the EU legal framework

For the past twenty years public authorities have been increasingly pressured to open their data for public use,

[1] The paper is a result of the TODO project based on received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 857592.

from the users' communities (NGOs, SMEs, etc.) but also from the highest political levels (e.g., endorsement of the Open Data Charter by G8, OGP etc.).

Open data is data which is available to the public in the machine-readable format for commercial or non-commercial (re)use (Van Loenen, Zevenbergen, de Jong, 2012). It is expected that open data will improve transparency of the public sector but also help boost the (data) economy, data-driven innovations, especially with respect to IT services industry and SMEs (Wessels 2012). In the EU, the issue of open data and the re-use of public sector information has been regulated by the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (Open Data Directive).

The new European strategy for data from 2020 has additionally emphasized the need for open data, even beyond the data held by public sector. Within this strategy the open data is perceived extremely important for strengthening the EU single market for data. The new Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act or DGA) complements the Open Data Directive but also goes beyond it. The Data Governance Act was proposed by the European Commission in 2020 and adopted by the European Parliament in 2022. The Act (EU) 2022/868 will apply from September 2023.

It extends the scope of openness of data held by public sector by imposing obligation of openness, under certain conditions, also to data which are subject to rights of others, including on the grounds of protection of personal data. The second important piece of legislation within the European strategy for data is the Proposal for a regulation of the European Parliament and of the Council on harmonized rules on fair access to and use of data (Data Act). The Proposal of Data Act extends the obligation of data openness to data held by private sector as well, thus substantially changing the paradigm of open data in EU.

The GDPR, as its full name suggests, is not only about protecting personal data but also about the free flow of data in the EU market. Similarly, Open Data Directive, Data Governance Act and Proposal of Data Act emphasize the necessity to comply with the GDPR and respect the right to protection of personal data. Nevertheless, it seems that data protection and open data are on the opposite sides of the openness/data protection legal debate. The question put forward is to what extent these opposing concepts can be balanced. What are the challenges to protection of personal data from the open data legislation, in particular as of the new European strategy of data, and vice versa? Is it possible to achieve adequate openness of data as required by the EU open data legislation, while at the same time preserving the right to protection of personal data as required by the EU data protection standards?

The open data and personal data protection come into conflict most often in cases when the users of open data, such as civil society organizations combating tax fraud and corruption or business organizations creating new content based on different databases, require datasets containing personal data (names) of politicians and civil servants or receivers of public funding such as subsidies, or demand full access to company registers in the open data format and similar. In these cases two issues have to be considered.

Firstly, whether the personal data is public information which could be accessible to public in any form. The question relates to the traditional balancing between transparency for the sake of accountability and efficiency of public organizations, and personal data protection with the purpose to protect privacy. The second issue concerns the measure to which this information, if public at all, should be made available in open data format to be further processed and disseminated outside the public bodies and their public functions for the purpose for which it is not initially collected. Usually, these issues are discussed and thought through in each particular case, depending on the wider context in which data is requested.

Any type of personal data processing, including re-use of personal data within the concept of open data, must rely on at least one of the legal bases from Article 6 GDPR (and Article 9 in case of special categories of personal data) to be lawful. When processing is based on (national) law imposing obligation to open data or public interest (Article 6.1.c) and e)), the certain quality of law is required. To what extant are these GDPR requirements of lawfulness recognized by the EU open data legislation? The Data Governance Act also introduces the concept of “data altruism” where individuals can consent to re-use of their personal data for altruistic purposes (e.g., in health, mobility). To what extent is consent approach from the Data Governance Act similar to the one from the GDPR? Also, we shall tackle to what extent are data protection principles affected with the open data concept.

Another line of argument and analysis concerns the advent of machine learning technologies, colloquially referred to as AI (Artificial Intelligence). The development and use of these technologies relies heavily on the use of personal data, and open data is seen as a supporting and contributing effort to further development of these critical systems for future information society products and services (Grafenstein, 2022).

At the same time, as the legal foundation of the European system of human rights, the Charter of Fundamental Rights incorporates substantial checks and constraints on the effect and aim of future EU AI regulation. Whenever and however this law is enacted, it must comply with and contain existing European legal standards pertaining to persons' fundamental rights in the EU. The European Commission's ethical guidelines and the opinions of data protection authorities and institutions such as the EDPS and the

EDPB lay out ethical standards based on recognized fundamental rights that future AI systems must follow in order to be trusted and secure from the standpoint of enacted principles of personal data protection.

The interplay between the data protection and open data, especially as regulated within the new European strategy for data, is very new. Against this background and in conclusion, the authors would like to identify problems and raise questions which seem to us especially relevant in balancing data protection and open data, acknowledging the scope of the problem and inviting further study and research.

3 Digital Governance Act

After several years in development, the European Commission released its new open data strategy document, the European Data Strategy in 2020, in which it detailed its vision for the building of a European data economy over the next decade.

Main goal of the strategy is to promote use of available data in all economic sectors as this is vital to the development of digital economy. Currently there are plenty of arguments why the EU market has not been able to leverage the available data in order to develop new information society products and services and previous research has identified numerous factors which have had a stifling influence on the development of information economy, such as general lack of data availability, market power imbalances, insufficient governance structures and technical infrastructure or the lack of adequate tools to enable data subjects (consumers, users of services, general public) to efficiently exercise their rights related to data sharing, such as the GDPR Article 20 Data Portability Right (Gellert, Graef 2021).

The Open Data Directive governs the re-use of publicly available information maintained by the public sector. The public sector, on the other hand, maintains huge volumes of protected data (e.g., personal data and commercially secret data) that cannot be re-used as open data but could be re-used under special EU or national legislation. A wealth of knowledge can be derived from such data without jeopardizing its protected status, and the Data Governance Act (DGA) includes guidelines and protections to facilitate such re-use if permitted by other laws.

The DGA is an important part of a the wider legislative effort the EU has undertaken to create rules for future digitalization, data economy, artificial intelligence, and other major policy goals sometimes referred to as digital sovereignty.

Because artificial intelligence research is reliant on having access to enormous pools of data, data is a critical component in this framework. Similarly, virtually all new information society services, digital goods, and apps rely on data availability for at least some of their performance. Lastly, the availability of enormous volumes of data is becoming increasingly important in scientific study in general.

The DGA's objectives are substantial. The major purpose is to promote the European data economy and to strengthen the EU's digital single market as it tries to play catch-up with United States and SE Asia (Voss 2020), birthplaces of most of today's innovative services and platforms. With this in mind, and with hope to empower the entrepreneurial culture of the European market, small and medium-sized firms (SMEs) and start-ups are given special attention. These entities especially benefit from planned data reuse and data sharing which provides new material for innovation in artificial intelligence and digital applications (Ruohonen, Mickelsson 2023).

Scientific research is also an essential component of the goals of the Digital Governance Act. Data is regarded as essential for combating a range of current economic and societal issues, also helping with adapting to and preventing climate change, supporting the green transition, as well as enhancing energy infrastructure, healthcare, and financial services. These objectives are to be met in specific "European style" of approaching data and the data economy, following that any proposed policy should ensure a fair and legal framework for reuse of data, especially from the perspective of data protection principles (Ruohonen, Mickelsson 2023).

The DGA itself is divided into eight chapters, from introductory provisions and definitions of key concepts and terms to provisions regarding the re-use of certain categories of protected data held by public sector bodies, requirements to data sharing services, data altruism, oversight of competent authorities over the application of the Regulation and establishment of the European Data Innovation Board.

4 From GDPR to DGA

The DGA defines the concept of data altruism, requiring Member States of the EU to adopt a horizontal regulation for the re-use of specific types of protected data stored by public sector organizations, the provision of data intermediation services and data altruistic services.

Data altruism is defined in the Article 2 of the Regulation as "...voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward that goes beyond compensation related to the costs that they incur where they make their data available for objectives of general interest as provided for in national law".

These provisions, according to Article 2 of the DGA, would relate to sectors such as healthcare, fighting climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the

provision of public services, public policy making or scientific research purposes in the general interest (Article 2 of the DGA).

Organizations that make relevant data available based on data altruism will be allowed to register as "data altruism organizations recognized in the Union" (DGA Art. 19 and 20). These institutions must be non-profit, fulfill transparency criteria, and provide particular protections to protect the rights and interests of individuals and businesses who share their data. Moreover, they must comply with the rulebook (prepared by the European Commission – the Commission will draft the rulebook in close collaboration with data altruistic organizations and other relevant parties at least 18 months after it becomes effective), which will include information requirements, technical and security requirements, communication roadmaps, and suggestions on interoperability standards (Recitals 46, 58 and Article 22 of the DGA). The organizations will be allowed to use the common symbol created for this purpose, and they will have the option of being included in the public register of data-altruism organizations. The Commission will create an EU-wide register of recognized data/altruism organizations for information purposes.

While referring to concepts and definitions already defined in the General Data Protection Regulation, such as personal data, data subjects, consent, and data processing (Article 4 of the GDPR), DGA also defines or updates definitions of specific terms and concepts such as re-use of data, data holders, data users, data sharing and data intermediation services in a way resembling the similar provisions in the GDPR (Papakonstantinou, De Hert 2021). As Papakonstantinou and de Hert already observed in 2021, the DGA constitutes a system remarkably similar to the one established by GDPR building on the concepts dating into the 1990s and Data Protection Directive, or even earlier to CoE Convention 108.

The DGA defines re-use as use by natural or legal persons of data held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data between public sector bodies purely in pursuit of their public tasks. Data holders are defined as legal persons - including public sector bodies and international organizations - or a natural person who is not a data subject with respect to the specific data in question, which have the right to grant access to or to share certain personal data or non-personal data in line with the applicable EU or Member State law. Data users are understood as natural or legal persons who have lawful access to personal or non-personal data and have the right to use that data for commercial or non-commercial purposes in accordance with applicable law (in case of personal data, the GDPR).

The DGA further develops the concept of data sharing as provisioning of data by a data subject or a data holder to a data user for the purpose of the joint or individual use of such data. The provisioning is based on voluntary agreements or Union or national law and can be facilitated directly or indirectly through an intermediary. The data might be available free of charge or be available under various open or commercial licenses. Many businesses are concerned that sharing their data would result in a loss of competitive advantage as well as the danger of potential abuse and litigation. The DGA establishes a set of rules for providers of data intermediation services (such as data markets) to follow in order to serve as trustworthy organizers of data sharing or pooling inside the shared European data spaces. The framework offers a new approach to the data-handling methods of Big Tech platforms, which have significant market influence due to their ownership over massive volumes of data.

In practice, data intermediates will serve as impartial third parties connecting individuals and data holders on one side with data users on the other. They will be unable to commercialize the data by selling it to another business or developing their own product based on it. Intermediaries will be subject to stringent standards in order to maintain their impartiality and avoid conflicts of interest. In reality, this implies that the data intermediation service and any additional services supplied must be structurally isolated (i.e. they must be legally separated). This innovative approach suggests a strategy based on the neutrality and openness of data intermediaries, while placing consumers and businesses in control of their data, in order to promote confidence in data sharing.

The DGA defines data intermediary services as services that establish commercial relationships between an undetermined number of data subjects and data holders one hand, and similarly an undetermined number of data user on the other hand. Intermediation services as defined by the DGA do not include services such as those that focus on intermediation of copyright protected content, services that are offered by public sector bodies that do not aim to establish commercial relationships, services exclusively used by one data holder or used by multiple legal persons in a closed group or services that obtain data from data holders and aggregate, enrich or transform the data for the purpose of adding substantial value to it and license the use of such data to data users (Article 2.11).

The similarities to the GDPR do not end with definitions of key concepts. DGA introduces principles of provisioning data sharing services in Article 11 such as transparency (Article 11.4 – 11.6), non-discrimination (11.6) and fair competition (11.6), special rights for individuals that want to participate in data altruism, such as regulation of the European data altruism consent form (Article 25), duty of data altruism organizations to inform data subjects on information pertinent to processing, such as the objectives of general interest and, if applicable, the specified, explicit and legitimate purpose for which

personal data is to be processed, and for which it permits the processing of their data by a data user (Article 21.1.a) etc. DGA also establishes a new permanent institution, the European Data Innovation Board (Article 29).

Article 3 of the DGA defines the types of data that can be reused. DGA applies to data kept by public sector organizations that are protected by commercial secrecy, statistical confidentiality, intellectual property protection, and personal data protection. As a result, personal data kept by public sector organizations is covered, and thus the GDPR also applies. This provision also recognizes certain exceptions, excluding "…data held by public undertakings, data held by public service broadcasters and their subsidiaries, data held by cultural establishments and educational institutions, data protected on the basis of national security and defense, and data falling outside the scope of the public sector bodies" public tasks.

Article 5 defines the requirements for data reuse, under the principles of non-discrimination, transparency, proportionality, and proper justification without efforts to limit competition. To guarantee data security, public sector organizations must ensure that personal data is anonymized, and commercially sensitive material is appropriately updated, aggregated, or otherwise managed with suitable disclosure restrictions.

The start of the application of the DGA will demand certain resources and institutional development in many of the Member States to reach the level of ability to adequately comply with the provisions of the GDPR. In this regard, the Commission foresees a need for Member States to invest in equipping their institutions to meet the requirements, especially coming from data protection law, in situations where data is being re-used. The measures may include developing a range of tools and technical solutions as well as adequate contractual provisions. If a public sector organization cannot allow access to specific data for re-use, it should help the potential user of data in obtaining the individual's agreement to re-use their personal data or the approval of the data holder whose rights or interests may be impacted by the re-use.

To make more publicly owned data available for re-use, the DGA restricts its usage of exclusive data re-use agreements (in which a public sector organization offers such an exclusive license to one enterprise) to certain circumstances of public interest. Also, while public sector organizations may charge fees for permitting re-use, they need to make sure that such fees do not exceed the necessary expenditures. Moreover, public sector organizations should reduce or even eliminate fees to encourage re-use for scientific research and other non-commercial objectives, as well as by SMEs and start-ups.

5 Competences and Role of European Data Innovation Board

In order to foster better implementation of the new data governance rules, the Article 29 of the DGA establishes the European Data Innovation Board (EDIB). The purpose of EDIB, established primarily as an expert group, will be to assist the European Commission in coordinating the practices and policies related to open data and supporting the cross-sector data use.

EDIB will consist of representatives of the competent authorities for data intermediation services and the competent authorities for registering altruism organizations. It will also include the EDPB, EDPS, ENISA and representation form other relevant institutions and organizations with specific expertise. According to Article 29, upon establishment the Board will divided into several subgroups, including one for stakeholder involvement, which should include representatives from industries such as health, environment, agriculture, transportation, energy, industrial manufacturing, media, cultural and creative sectors, and statistics, as well as research, academia, civil society, standardization organizations, relevant common European data spaces, and other relevant stakeholders and third parties. Even though the name of this working group resembles the European Data Protection Board, as Papakonstantinou and de Hert observe, its competences and role are not comparable. Where the EDPB has distinct powers to issue opinions, guidelines, recommendations and practices, it also monitors and ensures correct application of certain aspects of the GDPR (Article 70.1) whereas the EDIB is confined to an advisory role, advising, assisting and proposing the guidelines – much closer to the mandate of the former DPD Article 29 Working Party.

6 Discussion

While its always important to emphasize that the protection of personal data as a fundamental right in the EU is one of the more prominent achievements of the EU legislative development, development of the digital economy is also a necessity in the transformational process of building a postindustrial, information society. Availability of data for use should be followed by the general rise of trust in data intermediaries data-sharing mechanisms across the EU.

The regulation in question is increasingly divergent - the GDPR places a strong emphasis on safeguarding personal information and the rights of data subjects. On the other side, the DGA and Data Act are more concerned with promoting data exchange and developing a market for data. The competing goals might result in circumstances where data sharing prevails over data protection, a situation where data

protection authorities may feel a need to renew efforts in understanding these conflicting issues, perhaps by focusing more on promoting and participating in risk assessment, a role understood even in data protection systems far removed from EU and the GDPR (Mantelero 2016).

An important question at this time is whether the provisions of the DGA will contribute to this need to increase trust, or not. Certain provisions of the DGA, such as the definitions of data holders, data users, data sharing, seem to be in conflict or create confusion in application with regard to principals and other provisions of the GDPR (EDPB-EDPS Joint Opinion 03/2021). The joint opinion also mentions the definition of metadata as problematic, as it may imply processing of data that still qualifies as personal data.

More substantially, there appears to be a lack of clarity about a necessary legal basis for processing, where provisions of DGA, especially Article 5.6, 7.2.c, 11.11 and 19.3 refer to permission as something other than the regulated legal basis for processing of personal data under Article 6.1. of the GDPR. Same goes for instances of DGA provisions which seemingly may refer to other than legal basis prescribed by the Article 6 of the GDPR – however, the wording of applicable DGA provisions does not support the notion for DGA to be establishing a new, separate legal basis for processing nor does it in any case meet criteria for such situation set forth in Article 6.3 of the GDPR. According to that provision, such DGA provisions would have to specify "..inter alia: the general conditions governing the lawfulness of processing by the controller; the types of data which are subject to the processing; the data subjects concerned; the entities to, and the purposes for which, the personal data may be disclosed; the purpose limitation; storage periods; and processing operations and processing procedures etc."(Article 6.3.1 of the GDPR).

The EDPB and EDPS have also expressed concern on the lack of separation/distinction between personal data processing and processing of non-personal data in the provisions of the DGA, probably inspired by the provisions of the Regulation (EU) 2018/1807 of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, understanding how regulation of such processing is based and inspired on a different and often incompatible set of principles compared to processing of personal data. This reinforces the observation that such DGA provisions may create a regime of uncertainty and undermine the goals of the GDPR and the system of data protection in EU in general.

7 Conclusion

The EU has adopted the Data Governance Act (DGA) and is developing the Data Act proposal in an effort to foster the sharing of data across industries and further develop European common digital market for data. While the purpose of these activities is to strengthen the European digital economy, there is always risk that new legislation might endanger the data protection requirements set by the General Data Protection Regulation. The DGA and the Data Act proposal are part of the efforts to create a single market for data and promote data sharing across different sectors and although these regulations have been developed with GDPR standards in mind, they contain controversial provisions which might create conflicting situations or cause data sharing to take precedence over data protection, an established fundamental right in the EU.

The concept of intermediaries as introduced by the DGA as a neutral third parties facilitating data sharing between data holders and users may add another layer of complexity to the data sharing process and increase the risk of personal data misuse or data breaches, regardless of the GDPR obligations which intermediaries will have to operate under. Experience with data protection practice in the first five years of the GDPR application has repeatedly shown that complexity and number of stakeholders in a processing operation increase the chance of data breaches, while at the same time increasing administrative and technical load on organization compliance structures.

Another encouraged concept by the DGA, data altruism – sharing data for public interest persons – is also a potential source of risk as there are no adequate guarantees that would ensure that the data subjects are aware of the implications of sharing their data which in turn may prevent users from efficiently exercising their rights under the data protection law. As both DGA and Data Act in substantial part rely on data subject consent, experience in practical application of the GDPR shows how obtaining and maintaining consent may be a challenging endeavor.

Achieving an adequate balance between data sharing and data protection is essential in the EU legislation for reasons of economic growth, better protection of fundamental rights while achieving goals of public interest and other social benefits, furthering the state of legal harmonization and consistency among Member States and increasing competitiveness. Data sharing encourages innovation and subsequently economic growth through easier access to valuable information which in turns allows for more innovative research and development of new products ad services. On the other hand the EU has a goal of creating a consistent legal framework which will efficiently promote and protect recognized fundamental rights, among them privacy and data protection as distinctive rights of individuals in the EU. Any legislative intervention needs to contribute to better trust among all relevant stakeholders – citizens, businesses and governments. Ensuring adequate protection and enforcement of data subject rights can only help create an environment where individuals are more likely to share their data.

References

De Hert, P., Papakonstantinou, V: “The new General Data Protection Regulation: Still a sound system for the protection of individuals?”, Computer Law & Security Review, 32(2), 179-194, 2016.

Papakonstantinou, V., De Hert, P.:“Post GDPR EU laws and their GDPR mimesis. DGA, DSA, DMA and the EU regulation of AI“, European Law Blog: News and Comments on EU Law, europeanlawblog.eu, April 2021.

Mantelero, A.:”Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection”, Computer Law & Security Review, 32(2), p. 238-255, 2016.

Gonzalez Fuster, G:”The Emergence of Personal Data Protection as a Fundamental Right of the EU”, Springer, 2014.

Katulić, T., Vojković, G.:“From Safe Harbour to European data protection reform“, 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1447-1451, Opatija, 2016.

Ruohonen, J., Mickelsson, S.:"Reflections on the Data Governance Act", arXiv preprint arXiv:2302.09944, arXiv 2023.

Grafenstein, M.:"Reconciling Conflicting Interests in Data through Data Governance. An Analytical Framework (and a Brief Discussion of the Data Governance Act Draft, the Data Act Draft, the AI Regulation Draft, as well as the GDPR)", HIIG discussion Paper Series No.2022-02

Gellert, R., Graef, I:"The European Commission’s proposed Data Governance Act: some initial reflections on the increasingly complex EU regulatory puzzle of stimulating data sharing", TILEC Discussion Paper, ISSN 1572-4042, March 2021, Tilburg University

Van Loenen, B., Zevenbergen, J., de Jong, J. “Balancing open data and privacy in the design of the land administration domain model”, ISPRS International Journal of Geo-Information, 6(5), 137, 2012.

Voss, W. G.: "Cross-Border Data Flows, the GDPR, and Data Governance Cross-Border Data Flows, the GDPR, and Data Governance", Washington International Law Journal, 29 Wash. Int’l L.J. 485 (2020).

Wessels, B.:”Identification of the uses of open data for the development of new businesses”, Telecommunications Policy, 36(11), 997-1007., 2012.

EDPB-EDPS Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act), June 2021, available at: https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-032021-proposal_en

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and amending Regulation (EU) 2018/1724 (Data Governance Act), L 152/1, 3.6.2022.

The Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation; GDPR

Proposal for a Regulation of the European Parliament and of the Council on Harmonised rules on Fair Access to and Use of Data (Data Act), Com(2022) 68 Final, 2022/0047(Cod)

Inter- and Intra-Personal Differences, and Consistency of Decision Rules, in Multi-Criteria Modelling Method DEX: A Preliminary Study

Marko Bohanec
Jožef Stefan Institute, Department of Knowledge Technologies
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
and
University of Nova Gorica, Glavni trg 8, SI-5271 Vipava, Slovenia
marko.bohanec@ijs.si

***Abstract**. DEX (Decision EXpert) is a qualitative multi-criteria decision-modelling method in which decision alternatives are evaluated according to decision rules, elicited from individual decision makers. In this preliminary study, we assessed the differences between decision rules acquired from different subjects (inter-personal differences) and from the same subjects at different times (intra-personal differences). We also assessed the consistency of so-acquired rules and the ability of subjects to estimate the importance (weights) of criteria. The results indicate a high variability of decision rules, both inter- and intra-personal. Intra-personal drift is lower than inter-personal differences, but not substantially. The consistency of rules varied between a small decision table with clearly ordered criteria, where it was almost perfect, and a large decision table with less apparent preferential relations, where it was rather poor at the average level of 0.77. Criteria weighs also drifted at the rate 9–19% per month.*

Keywords. Multi-criteria decision modelling, method DEX, decision rules, weights, consistency, drift

1 Introduction

Multi-criteria decision modelling (MCDM) is a decision-making technique that involves the use of models to evaluate a set of decision alternatives based on multiple criteria or objectives (Greco, et al., 2016; Thakkar, 2021). MCDM is used in situations where the decision maker needs to balance the trade-offs between multiple, possibly conflicting criteria. MCDM typically involves three steps: (1) defining the decision problem and criteria, (2) identifying and evaluating the alternatives, and (3) synthesizing the results to make a decision. There are many MCDM methods that differ in how they represent criteria, evaluation/aggregation rules and alternatives, and how they acquire this information, which is often subjective, from decision makers. MCDM methods are typically named using acronyms, such as WSM, AHP, ANP, MAUT/MAVT, TOPSIS, VIKOR, MACBETH, PAPRIKA, PROMETHEE, ELECTRE, UTA, DRSA, DEX; see Greco, et al., (2016), Thakkar (2021) and Kulkarni (2022) for overviews and more information.

In this study, we are particularly interested in aggregation/evaluation aspects of multi-criteria models. In order to evaluate alternatives, a vast majority of MCDM methods employ the weighted sum

$$f(x_1, x_2, \dots, x_n) = w_1 x_1 + \cdots + x_n x_n$$

Here, x_i and w_i denote numerical criteria and their weights, respectively. The larger the weight, the more influential the criterion. Weights $w_1, w_2, \dots, w_n$ are often normalized so that their sum or maximum equals to some predefined number, typically 1 or 100. Generally, weights are subjective and need to be acquired from individual decision makers (Rezaei, et al., 2021; Silva, et al., 2021).

On one hand, MCDM methods strive to obtain weights that as accurately as possible represent decision maker's preferences. A good example is the method AHP (Analytic Hierarchy Process) (Saaty & Vargas, 2012), which proceeds by asking the user to assess relative importance of pairs of criteria, using the scale from 1 (equal importance) to 9 (extreme importance of one criterion over other). On this basis, AHP calculates criteria weights and assesses the consistency of user's information.

On the other hand, weights are subjective. Not only that they differ between different decision makers, they can change ("drift") also with the same person due to changes in the decision context, changes of their preferences or just their inability to express their preferences accurately enough. Thus, the question is how well can we assess criteria weights and what inter- and intra-personal differences should we expect.

This study is aimed at answering these questions in relation with the decision modelling method DEX (Decision EXpert) (Bohanec, 2022). DEX is a

qualitative MCDM method. It is somewhat specific in that it uses qualitative criteria and decision rules. Variables that represent criteria in DEX models are not numeric, but discrete and symbolic, using words as their values instead of numbers. For example, the criterion *Price* can be assessed using three categories "high", "medium", "low", and *Technical characteristic*s of some system can be assessed as "bad", "acceptable", "good", or "excellent". Consequently, in order to evaluate decision alternatives, DEX does not employ the weighted sum, but *decision rules* that take the general form:

if $x_1 = v_1$ and $x_2 = v_2$ and … and $x_n = v_n$
then $f(x_1, x_2, \ldots, x_n) = v_y$

Here, x_i are qualitative criteria and v_i are some categories taken from the corresponding value scales. Similarly as with weights, decision rules are acquired from the decision maker and conveniently represented in terms of *decision tables* (see example in Table 1).

In this study we addressed the following research questions:

A. *Inter-personal differences*: How much do decision tables acquired from different subjects differ?
B. *Intra-personal differences*: How much do decision tables acquired from the same subject at different times differ?
C. *Consistency*: Are decision rules, formulated by subjects, consistent and to which extent?
D. *Weight assessment*: Can subjects, who defined decision rules, also assess the weights of criteria and how well?

This study is considered preliminary because it has so far involved a relatively small number of respondents with narrow backgrounds, mostly students and researchers. Question B turned out particularly difficult, as it would require a well-controlled experimental setup with precise time differences between the trials, which we have not attempted so far.

2 Methods

The methodological approach is based on a questionnaire that asks respondents to define decision rules in two predefined decision tables, one assumingly easy and one more difficult. The experiment aims to exclude any tools that might help the respondents to formulate decision rules, thus the questionnaire is answered on paper. The response time is not limited (but is typically well within the 10-minutes range). All participants thus far were from Slovenia, therefore the questionnaire was formulated in the Slovenian language. An English translation is presented hereafter.

The first task (*Car*) is to define decision rules for evaluating a family car considering just two criteria: *Price* and *Technical characteristics*. An empty decision table consisting of 12 possible combinations of the criteria's discrete values is presented to the respondents, asking them to mark the corresponding values of *Car* (Table 1). In connection with this table, respondents are also asked to assess the weights of the two criteria, as shown in Table 2.

The second task (*Store*) is more difficult; it was inspired by the experiment designed by Vetschera, et al. (2014), but uses a reduced number of categories to keep the decision table reasonably small. The task is to assess the suitability/attractiveness of the store for daily purchases, primarily referring to purchases of groceries. Four qualitative criteria are suggested:

- *Store size*: "market" or "supermarket";
- Walking *distance* from home: "less than 10 minutes", "more than 10 minutes";
- *Price category*: "lower", "higher";
- Product *quality*: "lower", "higher".

There are the same four possible outcomes as with *Car*: "unacc", "accept", "good", and "excel".

Notice that the four *Store* criteria are binary (two-valued); this yields 16 possible value combinations, which are presented in the questionnaire in a table similar to Table 1. An analogous question to that from Table 2 is asked for the *Store* task, too.

The second task is considered more difficult than the first one because it involves twice as many criteria. The questionnaire itself does not interpret the criteria and their values any further, so we may expect subjective individual interpretations. In contrast with *Car*, where all the criteria are clearly preferentially ordered and we may expect that decision rules will reflect this order, this is much less so with *Store*. Buying habits largely differ between consumers, and while one may prefer a "lower" price category, some other may equally well prefer the "higher". Thus, we can hardly expect any clear preferential ordering of rules in the *Store* case.

Table 1: A table for acquiring decision rules for the evaluation of cars.

	Price	Tech.char.	Car			
			unacc	accept	good	excel
1	high	poor				
2	high	accept				
3	high	good				
4	high	excel				
5	medium	poor				
6	medium	accept				
7	medium	good				
8	medium	excel				
9	low	poor				
10	low	accept				
11	low	good				
12	low	excel				

Table 2: Question to assess *Car* weights.

Please assess criteria weights so that their sum equals 100:

Criterion	Weight
Price	
Tech.char.	
Sum	100

2.1 Differences between Decision Tables

Research questions A and B require the calculation of differences between two decision tables. A table T_t can be represented as:

$$T_t = \langle y_{t,1}, y_{t,2}, \dots, y_{t,k} \rangle$$

This is a vector of k ordinal numbers $y_{t,r} \in \{1,2,3,4\}, r = 1, \dots, k$, where r is the rule index, and k equals to 12 and 16 for the *Car* and *Store* tables, respectively. Notice that the lowest and highest possible vectors are $\langle \underbrace{111 \dots 1}_{k} \rangle$ and $\langle \underbrace{444 \dots 4}_{k} \rangle$. This gives the following formula for calculating the difference between two decision tables:

$$\Delta T_a, T_b = \frac{1}{k} \sum_{i=1}^{k} \frac{|y_{a,i} - y_{b,i}|}{3}$$

This formula yields the difference of 1 for the above extreme case, and 0 for two equal decision tables.

2.2 Consistency of Decision Rules

Whenever value scales of all involved criteria are preferentially ordered (from 'bad' to 'good' values or vice versa), we can assume that "rational" decision rules will be *consistent*: the better the input criteria (such as *Price* and *Technical characteristics*), the better the outcome (*Car*). Consequently, the decision table is expected to obey the *principle of dominance*, so that the aggregation function is *monotone*: improving or at least staying constant in the direction of each improving criterion.

In respondents' decision tables we observed and counted decision rules that violated the monotonicity constraint. In principle, well-defined *Car* decision tables are expected to be consistent. This is generally not true for *Store*, which was primarily aimed at assessing the average level of (in)consistency in that case.

2.3 Assessment of Weights from Decision Tables

Even though DEX is a qualitative method, for which the concept of criteria weights is somewhat unnatural, it is possible to approximately assess weights from a defined decision table. In this study, we used three methods. The first two, *Gini gain* (GG) and *Information gain* (IG), are routinely used in machine learning for determining the strength of features from data (Raileanu & Stoffel, 2004; Rokach & Maimon, 2015, pp. 62–63). These methods employ *impurity measures Gini* and *Entropy*, respectively, to assess the disorder of data (i.e., a DEX decision table T):

$$Gini(T) = 1 - \sum_{i=1}^{n} p^2(v_i)$$

$$Entropy(T) = -\sum_{i=1}^{n} p(v_i) \log_2 p(v_i)$$

Here, $p(v_i)$ is the probability/proportion of value v_i occurring in the table and n is the number of input criteria ($n = 2$ for *Car* and 4 for *Store*).

Using these measures, a relative weight RW of some criterion c is determined as

$$RW(c,T) = M(T) - \sum_{v \in S_c} \frac{|T_{c=v}|}{|T|} M(T_{c=v})$$

where M is an impurity measure (*Entropy* for IG and *Gini* for GG), v are values taken from S_c, the qualitative value scale of c, and $T_{c=v}$ denotes the part of T where $c = v$. $|T|$ denotes the size of table T in terms of the number of decision rules,.

The third method, *Linear approximation* (LA), is implemented in the software DEXi (Bohanec, 2020). It interprets decision rules as points in a multi-dimensional space and approximates them with a hyperplane using the least squares principle. Criteria weights are approximated from the slopes of the hyperplane: the higher the slope in the direction of a criterion, the higher the corresponding relative weight of the criterion. For more details about LA, see Bohanec & Zupan (2004, sec. 3.4) and Deguine, et al. (2021, sup. sec. 2).

Among the three methods, LA seems better suited for consistent, monotone and potentially linear decision tables, such as *Car*, while GG and IG might better capture the importance of variables in general data tables, such as *Store*.

2.4 Comparison of Weights

In order to compare weights $W_M = w_1, \dots, w_n$, estimated by some method M from decision rules, and weights $\Omega = \omega_1, \dots, \omega_n$ as given by the respondent, we used the formula:

$$\Delta W_M, \Omega = \frac{1}{n} \sum_{i=1}^{n} \frac{|w_i - \omega_i|}{100}$$

Here, n represents the number of criteria. The range of weights w_i and ω_i is [0,100]. Then, the range of $\Delta W_M, \Omega$ is from 0 (no difference) to 1 (extreme difference). The same formula is applicable for comparing weights obtained by two methods.

3 Preliminary Results

To date (May 2023), 34 participants answered the questionnaire, mostly students, colleagues at work and friends. Among these, 17 were approached about one month later and asked to answer exactly the same questionnaire; this served for the assessment of intra-personal differences between their first and second trials. The total number of collected questionnaires is 51. Although the number and composition of participants is inappropriate for drawing rigorous scientific conclusions, we already got interesting initial results and gained valuable experience to carrying on with larger-scale studies.

3.1 A: Inter-Personal Differences

Table 3 shows the minimal and maximal vectors observed in the experiment, and distances between them. It is clear that respondents' answers cover a large proportion of decision space between the possible extremes (which are all 1's and all 4's), and that the maximum distances are large, particularly with *Store*.

Table 3: Minimal and maximal observed vectors.

Task:	*Car*	*Store*
Min:	⟨111111221124⟩	⟨1112111111121111⟩
Max:	⟨123423443344⟩	⟨3444342434343444⟩
Distance:	0.53	0.81

Observing individual vectors also reveals a great variability of answers. In 51 questionnaires, there are as many as 47 different vectors for *Car*. This means that most of the vectors are distinct. There is a single most frequent vector ⟨111212341234⟩, which appears only 3 times. *Store* is even more extreme with 50 different vectors and only one appearing twice.

Average distances between these vectors are shown in the *Inter-personal* column of Table 4. They were calculated on all pairs of vectors of the first trial (561 pairs). The differences are roughly 14% and 24% for *Car* and *Store*, respectively.

Table 4: Average distances between vectors.

Task	*Inter-personal*	*Intra-personal*
Car	0.141 ± 0.062	0.114 ± 0.068
Store	0.237 ± 0.090	0.146 ± 0.069

3.2 B: Intra-Personal Differences

Intra-personal differences were assessed on 17 pairs of questionnaires that were answered by the same participants at two different times. Average distances between vectors in this case are shown in the third column of Table 4. They are about 11% and 15% for *Car* and *Store*, respectively.

As expected, intra-personal differences are smaller than inter-personal ones. However, they are not *substantially* smaller: the ratios between inter- and intra-personal distances in about one month time were $\frac{0.114}{0.141} = 0.81$ for *Car* and $\frac{0.146}{0.237} = 0.61$ for *Store*.

These results indicate that subjects' preferences drift a lot over time and that it is difficult for individuals to provide the same decision rules twice. At this point, this study cannot really explain the drift; we can only speculate about the effects of changed preferences, changed decision context, bad memory and even imprecise, inaccurate or otherwise "elusive" nature of decision rules.

3.3 C: Consistency of Decision Rules

As expected, the majority (45 of 51 ≅ 88%) of *Car* decision tables defined by respondents contain consistent decision rules. Even in the remaining 6 tables, the level of consistency is high: one 92%, one 94% and four 97%. This indicates that for relatively small decision tables (*Car*) that use preferentially ordered criteria, the users are generally capable of formulating decision rules of perfect or very high consistency without any help.

The consistency of decision tables for *Store* is much lower, as shown in Figure 1. The average consistency is only 0.77. Again, this was expected to some extent, but not that much. In this study, we cannot really explain the reasons, which might be due to using preferentially unordered criteria, non-monotone consumer behaviour, or just due to the sheer size of the decision space to cover (four binary criteria). This remains an open challenge for future research.

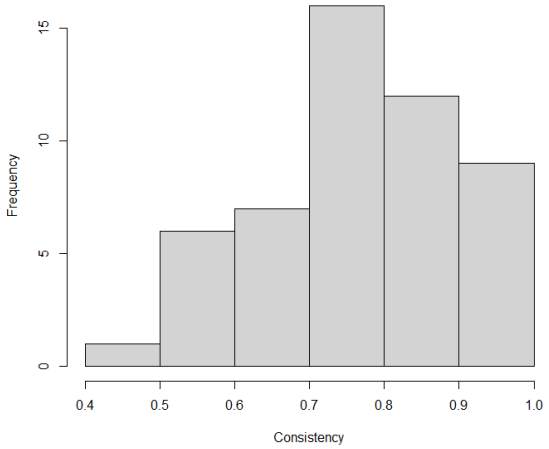


Figure 1: Consistency of *Store* decision rules.

Another interesting consistency-related question is: did those 17 participants that answered the questionnaire twice defined consistent (or, respectively, inconsistent) rules both times? For *Car*, the answer is no. None of the participants defined inconsistent rules both times; there were 3 participants who improved and 2 participants who degraded their consistency with time. With *Store*, most of the participants formulated inconsistent decision tables; 6 of them improved and 10 degraded the consistency measure.

3.4 D: Assessment of Weights

Figure 2 displays a boxplot of weights of *Car* decision rules as assessed by the participant (Ω) and by the three methods defined in section 2.3: LA, IG and GG.

On the one hand, we can see that participants, in average, assessed the two criteria, *Price* and *Technical characteristics*, almost equally, with a slight statistical leaning towards the latter. On the other hand, all the methods clearly indicated that, according to defined decision rules, *Technical characteristics* are far more important than *Price*. Here we can claim that users did not assess their weights really well, while the methods were largely consistent with each other.

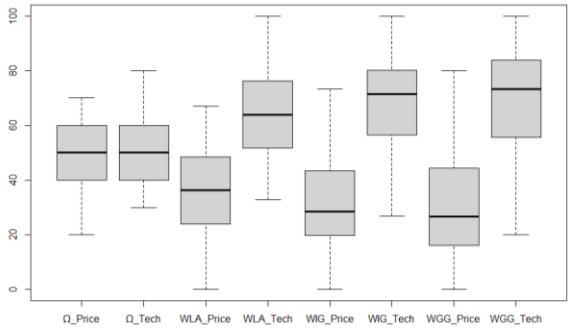


Figure 2: Weights of *Car* criteria assessed by the participant (Ω) and three methods, LA, IG and GG.

Results for *Store* are shown in Figure 3, where participants' weights (Ω) are compared with those assessed by the IG method. In contrast with *Car*, at least the order of criteria's importance was estimated correctly by participants (*Size* being the least and *Quality* the most important). Again, comparing human and algorithmic weights, the former are less extreme and all leaning towards 20–30%, while the latter are more extreme, ranging from about 10% to 50%. Yet again, the three methods turned out similar to each other (these results are not shown here).

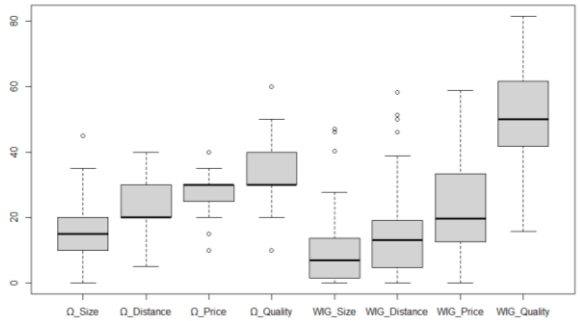


Figure 3: Weights of *Store* criteria assessed by the participant (Ω) and method IG.

Table 5: Differences between weight assessments of participants (Ω) and methods LA, GI and GG.

	Car			*Store*		
	LA	GI	GG	LA	GI	GG
Ω	0.086	0.137	0.159	0.062	0.099	0.116
LA		0.065	0.087		0.063	0.080
GI			0.038			0.034

Table 6: Weight differences between two trials (*Car*).

Car	Ω	LA	GI	GG
Ω	0.089	0.154	0.159	0.192
LA		0.127	0.114	0.152
GI			0.102	0.110
GG				0.126

Table 7: Weight differences between two trials (*Store*).

Store	Ω	LA	GI	GG
Ω	0.080	0.112	0.128	0.134
LA		0.098	0.121	0.121
GI			0.116	0.124
GG				0.126

The results of comparing all pairs of weight assessments (by Ω, LA, GI and GG) and calculating their difference using the formula from section 2.4, are shown in Table 5. Generally, differences are small, and particularly small differences (in the 4% range) are between GI and GG. Differences between LA vs. GI and GG are somewhat larger (about 7%). Differences between Ω and the remaining three methods are all greater than 6%. However, coming as a surprise and contrary to our expectations, the differences between Ω and LA turned out substantially lower than those of Ω vs. GI and GG. This indicates that the method LA, as implemented in DEXi, actually relatively well resembles user-assessed weights even for non-monotone decision tables.

Finally, Tables 6 (for *Car*) and 7 (for *Store*) compare weights between the first and second trials for the 17 participants that answered the questionnaire twice. The most important observation is that weights do change in time, however, the participants' assessments change less than those of the three methods. Roughly, we may expect about 9–19% weights change between subsequent trials, which are separated by about one month time.

4 Conclusion

At some point, we began calling this study *Aurora*. Considering the discrete nature of DEX decision tables, one might assume that there is little freedom in defining "vectors" of decision rules and that all decision tables should look the same. As we obtained more results from the study, it became increasingly evident that the opposite is in fact true. Humans are incredibly diverse and almost all decision tables, obtained in this experiment, were distinct. The search for "accurate" and "stable" preferences, represented with decision rules, turned to something like a quest for *Aurora Borealis*, which is real, but constantly changing its shape and colours.

The most important findings of this study are:

- Decision makers' preferences, expressed in terms of decision tables, are very diverse, peaking at inter-personal differences of 53% and 81% in the two investigated tasks (*Car* and *Store*), and the respective averages of 14% and 24%.
- Preferences of a single decision maker in time (intra-personal differences) also change a lot. Not as much as inter-personal differences, but close (11% and 15% in this study).
- For relatively small tables and preferentially ordered criteria (such as the *Car* task in this study), decision makers are perfectly capable of formulating consistent (or almost consistent) decision tables without any instructions and any supporting software tools.
- For larger decision tables with less clear ordering of criteria values, the consistency deteriorates

drastically. The reasons are yet to be determined by additional studies.

- Not only do the decision tables change over time, but the criteria weights, which are provided by the user and assessed from decision rules, also vary. The observed differences in criteria weights range between 9% and 19%.

Due to the small number and possibly biased composition of participants, these results should be considered preliminary. However, they have already provided some indications that contradicted our initial expectations, such as those that DEX decision tables are "rigid" and the variation of decision rules is small.

In the future, we wish to carry out a similar study on a larger and more diverse sample of participants. By doing so, we can establish concrete research conclusions and compare them with experience using common numeric weight-based MCDM methods.

The main challenge remains the question of how to "catch" participants, who are by default anonymous, for their second trial in a not too random time interval between the trials. Also, we wish to avoid respondents' "cheating", i.e., remembering and submitting exactly the same answers in both trials. Building on the results of this study, we aim to develop methods for identifying reasons for decision-rules drifting over time and their potential inconsistency. However, as we are pleased with the simplicity of the current questionnaire, we wish not to complicate it further with additional research questions.

Acknowledgments

The author acknowledges the financial support from the Slovenian Research Agency, research core funding P2-0103. The author also expresses his gratitude to students of the University of Nova Gorica, colleagues from Jožef Stefan Institute and friends who participated in the study.

References

Bohanec, M., & Zupan, B. (2004). A function-decomposition method for development of hierarchical multi-attribute decision models, *Decision Support Systems* 36, 215–233. doi: 10.1016/S0167-9236(02)00148-3

Bohanec, M (2020). *DEXi: Program for Multi-Attribute Decision Making, User's Manual, Version 5.04*. IJS Report DP-13100, Jožef Stefan Institute, Ljubljana, 2020. https://kt.ijs.si/MarkoBohanec/pub/DEXiManual504.pdf

Bohanec, M. (2022). DEX (Decision EXpert): A Qualitative Hierarchical Multi-criteria Method. In: A. J. Kulkarni (Ed.), *Multiple Criteria Decision Making* (Vol. 407, pp. 39–78). Singapore: Springer. doi: 10.1007/978-981-16-7414-3_3

Deguine, J.-P., Robin, M.-H., Camilo Corrales, D., Vedy-Zecchini, M.-A., Doizy, A., Chiroleu, F., Quesnel, G., Païtard, I., Bohanec, M., & Aubertot, J.-N. (2021). Qualitative modeling of fruit fly injuries on chayote in Réunion: Development and transfer to users. *Crop Protection* 139, doi: 10.1016/j.cropro.2020.105367.

Greco, S., Ehrgott, M., & Figueira, J. (2016). *Multi Criteria Decision Analysis: State of the art Surveys*. New York: Springer. doi: 10.1007/978-1-4939-3094-4

Kulkarni, A.J. (2022). *Multiple Criteria Decision Making*. Studies in Systems, Decision and Control 407, Singapore: Springer, doi: 10.1007/978-981-16-7414-3_3.

Raileanu, L.E., & Stoffel, K. (2004). Theoretical Comparison between the Gini Index and Information Gain Criteria. *Annals of Mathematics and Artificial Intelligence* 41(1):77-93. doi: 10.1023/B:AMAI.0000018580.96245.c6

Rezaei, J., Arab, A., & Mehregan, M. (2021). Equalizing bias in eliciting attribute weights in multiattribute decision-making: experimental research. Journal of Behavioral Decision Making, 35(2). doi: 10.1002/bdm.2262

Rokach, L., & Maimon, O. (2015). *Data Mining with Decision Trees: Theory and Applications*. New Jersey: World Scientific. doi: 10.1142/9789812771728_0001

Saaty, T.L., & Vargas, L.G. (2012). *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. New York: Springer. doi: 10.1007/978-1-4614-3597-6

Silva, F.F., Souza, C.L.M., Silva, F.F. et al. (2021). Elicitation of criteria weights for multicriteria models: Bibliometrics, typologies, characteristics and applications, *Brazilian Journal of Operations & Production Management* 18(4). doi: 10.14488/BJOPM.2021.014

Thakkar, J.J. (2021). *Multi-Criteria Decision Making*. Studies in Systems, Decision and Control, Vol. 336. Singapore: Springer. doi: 10.1007/978-981-33-4745-8

Vetschera, R., Weitzl, W., & Wolfsteiner, E. (2014). Implausible alternatives in eliciting multi-attribute value functions. *European Journal of Operational Research*, 234(1), 221–230. doi: 10.1016/j.ejor.2013.09.016

Perceived Threat of Cyber Attacks and its Role in the Adoption of Tablet Computers by Older Adults

Kaja Prislan Mihelič, Igor Bernik, Simon Vrhovec, and Anže Mihelič
University of Maribor
Faculty of Criminal Justice and Security
Kotnikova 8, 1000 Ljubljana, Slovenia
{kaja.prislan, igor.bernik, simon.vrhovec, anze.mihelic}@um.si

Abstract. *This study explores factors affecting tablet adoption among older adults in Slovenia. Findings highlight significant differences across three adoption groups: those without a tablet, with an unused tablet, and active users. Perceived usefulness and existing knowledge greatly influence adoption intention, while cybersecurity concerns and anxiety contribute to usage fear and potential benefits' rejection. As these concerns are universal across all adoption stages, strategies addressing cybersecurity, including clear cyber safety information, online safety training, and accessible cybersecurity infrastructure, are essential for promoting tablet use among older adults.*

Keywords. Technology adoption, older adults, tablets, tablet computers, cybersecurity, fear of cyber attacks

1 Introduction

The demographic transition towards an aging population represents a significant shift and an urgent societal concern in the current era. Individuals aged 65 and above constitute a substantial proportion of the population in Europe, a trend projected to escalate to a quarter of the population in the European Union and North America in the upcoming decades (United Nations, 2022; WHO, 2022). This rising trend spawns a multitude of challenges that demand immediate and creative solutions.

Addressing the issues associated with an aging population is pivotal for the successful realization of the United Nations' 2030 Agenda for Sustainable Development, considering its overlap with Sustainable Development Goals (SDGs) such as poverty eradication, health and wellbeing, gender equality, economic growth, decent work, reduced inequalities, and sustainable cities. In order to foster genuinely transformative, inclusive, and sustainable development outcomes, it is imperative to perceive older persons not solely as a vulnerable group but as active participants contributing to societal progress (United Nations Development Programme, 2017).

The proliferation of digital technologies, which have become increasingly intertwined with everyday life, presents unique solutions for issues associated with an aging population. Since 2019, the digital age has seen a marked surge in internet and information and communication technology (ICT) usage, notably involving mobile and smart devices. This growth was particularly pronounced during the pandemic (Deloitte, 2021; ITU, 2021).

However, a significant digital divide persists between younger and older generations, with the older adults demonstrating a lesser inclination to embrace technology compared to the broader population (Anderson & Perrin, 2017). Despite this, smart devices hold the potential to enrich the quality of life for all age demographics. Owing to their accessibility and user-friendly interfaces, these devices have led some individuals to forgo traditional computers, opting instead for mobile devices, particularly tablets (Chan et al., 2016).

The aim of this preliminary study is to investigate the adoption of tablets among older adults in Slovenia, i.e., individuals aged 65 and above, and to gauge their perceived threat level from cyber-attacks. The study seeks to understand the influence of various factors, such as technology adoption habits, behavioral tendencies, and security considerations, on this population's tablet usage. The survey also aims to analyze participant's experiences with using tablets, and to measure constructs like perceived usefulness, ease of use, and others, with the ultimate goal of understanding how these elements contribute to the adoption or rejection of tablets among older adults.

2 Theoretical Background

The penetration rate of tablet PCs, akin to other frequently employed ICT, has witnessed significant fluctuations over time. In 2012, the penetration rate in selected European countries varied from 14.9% to 22.3%. However, by 2020, this rate had seen a substantial upswing, ranging from 61% to 73.5%. The Netherlands registered the highest penetration rate of

tablet PCs in the European Union (66%), whereas Ukraine reported the lowest rate (14%) (Statista, 2023b, 2023a). In 2021, approximately 30% of individuals across 27 European Union countries utilized tablets for internet access, with 20% of individuals aged between 65 and 74 partaking in this practice (Eurostat, 2023a). Parallel studies conducted in the United States revealed that roughly one-third of seniors aged 65 and older owned tablet computers in 2016. Tablet ownership was more prevalent among seniors with advanced educational attainment and those residing in higher-income households (Anderson & Perrin, 2017).

While older adults may not exhibit the same level of proficiency with emerging technologies as younger generations, who grew up in the computer age, the utilization of smart devices, such as tablets, can proffer a multitude of benefits and flexible solutions for this demographic group (Vaportzis et al., 2017). Tablets can serve as a conduit for maintaining social connections, assist in medication management, facilitate setting reminders, and provide cognitive-stimulating games.

The user-friendly nature and adaptability of these devices are particularly advantageous for users with visual impairments, aiding in reading and accessing critical data. Different features enabling automated texting amplify the communicative potential for those with hearing and mobility limitations (Kim et al., 2022; Tsai et al., 2015; Vasconcelos et al., 2012). As a result, tablets can exert a positive influence on the cognitive abilities, health, and overall well-being of the older adults, particularly in alleviating feelings of isolation and depression often concomitant with decreased mobility (Chan et al., 2016; Vaportzis et al., 2018).

While tablets can be highly beneficial for older adults, their utilization is not without its limitations and barriers. These barriers can be classified into several categories, including a lack of instructions and guidance, limited knowledge and confidence, health-related issues, prohibitive costs, insufficient accommodation for sensory and motor impairments, and limited input from clinicians and researchers (Kim et al., 2022; Vaportzis et al., 2017; Wilson et al., 2022).

Moreover, the susceptibility to cyber threats and the risk of falling victim to cybercrime has become a significant concern for older adults operating smart devices and tablets. This demographic is particularly vulnerable to cyber-attacks due to their limited digital proficiency, restricted experiences, and a lack of knowledge in this arena (Blackwood-Brown et al., 2021; FBI, 2022; Virtanen, 2017). Furthermore, mild cognitive impairment can undermine their ability to discern scams, adhere to recommended security protocols, and evaluate the risks associated with their online behaviors (Mentis et al., 2019).

The latest Special Eurobarometer on Europeans' attitude towards cybersecurity revealed that up to a third of older users have personally experienced or been victims of cybercrime, with frauds, scams, and malicious software constituting the most prevalent types of crimes. Older users are also less likely to undertake security-related measures and feel less informed about cybercrime, thereby exacerbating their vulnerability. This digital divide is particularly acute among older, less affluent individuals with lower educational attainment, which underscores the need for enhanced support and assistance in using tablets and addressing their limitations and barriers (European Commission, 2020).

These findings from the Eurobarometer align closely with the data provided by Eurostat (2023b) emphasizing the pressing issue of internet security among the older adult population within the European Union (EU). In 2019, it was reported that approximately 19% of this demographic encountered issues related to online safety. The most prevalent issue was phishing scams or fraudulent emails. Pharming scams were the next most common concern, followed by the loss of documents, images, or other data due to computer infections, such as viruses, worms, or Trojan horses.

Therefore, the population of older adults requires additional support and guidance in the safe and secure usage of tablets and other digital technologies. By addressing these barriers and providing appropriate training and resources, we can help bridge the digital divide and enable senior citizens to harness the numerous benefits that tablets and technology can offer.

Thus far, an extensive body of research has been dedicated to investigating the engagement of the elderly population with smart devices. This collection of scholarly work focused on tablet usage can generally be compartmentalized into several distinct thematic categories.

Firstly, a prominent area of investigation has centered on the prevalence, adoption rates, and familiarity of tablets amongst the older adults, with a particular emphasis in the context of the COVID-19 pandemic (Kim et al., 2022; Vaportzis et al., 2018). These studies suggest that tablets have become an integral part of daily routines for many older users, concurrently advocating for enhancements in future training initiatives. Secondly, research examining the benefits and challenges associated with tablet use among the older adults has revealed several obstacles including a lack of technological familiarity, diminished confidence, and inadequate training. Conversely, tablets offer numerous advantages to this demographic, such as entertainment, social connectivity, and a platform for seeking information (Vaportzis et al., 2017; Wilson et al., 2022). Nonetheless, a note of caution has been raised regarding techno-optimism (Cid et al., 2020), with some studies suggesting that tablet use among elderly residents in long-term care facilities may not significantly ameliorate cognitive decline, loneliness, or self-esteem. A third area of research focuses on tablet design and interaction techniques (Petrie &

Darzentas, 2017; Sciarretta et al., 2015), emphasizing the potential of appropriate design methodologies in augmenting the value of tablet applications for older users. Lastly, emerging research topics include the investigation of adoption and motivation to use tablets among the older adults. Studies have found that perceived ease of use and usefulness, as well as positive perceptions toward the devices, serve as significant factors influencing adoption (Chee et al., 2022; Tsai et al., 2015).

Notwithstanding, recent studies have detected a degree of skepticism towards tablets and technology in general among senior citizens, which can pose substantial barriers to adoption and effective usage (Vaportzis et al., 2017). Furthermore, privacy and cybersecurity concerns have been linked to the usage of tablet applications among older adults (Moyle et al., 2020). Drawing upon previous research, it can be deduced that despite the generally positive reception of tablets, a number of obstacles obstruct greater or more rapid adoption rates within the elderly demographic.

While a considerable number of studies have explored smart devices and tablet adoption among the older adults, it is imperative to acknowledge the myriad factors influencing this process, intrinsically tied to their experiences and perceptions. Yet, existing research has rarely adopted a comprehensive approach to examine the diverse factors influencing tablet adoption among the older adults, including cybercrime fears. Therefore, the primary objective of this research is to delve into tablet adoption among older individuals, particularly focusing on discerning differences between adoption groups and correlated factors. Moreover, this study aims to scrutinize the role of perceived threat in relation to prospective adoption factors, with the ultimate objective of gaining a more profound understanding of the dynamics that influence tablet adoption among the older adults. The premise underpinning this research is that understanding the barriers and facilitators to tablet adoption among the older adults is crucial for the design of effective interventions. Addressing these factors could promote technology adoption, thereby enhancing the quality of life for this burgeoning demographic.

3 Method

To examine the adoption of tablet use among older adults and how much they feel threatened by cyber-attacks, we conducted a survey among older adults in Slovenia, explicitly targeting individuals aged 65 years and above. This survey was conducted in February and March 2023. The recruitment of participants for the study was carried out by students, who were given clear instructions to reach out to a diverse group of individuals in their immediate and extended circles. This included their grandparents, neighbors, acquaintances, and randomly selected individuals from their respective hometowns.

These students were trained to provide necessary assistance to the participants, with a key focus on ensuring that the participants had a full understanding of the survey's context and objectives. This involved clarifying the survey details outlined on the introduction page and explaining central concepts integral to the survey, such as the concepts of tablets and cyberattacks. Furthermore, the students provided support to participants who experienced difficulty in choosing a response, such as when expressing agreement or disagreement with a statement on a 5-point Likert scale. The goal was to ensure that participants felt confident in their responses and that their answers genuinely reflected their perceptions and experiences.

It's important to note that the students were strictly directed to assist in clarifying the content and in explaining how to provide responses. They were explicitly told not to sway the respondents' answers in any manner.

The survey questionnaire was structured to assess a range of factors related to technology adoption, behavioral tendencies, and security. These factors were categorized into several constructs, including perceived usefulness, ease of use, subjective norm, facilitating conditions, hedonic motivation, anxiety, fear, perceived threat, attitude towards use, intention to use, and rejection of benefits. We employed a mix of four items and three items to measure subjective norm, ease of use, and the other constructs, respectively. The questionnaire also incorporated questions specifically designed to understand participants' experiences with using tablets and their demographic characteristics.

The study was conducted in ethical compliance, as demonstrated by the approval of the Faculty of Criminal Justice and Security Research Ethics Committee at the University of Maribor.

The survey was successfully completed by 119 older adults. Of these, 54.6% were female, and 45.4% were male. The age distribution among the respondents was broad, with ages ranging from 65 to 93 years, and the median age being 70 years. Furthermore, the participants were almost equally distributed between urban (44.5%) and rural (55.5%) environments, offering a balanced perspective from both types of settings.

The reliability of the survey instrument was assessed using Cronbach's alpha (CA), resulting in values ranging from 0.850 to 0.979, which were comfortably above the standard threshold of 0.800 for all constructs. An exception was noted for the construct of facilitating conditions, with a CA of 0.357. Being a formative construct, it is composed of diverse dimensions. The lower reliability for this construct suggests that these dimensions may not be as interconnected as in other contexts, such as organizational settings, where the reliability for this construct is typically adequate. To analyze the data, we computed the mean for all constructs, except for facilitating conditions.

	All (N = 119)	Do not have (N_1 = 73)	Do not use (N_2 = 14)	Use (N_3 = 32)	*F*
Perceived usefulness	3.15	2.69	3.24	4.16	24.9831***[a,b]
Ease of use	2.82	2.43	2.39	3.91	17.4007***[a,b]
Subjective norm	3.43	3.08	3.70	4.11	14.9530***[b]
Hedonic motivation	3.50	3.09	3.43	4.46	15.8834***[a,b]
FC1: Financial status	3.35	2.95	4.00	4.00	9.2069***[b,c]
FC2: Knowledge	2.50	1.89	2.36	3.97	37.9726***[a,b]
FC3: Support	4.03	3.86	4.36	4.25	1.6026
Attitude towards use	3.46	2.98	3.50	4.52	25.1797***[a,b]
Perceived threat	3.00	3.18	3.19	2.52	2.9304
Anxiety	2.24	2.65	2.17	1.33	15.0772***[b]
Fear of use	1.90	2.28	1.69	1.13	12.0826***[b]
Rejection of benefits	2.57	2.76	2.71	2.06	6.5233**[b]
Intention to use	3.09	2.33	3.38	4.70	68.6113***[a,b,c]

FC - facilitating conditions; *** p < .001, ** p < .01,* p < .05; [a] difference between groups »use« and »do not use«, [b] difference between groups »use« and »do not have«, [c] difference between groups »do not use« and »do not have«

Table 1: Differences among adoption grous (pre-adoption, non-adoption, post-adoption).

4 Results

This section reports the results of our study, focusing primarily on two distinct areas: technology adoption and perceived cyberattack threat. We present a detailed examination of these aspects and their significant findings, providing insights into the behavior and perceptions of older adults regarding tablet use.

4.1 Technology Adoption Aspect

First, we focused on examining the differences between three distinct adoption groups: those who do not possess a tablet (pre-adoption), those who own but do not use a tablet (non-adoption), and those who actively use a tablet (post-adoption). This detailed investigation was facilitated by conducting an Analysis of Variance (ANOVA) for all the constructs that were measured during the survey. The ANOVA aimed to identify any substantial differences among the adoption groups. To further refine our analysis, we employed Bonferroni post-hoc tests. These tests were crucial in establishing whether there were significant differences between each individual adoption group. Our findings from these tests, which offer valuable insights into the three groups' behavioral tendencies and technology adoption habits, are presented comprehensively in Table 1.

The analysis of our survey results revealed statistically significant differences across adoption groups for all the constructs we measured (perceived usefulness, ease of use, subjective norm, facilitating conditions, hedonic motivation, anxiety, fear, perceived threat, attitude towards use, intention to use, and rejection of benefits) with the exception of perceived threat and support, a component of facilitating conditions. This finding suggests that feelings of threat from cyberattacks are universally experienced across all adoption groups, irrespective of their tablet usage status. In other words, the level of perceived threat from cyber incidents remains consistent whether respondents are pre-adoption, non-adoption, or post-adoption group members. Furthermore, our results show that all participants potentially have access to some form of support, likely due to the sampling strategy we used. This is illustrated by the high average score of 4.03 for item FC3, which was the highest mean score across all the items in the survey.

In addition to the analysis mentioned earlier, we also conducted a linear regression to delve deeper into the factors associated with the 'intention to use' (the dependent variable in our model). We included in our regression model all the factors outlined in Table 1, along with 'age' as a control variable, considering these as independent variables. The model explained over 70 percent of the variance, as indicated by an adjusted R-squared ($R^2 = 0.715$).

However, we found that only two factors – perceived usefulness ($\beta = 0.438$; $p < 0.001$) and knowledge (a dimension of 'facilitating conditions', ($\beta = 0.418$; $p < 0.001$) – showed statistical significance. This suggests that the respondents may be more inclined to use tablets if they perceive them as being useful and if they already possess the necessary knowledge to operate them. The findings underscore the critical role that the perceived utility of tablets and the existing knowledge of potential users play in the decision-making process related to tablet use.

4.2 Perceived Cyberattack Threat Aspect

To understand the role of 'perceived threats,' we explored various associations that could yield

meaningful insights. Specifically, we designed a linear regression model to predict the 'fear of use' (our dependent variable). In this model, we considered 'perceived threat,' 'anxiety,' and 'age' as independent variables. This model was successful in explaining more than 35% of the variance (adjusted $R^2 = 0.374$). Within these independent variables, 'perceived threat' ($\beta = 0.163$; $p < 0.05$) and 'anxiety' ($\beta = 0.489$; $p < 0.001$) emerged as significant, while 'age' did not. These findings offer valuable insights into the factors contributing to older adults' fear of using tablets. It appears that this fear could be partially rooted in anxiety related to tablet usage and partially driven by their perception of vulnerability when using a tablet. We believe that future research could benefit from further exploring these relationships to establish causal links.

Although seemingly pertinent for a considerable portion of older adults, the concept of 'rejection' has been less explored in previous studies. We, therefore, constructed a linear regression model to predict 'rejection of benefits' (our dependent variable) using the same independent variables as before. This model explained more than 25% of the variance (adjusted $R^2 = 0.258$). Consistent with the previous model, 'perceived threat' ($\beta = 0.309$; $p < 0.001$) and 'anxiety' ($\beta = 0.292$; $p < 0.01$) were found to be significant. Our findings suggest that the more threatened older adults feel by cyber-attacks or, the more anxious they are about tablets, the more likely they are to reject the potential benefits of tablets.

5 Discussion

Following the example in the results section, this section, explores two primary aspects: the technology adoption aspect and the perceived cyberattack threat aspect. We examine their relationship with the adoption and usage of tablets among older adults, and discuss how these factors can shape their behavior, attitudes, and decisions related to technology.

5.1 Technology Adoption Aspect

In our study, the classification of participants into three distinct groups - those who do not possess a tablet (pre-adoption), those who own but do not use a tablet (non-adoption), and those who actively use a tablet (post-adoption) - enabled us to examine the nuanced differences in behaviors, attitudes, and perspectives within each category.

The analysis revealed substantial differences among these groups across most of the constructs we investigated. This signifies that the technology adoption process among older adults is complex and layered, influenced by a broad array of attitudes, perceptions, and experiences related to technology. Interestingly, one of the standout findings of our study was the consistency of perceived threats from cyberattacks across all adoption groups. Regardless of whether the respondents already owned and used a tablet, had a tablet but did not use it, or did not possess a tablet at all, the fear of cyberattacks was universally present. This underlines the pervasiveness and influence of cybersecurity concerns in shaping technology adoption decisions among older adults. It underscores the necessity for adequate education and reassurance about safety measures when promoting tablet use among this demographic.

In addition, our results indicated that all respondents potentially had access to some form of support, with a high average score for the item FC3. Given that our sample comprises individuals from diverse backgrounds and locations, it is promising to see evidence of a support network. It is well-documented that having a support system can facilitate technology adoption, especially among older adults who might need assistance navigating new devices or features. Therefore, such a finding is an encouraging sign and suggests that our participants might not be as isolated from assistance as one might expect.

The linear regression model designed to examine the factors associated with the 'intention to use' tablets explained over 70% of the variance and highlighted perceived usefulness and existing knowledge about tablets as significant predictors of adoption. This finding suggests that the decision to use tablets among older adults is not a random or impulsive but a calculated decision influenced by their perception of the tablet's usefulness and their confidence in their ability to operate it. In essence, older adults are more likely to adopt tablets if they see a practical value in them, such as staying connected with family or accessing services, and if they feel they have the requisite knowledge to use them effectively. This highlights the critical need for both - promoting the benefits of tablet usage in a way that resonates with older adults and providing comprehensive and accessible education and training initiatives to equip them with the necessary skills and knowledge to use tablets.

5.2 Perceived Cyberattack Threat Aspect

The perception of threats from cyberattacks was consistent across all adoption groups, signifying its universal presence in the technology adoption equation among older adults. A particularly noteworthy is its role in shaping two specific constructs - the fear of use and the rejection of benefits - which have significant implications for tablet adoption. The linear regression model, which aimed to predict fear of use, revealed that both perceived threat and anxiety were significant contributors, explaining more than 35% of the variance. This suggests that the fear of using tablets among older adults is not solely a product of their unfamiliarity with the technology but is significantly influenced by their cybersecurity concerns and overall anxiety related to tablet usage.

The significant association between perceived threat and fear of use brings to the forefront the importance of cybersecurity concerns among older adults. It indicates that even when older adults are willing to adopt new technology, the fear of potential cyber threats can act as a formidable barrier, possibly causing them to avoid using tablets. This fear might be exacerbated by media reports on cyberattacks and data breaches, which could further fuel their doubts about the safety of using technology. Moreover, our study also found a significant association between perceived threats and anxiety, suggesting that heightened perception of cyber threats can contribute to increased anxiety about using tablets. This underscores the psychological toll cybersecurity concerns can have on older adults, possibly making them more hesitant about embracing and accepting digital technology.

The second regression model, predicting the rejection of benefits, also highlighted the role of perceived threats and anxiety. These two factors were significant predictors, accounting for more than 25% of the variance. This indicates that the more threatened older adults feel by cyberattacks, or the more anxious they are about tablets, the more likely they are to reject the potential benefits of tablets. This finding suggests that cybersecurity concerns might deter older adults from using tablets and lead them to dismiss the potential advantages these devices offer. This could include staying connected with family and friends, accessing information, online shopping, or using health services.

Our study suggests that the perceived threat of cyberattacks can have a pervasive and multifaceted impact on tablet adoption among older adults, influencing not only their usage habits but also their attitudes toward the benefits of technology. It highlights the need for targeted strategies to address these concerns, such as providing clear and comprehensible information on cyber safety measures, offering training on safe online practices, and establishing robust cybersecurity infrastructure and protocols that are understandable and accessible to older adults.

6 Limitations

Before drawing conclusions, we must acknowledge several limitations within our study. First, our findings are preliminary, forming part of an ongoing larger-scale study. As more data are gathered and more responses accumulate, we can conduct further and more in-depth analyses. For instance, examining each adoption group separately and more thoroughly will be possible.

Second, the adoption groups in our study were not uniformly distributed. This uneven distribution could potentially impact the reliability of our findings and interpretations. Future research might need to consider more balanced data collection strategies to ensure a more even distribution among the adoption groups. Such strategies would help derive more robust and reliable insights into older adults' adoption behaviors and attitudes toward tablet computers.

Third, despite our efforts to form survey items with the older demographic in mind, respondents could still find them challenging to comprehend. Future research should consider further simplifying the survey questionnaire to ensure our targeted older demographic easily understands it. An easier-to-understand questionnaire would not only enhance the validity of the responses collected but would also open up possibilities for different sampling strategies. These strategies could reach respondents who do not have any support while completing the survey, thereby broadening the participant base and enhancing the generalizability of the findings.

7 Conclusion

In conclusion, our research sheds light on the factors influencing tablet adoption among older adults in Slovenia. The results highlight the importance of perceived usefulness and existing knowledge in determining an individual's intention to use technology. The study also indicates that perceived threats and anxiety significantly influence the fear of using tablets and the tendency to reject their benefits. These findings provide a deeper understanding of the complex dynamics of technology adoption among older adults, further highlighting the need to address these challenges to facilitate their digital inclusion.

Future research should focus on conducting a more in-depth study of each adoption group, striving for a more balanced representation of participants across different adoption stages, and further simplifying the questionnaire for better comprehension. Future research should also employ alternative sampling strategies to reach older adults who may not have support while completing the survey. Additionally, future research should delve into the role of different forms of support, such as family, friends, or community resources, in the technology adoption process among older adults. Lastly, future research should investigate other factors that might influence technology adoption among older adults, including health status, previous technology experience, attitudes toward change, and the perceived relevance of technology in their day-to-day lives.

Acknowledgments

We thank the students who played a vital role in connecting with potential respondents and all the participants who dedicated their time to contribute to our study.

This work was partially funded by the Slovenian Research Agency [Grant Number J5-3111].

References

Anderson, M., & Perrin, A. (2017). *Tech Adoption Climbs Among Older Adults.* https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2017/05/PI_2017.05.17_Older-Americans-Tech_FINAL.pdf

Blackwood-Brown, C., Levy, Y., & D'Arcy, J. (2021). Cybersecurity Awareness and Skills of Senior Citizens: A Motivation Perspective. *Journal of Computer Information Systems*, *61*(3), 195–206.

Chan, M. Y., Haber, S., Drew, L. M., & Park, D. C. (2016). Training older adults to use tablet computers: Does it enhance cognitive function? *Gerontologist*, *56*(3), 475–484.

Chee, C. S., Cheo, J. J. Y., Chua, K. J. L., Chia, R. Z. J., & Lim, T. J. (2022). Evaluating the feasibility and perceptions of a senior-friendly tablet among older adults in Singapore: A mixed-methods study. *Gerontechnology*, *21*(1). https://doi.org/10.4017/GT.2022.21.1.597.10

Cid, A., Sotelo, R., Leguisamo, M., & Ramírez-Michelena, M. (2020). Tablets for deeply disadvantaged older adults: Challenges in long-term care facilities. *International Journal of Human Computer Studies*, *144*. https://doi.org/10.1016/j.ijhcs.2020.102504

Deloitte. (2021). *How the pandemic has stress-tested the crowded digital home*. https://www2.deloitte.com/content/dam/insights/articles/6978_TMT-Connectivity-and-mobile-trends/DI_TMT-Connectivity-and-mobile-trends.pdf

European Commission. (2020). *Europeans' attitudes towards cyber security - Special Eurobarometer 499*. https://op.europa.eu/en/publication-detail/-/publication/468848fa-49bb-11ea-8aa5-01aa75ed71a1/language-en

Eurostat. (2023a). *Individuals - devices used to access the internet*. https://ec.europa.eu/eurostat/databrowser/view/ISOC_CI_DEV_I/default/table?lang=en&category=isoc.isoc_i.isoc_ici

Eurostat. (2023b). *Security related problems experienced when using the internet*. https://ec.europa.eu/eurostat/databrowser/view/ISOC_CISCI_PB/default/table?lang=en&category=isoc.isoc_i.isoc_ci_sci

FBI. (2022). *FBI Springfield Warns of the Devastating Outcome of Elder Fraud — FBI*. https://www.fbi.gov/contact-us/field-offices/springfield/news/press-releases/fbi-springfield-warns-of-the-devastating-outcome-of-elder-fraud

ITU. (2021). *Internet use*. https://www.itu.int/itu-d/reports/statistics/2021/11/15/internet-use/

Kim, S., Yao, W., & Du, X. (2022). Exploring Older Adults' Adoption and Use of a Tablet Computer During COVID-19: Longitudinal Qualitative Study. *JMIR Aging*, *5*(1). https://doi.org/10.2196/32957

Mentis, H. M., Madjaroff, G., & Massey, A. K. (2019). Upside and downside risk in online security for older adults with mild cognitive impairment. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. https://doi.org/10.1145/3290605.3300573

Moyle, W., Jones, C., Murfield, J., & Liu, F. (2020). 'For me at 90, it's going to be difficult': feasibility of using iPad video-conferencing with older adults in long-term aged care. *Aging and Mental Health*, *24*(2), 349–352.

Petrie, H., & Darzentas, J. S. (2017). Older People's Use of Tablets and Smartphones: A Review of Research. In M. Antona & C. Stephanidis (Eds.), *Universal Access in Human--Computer Interaction. Design and Development Approaches and Methods* (pp. 85–104). Springer International Publishing.

Sciarretta, E., Ingrosso, A., Volpi, V., Opromolla, A., & Grimaldi, R. (2015). Elderly and Tablets: Considerations and Suggestions About the Design of Proper Applications. In J. Zhou & G. Salvendy (Eds.), *Human Aspects of IT for the Aged Population. Design for Aging* (pp. 509–518). Springer International Publishing.

Statista. (2023a). *Tablet penetration in Europe 2020, by country*. https://www.statista.com/forecasts/1169174/tablet-penetration-in-europe-by-country

Statista. (2023b). *Tablet penetration rate EU-5 countries 2012-2020*. https://www.statista.com/statistics/271001/penetration-rate-of-tablets/

Tsai, H. yi S., Shillair, R., Cotten, S. R., Winstead, V., & Yost, E. (2015). Getting Grandma Online: Are Tablets the Answer for Increasing Digital Inclusion for Older Adults in the U.S.? *Educational Gerontology*, *41*(10), 695–709.

United Nations. (2022). *World Population Prospects 2022 World Population Prospects 2022*. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf

United Nations Development Programme. (2017). *Ageing, Older Persons and the 2030 Agenda for Sustainable Development*. https://www.undp.org/sites/g/files/zskgke326/files/publications/UNDP_AARP_HelpAge_International_Ageing_Olderpersons and 2030 Agenda_revOct2018.pdf

Vaportzis, E., Clausen, M. G., & Gow, A. J. (2018). Older adults experiences of learning to use tablet computers: A mixed methods study. *Frontiers in Psychology*, *9*(SEP). https://doi.org/10.3389/fpsyg.2018.01631

Vaportzis, E., Giatsi Clausen, M., & Gow, A. J. (2017). Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. *Frontiers in Psychology*, *8*(OCT). https://doi.org/10.3389/fpsyg.2017.01687

Vasconcelos, A., Silva, P. A., Caseiro, J., Nunes, F., & Teixeira, L. F. (2012). Designing Tablet-Based Games for Seniors: The Example of CogniPlay, a Cognitive Gaming Platform. *Proceedings of the 4th International Conference on Fun and Games*, 1–10. https://doi.org/10.1145/2367616.2367617

Virtanen, S. M. (2017). Fear of Cybercrime in Europe: Examining the Effects of Victimization and Vulnerabilities. *Psychiatry, Psychology and Law*, *24*(3), 323–338.

WHO. (2022). *Ageing and health*. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

Wilson, S. A., Byrne, P., Rodgers, S. E., & Maden, M. (2022). A Systematic Review of Smartphone and Tablet Use by Older Adults With and Without Cognitive Impairment. *Innovation in Aging*, *6*(2), 1–19.

Self-Care Oriented Smartphone Apps for Type 2 Diabetes: A Comparative Analysis

Gosak Lucija
University of Maribor
Faculty of Health Sciences
Zitna ulica 15, Maribor, Slovenia
lucija.gosak2@um.si

Ooge Jeroen
KU Leuven
Department of Computer Science
Leuven
jeroen.ooge@kuleuven.be

Fijačko Nino
University of Maribor
Faculty of Health Sciences
Zitna ulica 15, Maribor, Slovenia
nino.fijacko@um.si

Kamenšek Janja
Lilimed d.o.o., specialist family clinic
Majšperk, Slovenia
janja177@gmail.com

Kocbek Primož, Debeljak Nika
University of Maribor
Faculty of Health Sciences
Zitna ulica 15, Maribor, Slovenia
primoz.kocbek@um.si,
nika.debeljak@gmail.com

Verbert Katrien
KU Leuven
Department of Computer Science
Leuven
katrien.verbert@kuleuven.be

Štiglic Gregor
University of Maribor
Faculty of Health Sciences, Faculty of Electrical Engineering and Computer Science
Zitna ulica 15, Maribor, Slovenia
University of Edinburgh
Usher Institute
Edinburgh, United Kingdom
gregor.stiglic@um.si

Abstract*. Type 2 diabetes can cause serious health complications due to sub-optimal blood glucose levels, so patients at risk need to monitor their lifestyle and improve self-care. Mobile apps for smartphones are the right tools for this. To provide patients with an evidence-based mobile app, we conducted a review of mobile stores. We used the AADE7 framework to assess the presence of self-care behavioural functions. uMARS scale was used to rate the overall quality of smartphone apps. Specifically, we identified 12 relevant apps and found that they included only 51 out of 88 possible behavioural features for self-care. In addition, most of the apps were rated as suboptimal in terms of inclusiveness, functionality, aesthetics and information. Based on the uMARS scoring scale, the average quality rating of the smartphone apps was 3.76. The highest average quality scores were given to forDiabetes: diabetes self-management app. Overall, we argue that diabetics can better control themselves with smartphone apps that implement enough relevant behavioural features for self-management.*

Keywords. mobile phone; mobile health; smartphone application; health behavior; application quality

1 Introduction

Diabetes is a slowly progressing chronic disease that affects millions worldwide and kills over one million people yearly, making it the ninth leading cause of death (Bratzke et al., 2015; Jacoby, 2019; Karimy et al., 2016). There exist two types of diabetes. *Type 1 diabetes mellitus* is caused by the autoimmune destruction of beta cells (Salsali & Nathan, 2006) and is the most common type of diabetes in children and adolescents (Dowling, 2021). In contrast, *Type 2 diabetes mellitus* (T2DM) is caused by resistance to insulin action, usually due to obesity and lack of insulin secretion. Concretely, T2DM disrupts blood glucose levels, which can lead to acute and macrovascular chronic complications (Nickerson & Dutta, 2012; Yau et al., 2012). Therefore, T2DM patients must constantly monitor their blood glucose levels to control them (Nyenwe et al., 2011; Shrivastavva et al., 2013) and focus on self-care to prevent complications.

However, high-quality self-care in the context of T2DM is not trivial and requires knowledge about the disease and its management, behavioral interventions such as diet and insulin adjustments, and emotional aspects such as self-efficacy (Bigelow & Freeland,

2017; Siguroardóttir, 2005). First, knowledge about T2DM is a form of health literacy, a *social determinant* of health that can be influenced by social policies (Braveman & Gottlieb, 2014) and that is important for people to control their health (Kendir & Breton, 2020). The Internet has significantly impacted health literacy by increasing access to online health information (Lee et al., 2021). Second, behavioral interventions require careful monitoring of blood glucose levels to avoid, identify, and treat hypo- and hyperglycemia, and prevent morbidity and mortality (Asgari & Nazari, 2019; Rezaei et al., 2019). Third, self-care adherence can be stimulated by actively involving patients in managing their condition (Izahar et al., 2017), thus improving their satisfaction, happiness, and overall quality of life (Amelia, 2018).

Mobile health (m-health) offers promising solutions for maintaining self-sufficiency in chronic diseases such as T2DM (Rossi & Bigi, 2017; Scott et al., 2020; Srivastava et al., 2015). By providing customized information, instructions, graphics, and reminders, smartphone apps can improve self-care behavior, health knowledge, and commitment (Asgari & Nazari, 2019; Gardsten et al., 2017; Adu et al., 2020; Bene et al., 2019; Kebede & Pischke, 2019). Furthermore, smartphone apps can simplify remote health management by providing tailored self-care recommendations and facilitating communication between patients and care providers (Liu et al., 2020).

Nevertheless, smartphone apps targeting T2DM patients vary in supported features, availability, required device type or operating systems, etc. (Zhang et al., 2020). So far, a clear overview, comparison, and evaluation of these smartphone apps are missing. Our systematic review fills this gap: we assess and compare apps in terms of supported self-care behavior functions and overall quality. Moreover, we point out shortcomings in currently available smartphone apps and provide suggestions for improvement.

This app review aimed at providing patients with type 2 diabetes with relevant and evidence-informed mobile applications for self-care and disease management.

2 Methods

We formulated our research goals with the PIO model (Polit & Beck, 2012): the *target population* (P) consisted of T2DM patients older than 18; the *intervention* (I) was to identify smartphone apps that aim to control T2DM and record measured glucose values, daily activity, dietary intake, and medications; and the *outcome* (O) was an assessment of apps regarding self-care behavior functions and quality. This section describes how we collected, selected, and assessed apps that target T2DM patients. Our assessment encompasses both supported self-care behavior functions and overall quality.

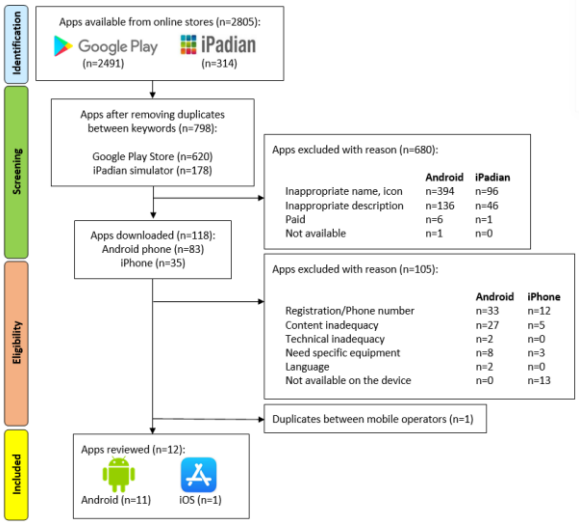


Figure 1. Flow diagram of our search process.

2.1 Search Strategy and Selection Process

Following the PRISMA guidelines (Moher et al., 2009), we systematically searched for smartphone apps in the Google Play Store (Android) and the iPadian simulator (iOS). In January 2021, the first two authors independently searched for apps with the keywords "diabetes" and "T2DM" and transferred the details of hits to Microsoft Excel. Then, they screened the name, icon, and description of collected apps based on preset inclusion criteria, consulting a third researcher in case of disagreement. The inclusion criteria were as follows: the used language was English, no registration or connection devices were required, no in-app purchases were necessary, and the app fostered T2DM self-care and allowed to enter the health measurements mentioned in our *intervention* criterium.

We only included mobile apps that are not paid for and do not require registration or integration, as we wanted to give all patients with type 2 diabetes the same options for use as much as possible. Using mobile apps requires technical skills such as setting up an account, logging into the mobile app, which is too complex for those who are less technically proficient (Campos-Castillo & Mayberry, 2022; Nelson et al., 2022). Participants who have a lower income or are less health literate are less likely to have a smartphone and use mobile apps (Nelson et al., 2022). Mobile health literacy has a direct impact on knowledge and use of mobile technology, which in turn affects patient health outcomes (Guo et al., 2021).

Only the Android version was included for apps available on both Android and iOS. Finally, we installed all included apps on a Huawei Pro 20 (Android 9.0) or an iPhone 7 (iOS 12.3.1).

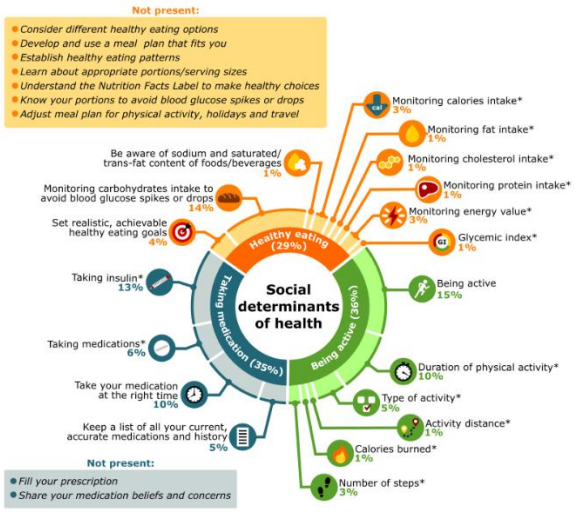


Figure 2. Present self-care behavior functions in *Awareness of social determinants of health*.

2.2 Assessing Self-Care Behavior Functions

The first two authors independently assessed the presence of self-care behavior functions according to the AADE7 framework (AADE, 2020; Burson & Mran, 2014; Eaton, 2020; Ye et al., 2018), which classifies self-care behavior functions into three main categories. First, **Awareness of social determinants of health** contains three subcategories: *healthy eating* (16 functions), *taking medication* (6 functions), and *being active* (6 functions). Second, **Integration of technology into self-care** encompasses *monitoring* (16 functions). Third, **Role of the diabetes care and education specialist** contains three subcategories: *problem-solving* (3 functions), *risk reduction* (31 functions), and *healthy coping* (10 functions). Found self-care behavior functions not yet part of the AADE7 framework were added to the most suitable category (marked with * in Figures 2 and 3).

2.3 Assessing Overall Quality

To assess the overall quality of smartphone apps, we used the uMARS scale (Stoyanov et al., 2016), which contains four constructs: *Engagement*, *Functionality*, *Aesthetics*, and *Information*. First, *Engagement* assesses whether apps are fun and interesting to use, interactive, and allow for customization. Then, *Functionality* evaluates the apps' performance, ease of use, navigation flow, and gestural design. Next, *Aesthetics* evaluates the layout, graphical form, and visual appeal. Finally, *Information* assesses whether the provided information is sufficient and of high quality, comes from a credible source, and is displayed visually. Two graduate nurses, experienced with managing chronically ill patients, used all apps for a few minutes and then rated the four uMARS constructs with scores ranging from 1 (inadequate) to 5 (excellent). The final uMARS quality score is the average of these four constructs, ignoring questions answered with N/A (not applicable).

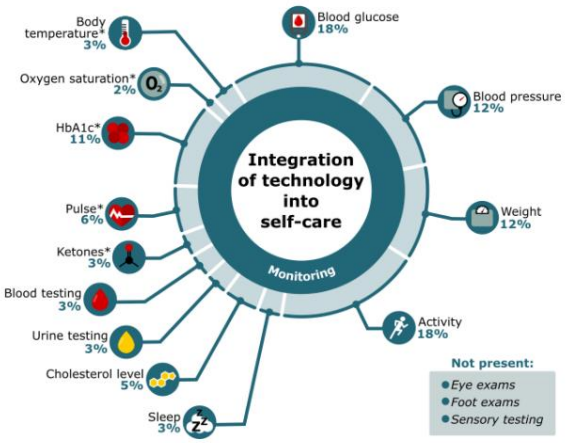


Figure 3. Present self-care behavior functions in *Integration of technology into self-care*.

3 Results

Figure 1 shows the PRISMA flow diagram of our search process. In total, 2805 smartphone apps matched our search queries. After removing duplicates and eliminating apps not complying with our inclusion criteria, 118 apps remained. After downloading all these apps on an Android or iOS smartphone, we removed 106 of them from our sample due to mandatory registration (e.g., email, phone number; n=45), substantive inadequacy (e.g., did not contain a component for tracking blood sugar, diet, activity, and medications; n=32), technical inadequacy (e.g., malfunction; n=2), need for additional technological devices (e.g., blood sugar meters that transfer data to smartphones via Bluetooth; n=11), non-English language (e.g., Spanish; n=2), and unavailability for the used device (e.g., only for iPad; n=13). Ultimately, our screening process resulted in 12 smartphone apps suitable for evaluation: four apps for both platforms (33%), one for iOS only (8%), and the remaining seven for Android only (58%). Four apps (33%) were classified under the *Health & Fitness* category in the mobile store, the others under *Medical*.

3.1 Self-Care Behavior Functions

Figures 2 to 4 visualize self-care behavior functions present in the reviewed smartphone apps, categorized according to the AADE7 framework.

Table 1 presents the distribution of the self-care behaviour functions into three main categories within their sub-categories.

Awareness of social determinants of health (see Figure 2) contains 28 unique self-care behavior functions, but we only found 19 of them (68%) in the reviewed apps. In total, apps supported only 78 out of

$12 \times 28 = 336$ functions, which is about 23%. The most common functions were *being active*, *monitoring carbo-hydrates intake*, and *taking insulin*: they were present in 12/12 (100%), 11/12 (92%), and 10/12 (83%) apps, respectively. Overall, *forDiabetes: diabetes self-management app* contained the widest range of self-care behavior functions.

Table 1. Self-care behavior functions

Awareness of social determinants of health	
Being active	28/78 functions
Taking medication	27/78 functions
Healthy eating	23/78 functions
Integration of technology into self-care	
	65/192 functions
Role of the diabetes care and education specialist	
Problem solving	12/148 functions
Risk reduction	101/148 functions
Healthy coping	35/148 functions

Integration of technology into self-care (see Figure 3) contains 16 unique self-care behavior functions and we recorded 13 of them (81%) in the reviewed apps. Yet, in total, apps supported only 65 out of the $12 \times 16 = 192$ functions; about 34%. All apps allowed to monitor blood glucose and activity, together accounting for 36% (24/65) of all recorded functions. Furthermore, 8/12 apps (67%) allowed to monitor both blood pressure and body weight, together accounting for 24% of all recorded functions. Finally, 7/12 apps (58%) supported monitoring HbA1c (i.e., glycosylated hemoglobin), accounting for 11% of all recorded functions. Two exemplar apps in this category were the *WeCheck -Diabetes lifelog, Blood Glucose Level* and the *Smart e-SMBG-Diabetes lifelog, Blood Glucose* apps, which allowed to monitor blood glucose levels, weight, body fat, HbA1c, feelings, mood, and hypoglycemia awareness.

Role of the diabetes care and education specialist (see Figure 4) includes 44 unique self-care behavior functions. However, we only found 18 of them (41%) in the reviewed apps, which supported 148 out of $12 \times 44 = 528$ functions (28%). In the *problem solving* subcategory, 2/3 self-care behavior functions (67%) were represented: 11/12 apps (92%) facilitated problem identification, but only 1 app (8%) encouraged patients to act. Furthermore, r*educing risks* represented 14/31 functions (45%) and all reviewed apps (12/12) monitored food and medication and advised to take prescribed medications and measure blood sugar. Finally, *healthy coping* represented 3/10 functions (30%): all smartphone apps (12/12) encouraged body movement and physical activity, and 11/12 apps (92%) allowed journalling. Overall, the *WeCheck-Diabetes lifelog, Blood Glucose Level* and the *Smart e-SMBG-Diabetes lifelog, Blood Glucose* apps supported the most self-care behavior functions.

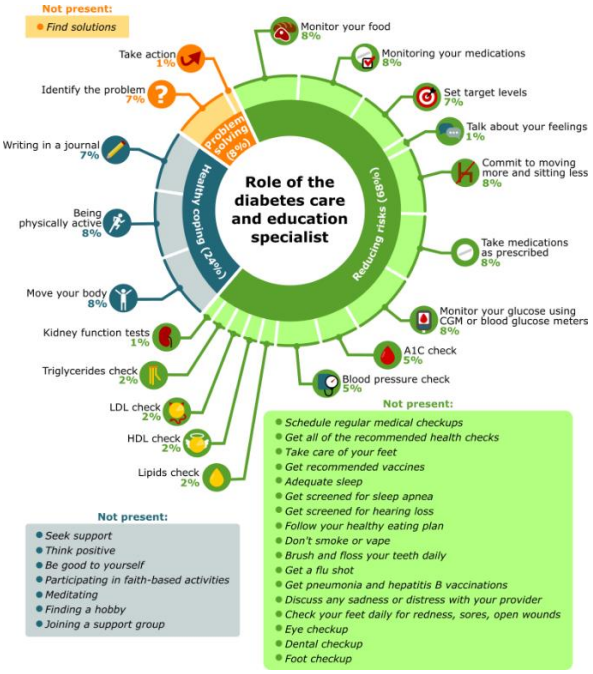


Figure 4. Present self-care behavior functions in *Role of the diabetes care and education specialist*.

3.2 Overall Quality

Based on the uMARS scoring scale, the average quality rating of the smartphone apps was 3.76/5 (SD 0.31). The highest average quality scores were given to *forDiabetes: diabetes self-management app* (4.54, SD 0.33) and *Diabetesdagboka* (4.09, SD 0.10). Overall, the lowest rated category was *Aesthetics*, scoring 3.68 on average (SD 0.47), and *Engagement* received the highest average rating (3.86, SD 0.34).

Regarding engagement, the most interesting and interactive smartphone app was *forDiabetes: diabetes self-management app* (4.5, SD 0.14), whereas Diabetes Journal (3.3, SD 0.42) got the lowest score. Regarding functionality, only 3 smartphone apps scored at least 4.00: *forDiabetes: diabetes self-management app*; *WeCheck -Diabetes lifelog, Blood Glucose Level; Smart e-SMBG -Diabetes lifelog, Blood Glucose*. The *SiDiary Diabetes smartphone app* was rated for the least functional application (3.50, SD 0.71*). SiDiary Diabetes* was considered the least aesthetically pleasing smartphone app (2.67, SD 0.47). *forDiabetes: diabetes self-management app* was rated with the highest score for information in the smartphone app (4.88, SD 0.18).

4 Discussion

Smartphone apps facilitate chronic disease management and thus show great potential for self-care education (Zhang et al., 2020), improving motivation (Fijačko et al., 2020) and reducing patient experience

(Lu et al., 2018), assisting with strict treatment requirements (Lee 2014), and reducing the cost of care (Lewinski, et al., 2019). However, a lack of knowledge about their actual use (Trawley et al., 2017), insufficient evidence of effectiveness in changing behavior (Milne-Ives et al., 2020; Zhao et al., 2016), and insufficient evidence of clinical precision limit their effectiveness (Fleming et al., 2020).

We systematically assessed smartphone apps for T2DM patients regarding supported self-care behavior functions and overall quality. Overall, we found 12 useful apps for T2DM self-care with above-average quality but lacking self-care behavior functions. They included only 51 of the 88 possible self-care behavior functions. The least well represented are the functions that focus on integrating the role of the expert in the care and education of patients with diabetes.

App name	Engagement	Functionality	Aesthetics	Information	Average
forDiabetes: diabetes self-management app	4.50 (0.14)	4.13 (0.53)	4.67 (0.47)	4.88 (0.18)	4.54 (0.33)
Diabetesdagboka	4.20 (0.28)	3.88 (0.18)	4.17 (0.24)	4.13 (0.18)	4.09 (0.1)
Diaguard: Diabetes Tagebuch	4.20 (0.28)	3.88 (0.53)	4.17 (0.24)	3.38 (0.18)	3.90 (0.10)
Diabetes:M - Management & Blood Sugar Tracker App	3.90 (0.71)	3.88 (0.53)	3.50 (0.71)	4.13 (0.53)	3.85 (0.62)
WeCheck -Diabetes lifelog, Blood Glucose Level	3.80 (0.28)	4.13 (0.18)	3.50 (0.71)	3.63 (0.18)	3.76 (0.25)
Smart e-SMBG -Diabetes lifelog, Blood Glucose	3.80 (0.28)	4.13 (0.18)	3.50 (0.71)	3.63 (0.18)	3.76 (0.25)
MEET ME @7 - Diabetes Self-management Tool for Patients and Caregivers	3.90 (0.71)	3.88 (0.18)	3.67 (0.47)	3.50 (0.35)	3.74 (0.43)
SmartLog	4.10 (0.42)	3.63 (0.18)	3.50 (0.71)	3.38 (0.53)	3.65 (0.46)
BG Monitor Diabetes	3.60 (0.57)	3.38 (0.18)	3.83 (0.24)	3.38 (0.53)	3.55 (0.29)
Betes - Your Diabetes Diary	3.50 (0.71)	3.75 (0.35)	3.50 (0.71)	3.38 (0.53)	3.53 (0.57)
Diabetes Journal	3.30 (0.42)	3.88 (0.18)	3.50 (0.71)	3.25 (0.71)	3.48 (0.42)
SiDiary Diabetes Management	3.50 (0.14)	3.50 (0.71)	2.67 (0.47)	3.63 (0.18)	3.32 (0.07)
Overall mean (SD)	*3.86 (0.34)*	*3.83 (0.23)*	*3.68 (0.47)*	*3.69 (0.45)*	*3.76 (0.31)*

Figure 5. uMARS mean scores for all included smartphone apps. Column-wise maximal and minimal values are highlighted in green and red, respectively.

4.1 Integrating More Diverse Self-Care Behavior Functions in Apps

We discuss the three AAED7 subcategories of self-care behavior functions separately.

4.1.1 Awareness of Social Determinants of Health

Of all self-care behavior functions, existing apps best support the awareness of social determinants of health. This is a positive finding because physical activity and proper diets are essential to prevent complications due to T2DM (Anjana & Mhan, 2016; Kim et al., 2022).

All smartphone apps allow recording physical activity, for example, activity type and duration. However, only one app shows an estimate of the number of calories consumed during physical activity. This is because the apps we reviewed are not sport tracking apps but are instead tailored towards T2DM self-care. Most apps allow monitoring carbohydrate intake to avoid fluctuations in blood glucose and thus hyper- or hypoglycemia. Yet, no existing apps monitor meal size or explain dietary labels. Adding such functionalities would make it easier for T2DM patients to compose a proper diet and encourage them to eat healthier (Christophet et al., 2017).

Finally, most apps allow to monitor insulin intake, but none consider carbohydrate intake or blood glucose measurements. Yet, more fine-grained monitoring is essential to avoid medication errors: by tracking nutrient intake during the day, T2DM patients can better determine the required insulin dose, thus avoiding hyper- or hypoglycaemia. Insulin dose calculators can further avoid medication errors by indicating the number of required insulin units.

Studies have shown that apps designed to monitor physical activity can benefit diabetic patients too (Kordonouri & Riddell, 2019). We therefore recommend that app developers implement calorie counting as this is useful for determining optimal nutrient intake (Scarry et al., 2022).

4.1.2 Integration of Technology into Self-Care

The most common measurement inputs were blood pressure, body weight, and HbA1c. In a similar study, researchers found that smartphone apps focused primarily on monitoring body weight, blood glucose, HbA1c, and the units and types of insulin used (Ersotelos et al., 2018).

Although many apps' features are aimed at monitoring and controlling T2DM, they do not include tests for early detection of the risk of developing chronic complications such as neuropathy (Jamin et al., 2021), ocular retinopathy (Khura et al., 2021), and diabetic foot (Al-Rubeaan et al., 2015; Ploderer et al., 2018; Wang & Lo, 2018). Furthermore, none of the apps we analyzed monitored important indicators of complications (e.g., eye exams, foot exams, sensory testing). Early detection could reduce chronic diabetes complications and associated high costs (Chawla et al., 2020; Farshchi et al., 2014). We believe smartphone apps could help detect risks of developing complications more quickly, thus reducing costs and improving treatment outcomes and patients' quality of life. Therefore, we recommend that app developers include estimated risks of complication onset, thus contributing to early screening and preventing health deterioration.

4.1.3 Role of the Diabetes Care and Education Specialist

Our results represent a higher number of self-care behavior functions regarding *problem solving*, *risk reduction*, and *healthy coping*. We speculate that smartphone apps in the past were more focused on T2DM and were not patient-oriented because patient-centered care is still an evolving concept (Fix et al., 2018; Santana et al., 2018). Therefore, app designers have not yet integrated it into the mobile environment

to such an extent. Patient-centered smartphone apps would place more emphasis on improving patient health and well-being and preventing complications rather than diagnosis. Nevertheless, there is still a lack of risk prevention regarding T2DM. Smartphone apps could provide tips for a healthy lifestyle, such as taking care of proper oral hygiene, avoiding cigarette smoke, etc. There is evidence that smartphone apps can effectively influence lifestyle changes in T2DM (Wu et al., 2019).

4.2 Improving the Overall Quality of Apps

Unfortunately, the overall quality of smartphone apps in our sample is rather average. We propose several recommendations about aesthetics, information amount, and engagement to improve apps' quality.

First, it is concerning that apps generally scored low in terms of aesthetics, because apps' likeability depends on their visual design (Trapp & Wienrich, 2018). One potential solution is to display information more visually, as this can increase motivation for self-management (McMillan et al., 2017). In this regard, *forDiabetes: diabetes self-management app* is a great example: its layout is simple and clear, its graphics are of high quality, and its visuals are attractive and memorable.

Second, the quality and quantity of information provided in apps were deemed moderately relevant and comprehensive. To further improve this aspect, apps could inform patients with more evidence-based information on T2DM; for example, symptoms, risk factors, prevention, and coping strategies. This additional information could, for example, be provided as personalized advice or links to relevant health articles. In all cases, simple language should be used to ensure patients understand the contents, and information should be tailored to patients' needs and knowledge (Torbjørnsen et al., 2019).

Third, we found that apps were deemed rather interesting and useful for patients, facilitating adjusted settings. However, with an average score of 3.86 out of 5, there is still room for improvement. To make apps more engaging, we propose integrating motivational techniques into smartphone apps. For example, gamification can improve user engagement and motivation by incorporating game-design elements (e.g. leaderboards, badges, and points) into non-gaming contexts (King et al., 2013). We found that only the *Diabetesdagboka* app engages users to achieve self-chosen goals with gamification (e.g., rewarding regular exercise with trophies).

4.3 Smartphone Apps and the Future of m-Health

Finally, we want to take a step back and briefly discuss our research in the light of m-health.

Information and communication technology are increasingly being integrated into healthcare, requiring a higher health literacy and a more active role of patients (Kim & Xie, 2015; Santana et al., 2021). This shows the urgency of clearly providing the right amount of information to patients. Whereas the smartphone apps we reviewed target adults diagnosed with T2DM, children might require less complex apps. We believe this is particularly important as children too are an important target group for future m-health applications. For example, smartphone apps could help motivate young diabetic patients to efficiently monitor their physical activity, diet, and blood sugar (Majeed-Ariss et al., 2015; Paramanik et al., 2019).

In general, we believe m-health holds a lot of potential for monitoring lifestyle and improving self-care, which are both especially relevant in the context of T2DM. However, given our findings, we agree with researchers such as Veazie et al. (2018) that there is still much room for improving existing T2DM self-care smartphone apps, for example by supporting better collaboration between doctors, patients, and researchers. At the same time, as m-health is increasingly integrated into patients' daily lives, it is necessary to promote awareness of how smartphone apps can benefit self-care and what its potential risks are (Gong et al., 2020).

4.3 Limitations

Our study has several limitations. First, we assigned several self-care behavior functions to AAED7 categories ourselves, which hampers comparison with other work. At the same time, this contributes to the field, as we have identified some new functions. Second, although we relied on clear definitions for the different self-care behavior functions and involved two graduate nurses in our study, assessing apps remains subjective. Other health experts may assess apps differently. Third, we restricted our review to free smartphone apps that do not require a username or login for regular operation. Doing so, we may have missed interesting apps. Fourth, we only included smartphone apps for iOS and Android, the most prominent smartphone operating systems (Götz et al., 2017). Fifth, we carried out our review in 2021, so we allow for the possibility that new apps for diabetes self-management have been developed since then.

5 Conclusion

Although m-health for maintaining T2DM is actively being developed, our systematic analysis shows that the wide range of available smartphone apps still has shortcomings. Addressing these according to our observations and recommendations would help patients to manage their condition better and reduce complications. In sum, we encourage app developers to pay more attention to educating users about the dangers of life-threatening and chronic complications. Furthermore, to motivate patients to monitor their

condition closely and adopt a healthy lifestyle, we also see an opportunity for integrating motivational techniques and visual data representations in apps. In this way, we believe that smartphone apps can contribute to improving patients' health and preventing health deterioration.

Acknowledgments

The authors acknowledge partial support from the Slovenian Research Agency (grant numbers N3-0307 and P2-0057). The project "Digital Toolbox for Innovation in Nursing Education (I-BOX)" has been funded with support from the European Commission (2019-1-ES01-KA203-065836) under the Erasmus+ program. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein. The authors acknowledge partial support from the Slovenian Research Agency (grant numbers N3-0307 and P2-0057).

References

Adu, M. D., et al. (2020). The development of My Care Hub mobile-phone app to support self-management in Australians with type 1 or type 2 diabetes. *Scientific Reports*, 10(1), 1-10.

Al-Rubeaan, K., et al. (2015). Diabetic foot complications and their risk factors from a large retrospective cohort study. *PloS one*, 10(5), e0124446.

Anjana, R. M. & Mohan, V. (2016). Diabetes and physical activity. *The Indian Journal of Medical Research*, 143(4), 530.

Amelia, R. (2018). The model of self-care behavior and the relationship with quality of life, metabolic control and lipid control of type 2 diabetes mellitus patients in Binjai City, Indonesia. *Open access Maced J Med Sci*, 6(9), 1762–1767.

American Association of Diabetes Educators. (2020). An effective model of diabetes care and education: Revising the AADE7 self-care behaviors®. *The Diabetes Educator*, 46(2), 139–160.

Asgari, P. & Nazari, E. (2019). Review of the effectiveness of Information and Communication Technology (ICTs) in the treatment of type 1 and type 2 diabetes. *Front Heal Informatics*, 8(1), 12.

Bansal, V., et al. (2006). Diabetic neuropathy. *Postgraduate medical journal*, 82(964), 95-100.

Bene, B. A., et al. (2019). Impact of mobile health applications on self-management in patients with type 2 diabetes mellitus : protocol of a systematic review. *BMJ Open*, 9(6), e025714.

Bigelow, A. & Freeland, B. (2017). Type 2 diabetes care in the elderly. *The Journal for Nurse Practitioners*,13(3), 181–186.

Bratzke, L. C., et al. (2015). Self-management priority setting and decision-making in adults with multimorbidity: A narrative review of the literature. *International Journal of Nursing Studies*, 52(3), 744–755.

Braveman, P. & Gottlieb, L. (2014). The social determinants of health: it's time to consider the causes of the causes. *Public health reports*, 129(1_suppl2), 19-31.

Burson, R. & Moran, K. J. (2014). The AADE 7. *Home Healthcare Now*, 32(9), 556.

Campos-Castillo, C., & Mayberry, L. S. (2022). Disparities in digital health in underserved populations. In Diabetes Digital Health and Telehealth (pp. 269-280). Academic Press.

Chawla, R. et al. (2020). RSSDI-ESI clinical practice recommendations for the management of type 2 diabetes mellitus 2020. *Indian journal of endocrinology and metabolism*, 24(1).

Christoph, M. J., et al. (2017). Nutrition Facts: who is using them, what are they using, and how does it relate to dietary intake?. *J Acad Nutr Diet*.

Dowling, L. (2021). Effective management of type 1 diabetes in children and young people. Nursing Children and Young People, 33(4).

Eaton, M. (2020). *AADE7 Self-Care Behaviors Revised in New Guidance from the Association of Diabetes Care & Education Specialists.*

Ersotelos, N. T., et al. (2018). Review of mobile applications for optimizing the follow-up care of patients with diabetes. *Hormones*, 17(4), 541-550.

Fijačko, N., Gosak, L., Cilar, L., Novšak, A., Creber, R. M., Skok, P., & Štiglic, G. (2020). The effects of gamification and oral self-care on oral hygiene in children: systematic search in app stores and evaluation of apps. *JMIR mHealth and uHealth*, 8(7), e16365.

Fix, G. M., et al. (2018). Patient-centred care is a way of doing things: How healthcare employees conceptualize patient-centred care. *Health Expectations*, 21(1), 300-7.

Fleming, G. A., Petrie, J. R., Bergenstal, R. M., Holl, R. W., Peters, A. L., & Heinemann, L. (2020). Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) *Diabetes Technology Working Group. Diabetes care*, 43(1), 250-260.

Gardsten, C., et al. (2017). Designing an ICT self-management service: suggestions from persons with type 2 diabetes. *Health and Technology*, 7(2), 197–206.

Gong, E., et al. (2020). Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. *JMIR mHealth and uHealth*, 8(4), e14836.

Götz, F. M., et al. (2017). Users of the main smartphone operating systems (iOS, Android) differ only a little in personality. *PloS one*, 12(5), e0176921.

Guo, S. H. M., et al. (2021). Relationships between mobile eHealth literacy, diabetes self-care, and glycemic outcomes in Taiwanese patients with type 2 diabetes: cross-sectional study. JMIR mHealth and uHealth, 9(2), e18404.

Huckvale, K., et al. (2015). Smartphone apps for calculating insulin dose: a systematic assessment. *BMC medicine*, 13(1), 1-10.

Izahar, S., et al. (2017). Content Analysis of Mobile Health Applications on Diabetes Mellitus. *Front Endocrinol (Lausanne)*, 8.

Jacoby, J. (2019). Systematic Review of Mobile Phone Apps Currently Available to Norwegian Users to Support Diabetes Self-management. *International Conference on Human-Computer Interaction*, 457–466.

Jasmin, M., et al. (2021). Validity and reliability of a vibration-based cell phone in detecting peripheral neuropathy among patients with a risk of diabetic foot ulcer. *The International Journal of Lower Extremity Wounds*, 15347346211037411.

Karimy, M., et al. (2016). Determinants of adherence to self-care behavior among women with type 2 diabetes: An explanation based on health belief model. *Med J Islam Repub Iran*, 30, 368.

Khurana, R.N. et al. (2021). A smart mobile application to monitor visual function in diabetic retinopathy and age-related macular degeneration: the CLEAR study. *American Journal of Ophthalmology*, 227, 222-230.

King, D., Greaves, F., Exeter, C., & Darzi, A. (2013). 'Gamification': Influencing health behaviours with games. *Journal of the Royal Society of Medicine*, 106(3), 76-78.

Kim, H.J. et al. (2022). The effect of a mobile and wearable device intervention on increased physical activity to prevent metabolic syndrome: observational study. *JMIR mHealth and uHealth*, 10(2), e34059.

Kordonouri, O. & Riddell, M. C. (2019). Use of apps for physical activity in type 1 diabetes: current status and requirements for future development. *Therapeutic advances in endocrinology and metabolism*, 10, 2042018819839298.

Kebede, M. M. & Pischke, C. R. (2019). Popular diabetes apps and the impact of diabetes app use on self-care behavior: a survey among the digital community of persons with diabetes on social media. *Frontiers in Endocrinology*, 10, 135.

Kendir, C. & Breton, E. (2020). Health literacy: from a property of individuals to one of communities. *International journal of environmental research and public health*, 17(5), 1601.

Kessler, H. & Wunderlich, S. M. (1999). Relationship between use of food labels and nutrition knowledge of people with diabetes. *Diabetes Educ*, 25(4), 549-559.

Kim, H. & Xie, B. (2015). Health literacy and internet-and mobile app-based health services: A systematic review of the literature. *Proceedings of the Association for Information Science and Technology*, 52(1),1-4.

Lee, H.Y., et al. (2021). Role of health literacy in health-related information-seeking behavior online: Cross-sectional study. *J Med Internet Res*, 23(1), e14088.

Lewinski, A. A., Patel, U. D., Diamantidis, C. J., Oakes, M., Baloch, K., Crowley, M. J., ... & Bosworth, H. B. (2019). Addressing diabetes and poorly controlled hypertension: pragmatic mHealth self-management intervention. *Journal of medical Internet research*, 21(4), e12541.

Liu, K., et al. (2020). Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. *JMIR mHealth and uHealth*, 8(8), 15779.

Lu, C., Hu, Y., Xie, J., Fu, Q., Leigh, I., Governor, S., & Wang, G. (2018). The use of mobile health applications to improve patient experience: cross-sectional study in Chinese public hospitals. *JMIR mHealth and uHealth*, 6(5), e9145.

McMillan, K. A., et al. (2017). A systematic and integrated review of Mobile-based technology to promote active lifestyles in people with type 2 diabetes. *J Diabetes Sci Technol*, 11(2), 299–307.

Majeed-Ariss, R. et al., 2015. Apps and adolescents: a systematic review of adolescents use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. *Journal of Medical Internet Research*, 17(12), 1–16.

Milne-Ives, M.. et al. (2020). Mobile apps for health behavior change in physical activity, diet, drug

and alcohol use, and mental health: systematic review. *JMIR mHealth and uHealth*, 8(3), e17046.

Moher, D., et al. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7).

Nelson, L. A., et al. (2022). Disparities in mobile phone use among adults with type 2 diabetes participating in clinical trials 2017–2021. JAMIA open, 5(4), ooac095.

Nickerson, H.D., Dutta, S. (2012). Diabetic complications: Current challenges and opportunities. *Journal of cardiovascular translational research*, 5, 375–379.

Nyenwe, E. A., et al. (2011). Management of type 2 diabetes: evolving strategies for the treatment of patients with type 2 diabetes. *Metabolism*, 60(1), 1-23.

Ploderer, B. et al. (2018). Promoting self-care of diabetic foot ulcers through a mobile phone app: user-centered design and evaluation. *JMIR diabetes*, 3(4), e10105.

Polit, D. & Beck, C. *Nursing research: Generating and assessing evidence for nursing practice*. Lippincott Williams & Wilkins, editor (2012).

Pramanik, B. K. et al. (2019). Smartphone app as motivational intervention to improve glycemic control in adolescents with type 1 diabetes. *Indian Journal of Pediatrics*, 86(12), 1118–1123.

Rezaei M, Valiee S, Tahan M, et al. Barriers of medication adherence in patients with type-2 diabetes: a pilot qualitative study. Diabetes Metab Syndr Obes; 2019;12:589. PMID:31118722 doi:https://doi.org/10.2147/DMSO.S197159

Rossi, M. G. & Bigi, S. (2017). mHealth for diabetes support: a systematic review of apps available on the Italian market. *mHealth*, 3, 16.

Salsali, A., & Nathan, M. (2006). A review of types 1 and 2 diabetes mellitus and their treatment with insulin. American journal of therapeutics, 13(4), 349-361.

Santana, M. J., et al. (2018) How to practice person-centred care: A conceptual framework. *Health Expectations*, 21(2), 429-40.

Santana, S. et al. (2021). Practice full report: Updating health literacy for healthy people 2030: Defining its importance for a new decade in public health. *Journal of Public Health Management and Practice*, 27(6), S258.

Scarry, A., Rice, J., O'Connor, E. M., & Tierney, A. C. (2022). Usage of Mobile Applications or Mobile Health Technology to Improve Diet Quality in Adults. *Nutrients*, 14(12), 2437.

Scott, I. A., et al. (2020). Going digital: A narrative overview of the effects, quality, and utility of mobile apps in chronic disease self-management. *Australian Health Review*, 44(1), 62-82.

Shrivastava, S. R., Shrivastava, P. S. & Ramasamy J. (2013). Role of self-care in management of diabetes mellitus. *J Diabetes Metab Disord*, 12(1), 14.

Siguroardóttir, Á. K. (2005). Self-care in diabetes: Model of factors affecting self-care. *J Clin Nurs*, 14(3), 301–314.

Srivastava, P. K., et al. (2015). Role of Ayurveda in management of diabetes mellitus. *Internal Research Journal of Pharmacy*, 6(1), 8–9.

Stoyanov, S., et al. (2016). Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR mHealth and uHealth*, 4(2), e72.

Torbjørnsen, A., et al. (2019). Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. *BMC health services research*, 19(1), 1-14.

Trapp, A. K. & Wienrich, C. (2018). App icon similarity and its impact on visual search efficiency on mobile touch devices. *Cognitive research: principles and implications*, 3(1), 1-21.

Trawley, S., Baptista, S., Browne, J. L., Pouwer, F., & Speight, J. (2017). The use of mobile applications among adults with type 1 and type 2 diabetes: results from the second MILES—Australia (MILES-2) study. *Diabetes technology & therapeutics*, 19(12), 730-738.

Zhang, Y., et al. (2020). Effectiveness of Lilly Connected Care Program (LCCP) app-based diabetes education for patients with type 2 diabetes treated with insulin: Retrospective real-world study. *JMIR mHealth and uHealth*, 8(3), 1–12.

Zhao, J., Freeman, B., & Li, M. (2016). Can mobile phone apps influence people's health behavior change? An evidence review. *Journal of medical Internet research*, 18(11), e287.

Yau, J. W. Y, et al. (2012). Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*, 35(3), 556–564.

Ye, Q., et al. (2018). An analysis of diabetes mobile applications features compared to AADE7TM: Addressing self-management behaviors in people with diabetes. *Journal of diabetes science and technology*, 12(4), 808–816.

Veazie, S., et al. (2018). Rapid evidence review of mobile applications for self-management of diabetes. *J Gen Intern Med*, 33(7), 1167–1176.

Wang, W. & Lo, A. C. (2018). Diabetic retinopathy: pathophysiology and treatments. *International journal of molecular sciences*, 19(6), 1816.

Wu, X., et al. (2019). The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. *JMIR mHealth and uHealth*, 7(1), e12297.

Digital Human Resource Management - Designing a digital platform for educational institutions' HR process

Katarina Tomičić-Pupek, Martina Tomičić Furjan, Igor Pihir
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{katarina.tomicic, martina.tomicic, igor.pihir}@foi.unizg.hr

Abstract. *Human resource management (HRM) systems have been of interest in research for many years, with a slightly difference in frequency in two waves. With technology development new opportunities to harvest value from introducing digital platforms have arisen. This paper addresses the role of digital HRM platforms by presenting a case study of designing a conceptual architecture model of an HR digital platform for educational institutions, more specifically for supporting the process of assessment and promotion of teachers in primary and secondary schools in Croatia. The case study describes starting value propositions for main actors, platform's design criteria, a conceptual model of the platform, a wireframe model illustrating technological extensions, all oriented for delivering the set value propositions. The case study shows that an integrated HR platform designed in such a manner, would be able to meet the value proposition of supporting the described process, and provide better collaboration options of involved stakeholders.*

Keywords. HRM, platforms, promotion, schools

1 Introduction

Human Resource Management encompasses different perspectives of managing employees and their role in an organization. It is based on data covering: a) administration (i.e. personal data, contract data, demographic data, academic data, previous work experience and similar), b) interrelations (like social interaction, team working, collaboration, and interpersonal relationships (Skinner, 1981)), c) stages of a work-lifecycle (like recruitment, selecting, training, onboarding, promotion, rewarding and assessment, orientation, (re)placement (Sukawati et al., 2020), retirement, resignation), d) skills and competencies (from vocational, professional, technology-related, industry-related, digital), e) performance and operational activities related to specific work positions. Basic classification of the HR data is into master data collected and used as common registries in integrated environments, and transactional data capturing interactions and performance.

With technology development, existing traditional use of HRM systems is offering new design features for supporting specific organizational needs in HR management. Some examples of applying technology in HRM systems can be found in applying IoT and cloud computing (Chai, 2022), Big data technology in human resource management for supporting decision making for employee-job matching (Chen & Du, 2022) or based on fuzzy data mining for different stages of HR Management (Peng, 2022), and they show that this topic gains value as more data on employees is available.

Social media platforms are another example of extending the insights into traditional HRM, introducing professional social media (PSM), and emergent themes, and subsequently, related research questions (Ruparel et al., 2020).

With social disruptions and developments other challenges arise as well. The phenomena of short-term and temporary positions, independent workers, flexibility in time and space management of jobs, gig economy, remote work, digital nomads and other, offer new paradigms of managing HR. As a response to these challenges, platform-based tools are being introduced.

HRM activities in platform ecosystems enable new roles and contributions of key actors in ensuring value co-creation (Meijerink & Keegan, 2019), aligning thereby multilateral exchanges of three key gig economy actors: gig workers, requesters, and intermediary platform firms, for ensuring value co-creation. Although there is a need to criticality examine the role of HRM (Kaufman, 2020) or platforms (especially management-by-algorithm in gig economy forms of introducing platforms (Duggan et al., 2020)) there is still space for improving existing HRM systems.

Human resource management systems created either as simple applications or as complex platforms should support organizations and employees to actively participate in achieving set goals and improve

the quality of services as an opportunity to advance, both collectively and individually. In many cases teachers and other employees in education are missing a proper digital platform which would support them in tracking their achievements in their Carrier Lifecycle. This impacts their administrative workload heavily. Generic design criteria for digital platforms can be translated into specific functional requirements addressing roles and needs of educational institutions offering to tailor custom solutions and providing thereby new value for "customers".

A digital platform that is custom designed for creating value by minimizing the administrative burden and that is strongly in compliance with regulatory frameworks, should enable individual user experience paths, provide collaboration with other stakeholders, aim to assist the personal development as well as the development of communities within the ecosystem. The latter is the motivation for this research, namely investigating the role of platforms in supporting HRM in schools, as a part of a wider project of e-schools in Croatia.

2 Research insights about publications related to HRM and digital platforms

For the purpose of secondary data collection, desk research was conducted during April 2023, starting with the identification of publications in two relevant databases - Web of Science and Scopus. The literature review was performed based on a search using the combination of keywords "human", "resource" and "platform" in the publication titles. The search resulted in 63 results in Scopus and 33 results in Web of Science (i.e. 96 in total), whereby 27 papers were covered by both databases, leaving 69 different papers for further analysis (listed in the Appendix I). Analysis was performed on those 69 papers, first by year, document type and source (as shown in fig. 1). This basic analysis was performed to identify main trends in this field, and the results revealed that until 2018, 31 papers were published, followed by a significant drop in 2019, and then from 2019 to 2023 additional 38 publications arose. This second set indicates that there were two waves, which can be a result of rethinking HRM in disruptive environments challenged by Covid-19 pandemic.

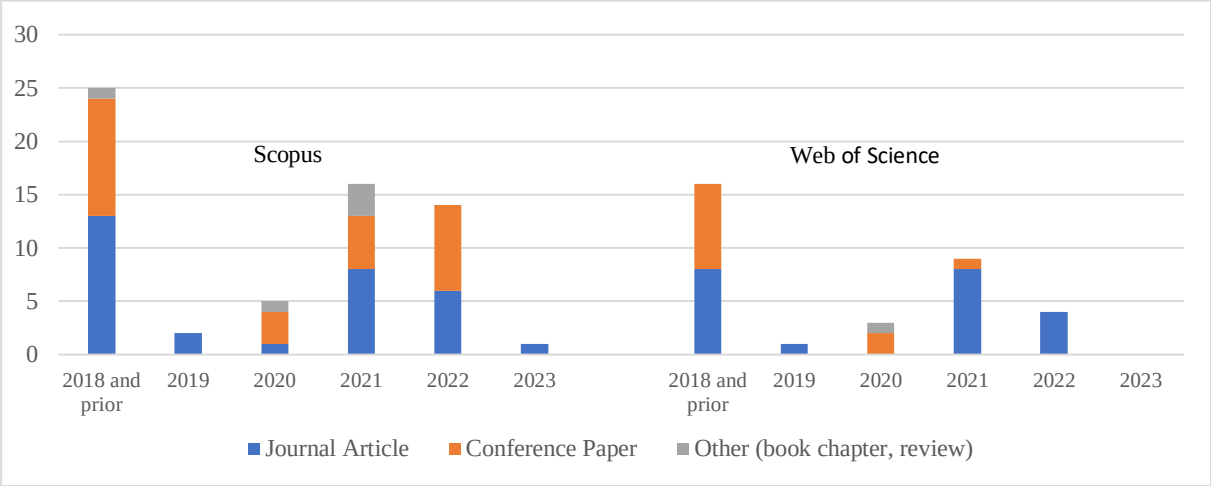


Figure 1. Publications by year, type and source

To explore the difference more in depth between these two waves, text-mining techniques were applied, for a more qualitative analysis. For the text-mining R Studio tool was used, for two runs: for papers published before and in 2018, and for papers from 2019 until 2023. The text-mining was performed on publications' abstracts, with application of cleaning methods, both manual and automatic. Although this research is not exhaustive, and it did not include whole publication texts, some results seem to be interesting for gaining some insights. A full list of words with appearance frequency of 10 and more for both waves is given in table 1, and respective word clouds in figures 2 and 3.

The results indicate that the pool of words doubled in the second wave (when exploring only words appearing more than 10 times). This could indicate a rise of HRM complexity, since 2019. Some words were appearing more often: human 85→128, resource 85→140, platform 49→65, management 41→88, data 19→71, technology 13→23, design 7→17.

Table 1. Words by appearance

Publications till and in 2018				Publications from 2019	
human	85	resource	140	enterprise	16
resource	80	human	128	teaching	16
platform	49	management	88	application	13
management	41	data	71	SaaS	13
information	29	platform	65	service	13
development	24	HRM	28	support	13
data	19	decision	27	traditional	13
service	19	development	27	computing	12
system	18	information	23	model	12
technology	13	system	23	process	12
cloud	12	technology	23	labor	11
evaluation	11	analysis	21	method	11
e-learning	10	research	20	algorithm	10
network	10	results	18	economy	10
research	10	design	17	online	10
science	10	network	17	processing	10
		performance	17	technologies	10
		activities	16		

The results also indicate that some new terms are entering the HRM arena (like complexity 0→10, SaaS (software as a Service) 0→13.), and some more emphasis is put on the management of resources, rather than the human aspect. A word which did not meet the threshold of 10 and more appearances is "digital", moving from not appearing till 2018, to 9 times in the second wave.

Figure 2. Wordcloud from publications till 2018 [Author's work, using R studio]

Figure 3. Wordcloud from publications from 2019 [Author's work, using R studio]

Based on the insights gained through the text mining, and previous research in this field, further sections of this paper are addressing the design of a HRM platform which aims supporting one important stage of HRM, the promotion stage. This specific stage is selected because in involves many actors within the process of assessing teachers and their efforts in terms of meeting general conditions and excellence conditions related to organizing student competitions, attending educational programs, participating in professional councils and

associations, publishing teaching materials and professional papers, working on projects, improving school work and elements of the education system. For that purpose, first a conceptual architecture model of an HR digital platform for schools is presented, offering to describe design criteria based on specific factors of decision-making for using a digital platform to run the process of assessment and promotion, by relying on existing research on perspectives of joining, adopting, or implementing a digital platform. Next, to illustrate how such a platform could include certain technological elements, a wireframe illustrates how this platform could be realized in order to integrate other existing HR-related systems into a single operating platform. At the end of the paper, a discussion and concluding remarks are given.

3 A case study of designing a conceptual architecture of a digital platform for schools

A general conceptual model implying convergent or divergent relationships between various stakeholders, influencing thereby their perceived balance of value or risk and the overall decision on joining, adopting, or implementing a digital platform (Kutnjak et al., 2022) is the basis for designing the Conceptual architecture of an HR digital platform (DP) for schools. It encompasses four actors, divided into 3 roles: DP provider, DP customer, DP producer. DP provider for schools in Croatia is CARNET, a national agency for supporting schools and other academic institutions with ICT related products and services. DP customers (customers in a sense of using the platform to get assessed) are teachers in schools, performing their daily activities and consequently generating inputs for their assessment within the promotion process i.e. in their professional advancement and promotion related to their work excellence. The role of DP producers (producers of the output of the assessment and promotion process) is primarily intended for the Education and Teacher Training Agency, responsible for the assessment process in teacher promotions, but also school management involved in administrative activities within that process. Starting point for a platform design is shaping the Value proposition for stakeholders and future platform actors. The value proposition for DP customers consists of following elements:

- Having everything in place, demanding less time and effort to prepare for assessments ran in cycles or on demand;
- More experience exchange opportunities with peers within communities related to usefulness on offered activities (educational and training programs, association events, projects) delivered by DP producers and other stakeholders;
- More convenient access to data needed for planning personal achievements and more efficient options for tracking progress of achieving them.

The value proposition for DP producers consists of following elements:

- More detailed overview and insights of delivered educational and training programs, association events, projects;
- More detailed data available for planning future policies, events and activities;
- Opportunity to provide assistance and support not only at the end of an assessment cycle, but also continuously.

The value proposition for DP provider is to generate design criteria detailed enough to enable them to develop and implement a new HRP digital platform. In order to enable all actors to gain their value proposition and claim the, for them intended roles, design criteria for a future HR digital platform were designed. The design criteria have to meet specific factors of decision-making for using the platform to run the process of assessment and promotion, which is explained in the following sections. From the DP provider's perspective, factors for implementing the platform are translated into following design criteria:

- *Resource dependencies & bottlenecks:* The provider enables integration with other systems in order to ensure none or only minimum intentional redundancies, as well as easy and convenient access to the multiple services through different devices (Ojala et al., 2020).
- *Purpose and functionality dependencies:* Providers follow functional specifications to support the current process of promotion assessment, by including functionalities for different stakeholders (school management, associations for conducting the assessment and teachers) and their interaction.
- *Platform multi-layer architecture*: The resource and functionality dependencies have to be designed as an open modular multi-layered structure consisting of a stable core with modular services (Riefle et al., 2021).
- *Value creation options:* The open modular structure of the platform provides opportunities for growth and adopting new services within the ecosystem by introducing new stakeholders and their complementary services.
- *Co-provider relations across layers:* Opportunities for growth and adopting new complementary services can contribute to more inclusiveness, motivation to use and client loyalty. Connectors to social media can be a fair example of co-provider relation.
- *Access control & limitations:* Limited access and access control must be provided in order to ensure data protection and support value recognition of intellectual property in publishing teaching

material and other publications related to teachers or student work. Integration with other services like portfolio management can ease the amount of data to store as long as data ownership and data sharing policies are addressed.

- *Efforts through platform evolution phases:* Each platform must undergo certain stages of maturity before it can reach its full potential; nevertheless, some challenges (e.g. technical bottlenecks or user onboarding) can be successfully managed by providing a community support.
- *Incentives for active interaction:* Besides successfully completing the assessment process, an incentive for active interaction after onboarding, i.e. in the retention phase, co-provider relations in social communities of teachers and practitioners can provide high quality data about existing services and programs, offering thereby an opportunity to better forecast future support for developing competencies.
- *Behavior modelling and experience design:* The behavior and experience of stakeholders in the retention phase make adequate preparation of scenarios for future experience design in order to trigger or engage existing or new designed customers' preferences and needs (Ojala et al., 2020).
- *Accountability, regulatory and legal compliance:* The platform provider needs to address policies about labor protections and regulatory frameworks (Hardaker, 2022) as educational systems can interact or handle personal data about teachers and to some extent also the data about students.

From the DP customer's perspective, factors for joining the platform are translated into following design criteria:

- *Zero negative impact:* Low resource consumption by minimizing efforts to store and manage data across multiple systems is a paradigm of interest for customers, since their primary job is to educate.
- *Delivery location & time:* Customers expect remote access and 24/7 availability of any content stored or accessible through the platform.
- *Relationship history with the producer:* Existing experiences, response time and response quality in cases of handling challenges, social interactions and flexibility in build sustainable social relations provides customers the perception of belonging and care within the ecosystem.
- *Payment options:* payment options are in this sense related to covering costs occurring while consuming activities provided by actors in the wider ecosystem (like educational program providers, associations for teacher training, cost related to working on projects).
- *Comfort & convenience:* Usefulness and 24/7 availability are contributing to the customer's sense of comfort and convenience, especially if they are supported by gamification elements contributing to overall user experience, as well as combing the access control and single sign-in or identity management.
- *Recommendations (C2C2C):* Customer-to-customer-to-customer recommendations refers to the support for existing and new complementary communities of teachers and practitioners in the educational ecosystem for exchange of experiences, reviews of educational programs, best practices.
- *Community support:* Within existing and for building new complementary communities, teachers could use social media connectors to establish communication and visibility channels in order to communicate and promote their achievements, recommendations and questions.
- *Producer's reliability:* Customers are assessing the delivered value in relation to their efforts of using another platform and in that sense clear value proposition must be implemented in order to ensure the desired impact.
- *Trust & traceability:* ability to manage, track and review all performance elements which are being assessed through the value chain including administration, preparation, production (e.g. teaching materials), sharing, delivery, while maintaining an acceptable level of trust in social interactions with others.
- *Product or Service safety:* As stated before, teacher's job is primarily to teach and educate, and in that process certain outputs are generated as a result of those and other activities performed at their workplace (including self-education and competencies building). Perception about the quality of the platform to safeguard and manage outputs is therefore important factor and incentive to use a platform.

From the DP producer's perspective, factors for adopting the platform are translated into following design criteria:

- *Sales channels:* In the context of educational institutions and especially the processes of assessment and promotion, "sales" channels are all channels of communication with the "customers" i.e. teachers as consumers of services delivered or managed by producers.
- *Product or Service safety:* In relation to teacher's "job" of educating selves and others, the producer's job is to ensure a stimulating environment for performing activities to gain new competencies and skills prescribed by assessment and promotion rules. Perception about the quality of the platform to safeguard and manage performance measurement data is therefore important factor and incentive to use a platform.

- *Production or Service delivery technologies:* Efficient and effective feasible value creation functionalities, across multiple services through different devices, assuring expected or designed outcomes available 24/7 and remotely with connectors to other systems and social media.
- *Product or Service Quality:* This factor deals with platform's ability to provide feasible value creation scenarios, building thereby the perception on comfort, convenience and trustfulness of available functionalities.
- *Resources:* Resources needed for implementing appropriate digital technologies to empower, support or simplify the process, minimize actors' effort in using the platform and providing value for all stakeholders.
- *Inbound logistics / Supply Chain:* Producer's input is related to opportunities of supporting the teachers in acquiring new skills and competencies, organizing educational and training activities, providing feedback on existing performance milestones, assessing educational programs provided by other stakeholders which can be offered to existing customers and similar inbound services.
- *Innovations:* Based on data mining and existing evaluation protocols and results, the platform should enable options for building innovation potentials, design and adoption of new value creation paradigms, reacting to new technologies and trends, as well as incorporating them into the assessment and promotion process.
- *Outbound logistics / Distribution chain:* Producer's output is related to delivering clear and traceable decisions and evaluation of running and past organized educational and training activities, providing constructive feedback on achieved performance milestones, assessing participation scores on project activities and similar outbound services, while assuring substitute distribution channels in case of disruptions, and supporting collaboration among actors.
- *Incentives and sustainability:* Actors acting as producers should design, build and support efforts contributing to raising readiness to adopt new technologies, skills and competencies, as well as ensure availability of motivating initiatives for providing a variety of incentives for active involvement of customers in the platform.
- *Regulatory compliance:* Producers should design, build and ensure scenarios for managing risks and changing conditions in performing activities related to the assessment and promotion process, as well as contribute to platform's regulatory and legal compliance.

Figure 4 visualizes the Conceptual architecture of an HR digital platform for schools by presenting an overview of different roles, their design criteria and perspectives.

Figure 4. Conceptual architecture of an HR digital platform for schools [Author's work, using storyset.com]

A conceptual architecture model of an HR digital platform for schools for supporting the process of assessment and promotion presented in Figure 4, shows only the most relevant design criteria. Since the design criteria cover most of important functional specification for understanding value propositions for relevant stakeholders, they provide some design input for deciding on future development of such a platform. In order to specify with more detail specific design criteria not occurring in existing systems contributing to HR data in schools, a wireframe illustrating technological extensions is given in figure 5, showing a simplified DP customer view.

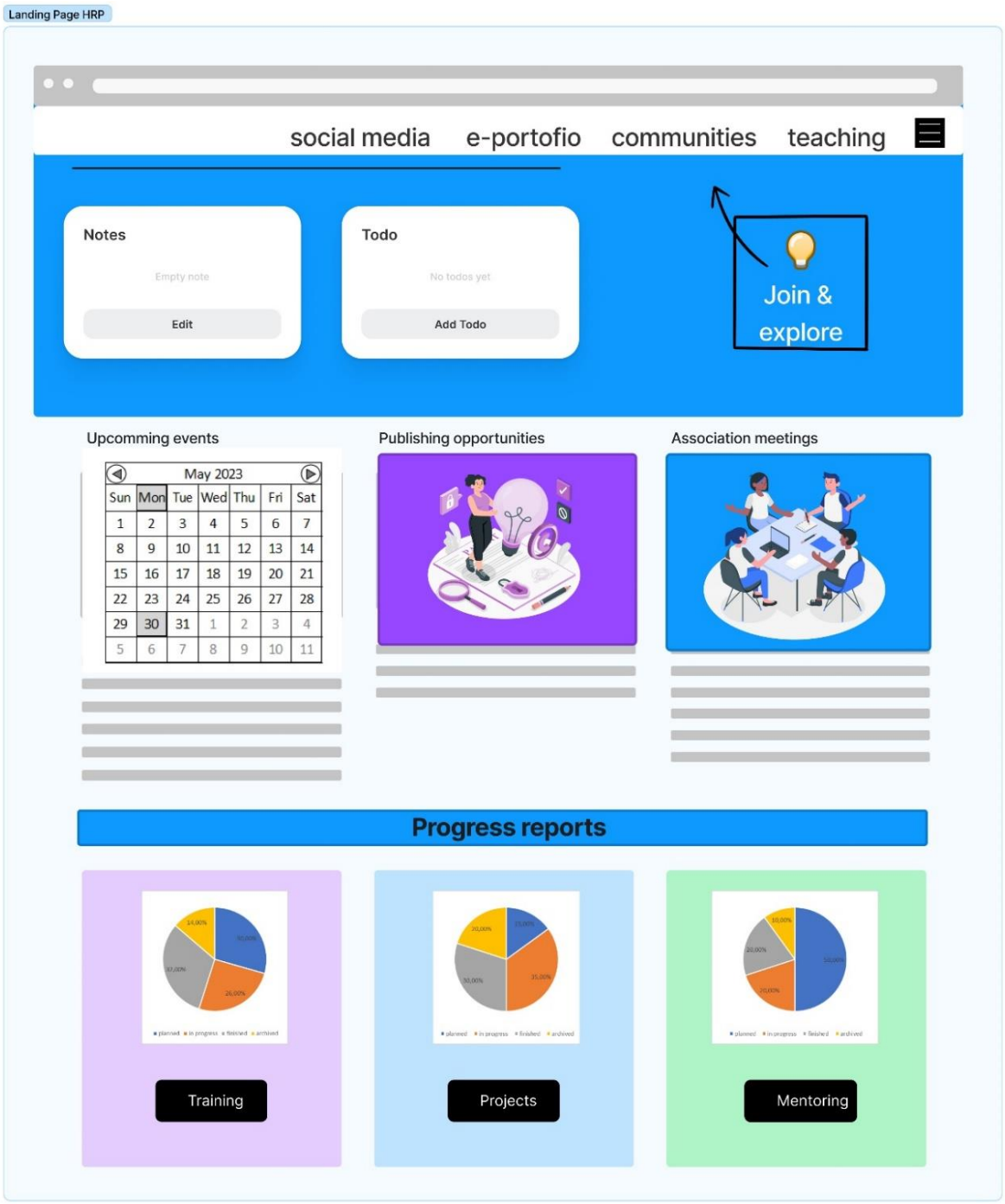


Figure 5. Wireframe example for an HR digital platform for schools – DP customer view [Author's work, using storyset.com and Figma]

4 Evaluation the conceptual architecture of an HR digital platform for schools

A conceptual architecture model of an HR digital platform for and the illustrative wireframe model, are based on the starting value propositions for main actors and cover most design criteria which are intended to contribute in delivering the set value propositions. DP customers could receive value from integration with other systems (like e-portfolio), social media connectors for more collaboration with peers (through communities and direct usage of social media platforms integrated into the HR platform), as well as from gamification elements illustrating progress among selected categories for tracking their personal achievements (e.g. progress report). DP producers could harvest more value through data analytics based on single platform integrating teaching and excellence parameters, as well as more inclusion into the progress of consuming offered activities and events. DP provider could harvest benefit from confirming their role of being a key provider of services and infrastructure, valuable to core actors of the ecosystem.

5 Conclusion

A conceptual architecture model of an HR digital platform for schools for supporting the process of assessment and promotion presented in this paper, is based on the starting value propositions for main actors and described design criteria which are intended to contribute in delivering those value propositions. The design criteria are derived from specific factors of decision-making for using the platform from existing research and seam to cover the basic architectural elements defining the main functional specifications.

Since the specific factors for decision-making of joining, adopting, or implementing a digital platform from existing research offer wide enough but also translatable factors, it can be concluded that such an HR platform could cover important perspectives of relevant stakeholders and, together with a wireframe illustrating technological extensions, provide enough design input for the future development of such a platform. Also, the open and modular architecture set as design criteria from the start, would be a mean for integrating existing systems into a more comprehensive tool designed to aid teachers in managing their educational and other outputs. Such an integrated HR platform would be able to meet the value proposition of supporting the described process of assessment and promotion in schools, and provide collaboration with other stakeholders, aiming to assist the personal development as well as the development of communities within the ecosystem.

Acknowledgments

This research was conducted and supported by EU funded project "e-Schools: Development of a system of digitally mature schools (Phase II)" – Project A: financed from the Operational Programme Competitiveness and Cohesion (OPCC), European Regional Development Fund (ERDF).

References

Chai, M. (2022). Design of Rural Human Resource Management Platform Integrating IoT and Cloud Computing. *Computational Intelligence and Neuroscience, 2022*, 4133048. https://doi.org/10.1155/2022/4133048

Chen, W., & Du, C. (2022). Human Resource Decision-Making and Recommendation Based on Hadoop Distributed Big Data Platform. *Mathematical Problems in Engineering, 2022*, 1–9. https://doi.org/10.1155/2022/8325677

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal*, *30*(1), 114–132. https://doi.org/10.1111/1748-8583.12258

Hardaker, S. (2022). More Than Infrastructure Providers – Digital Platforms' Role and Power in Retail Digitalisation in Germany. *Tijdschrift Voor Economische En Sociale Geografie*, *113*(3), 310–328. https://doi.org/10.1111/tesg.12511

Kaufman, B. E. (2020). The real problem: The deadly combination of psychologisation, scientism, and normative promotionalism takes strategic human resource management down a 30-year dead end. *Human Resource Management Journal*, *30*(1), 49–72. https://doi.org/10.1111/1748-8583.12278

Kutnjak, A., Tomičić-Pupek, K., & Strahonja, V. (2022). Insights into understanding the perspective of digital platform providers. *CECIIS.* https://submissions.foi.hr/index.php/ceciis/article/view/127

Meijerink, J., & Keegan, A. (2019). Conceptualizing human resource management in the gig economy: Toward a platform ecosystem perspective. *Journal of Managerial Psychology*, *34*(4), 214–232. https://doi.org/10.1108/JMP-07-2018-0277

Ojala, A., Rollins, M., Fraccastoro, S., & Gabrielsson, M. (2020). *The Internationalization of B2B Digital Platform Providers: The Role of Cross-National Distance and Digital Characteristics.* http://hdl.handle.net/10125/64313

Peng, Q. (2022). Design of Intelligent Information Platform for Human Resource Allocation Based on Fuzzy Data Mining Algorithm. *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 1159–1162. https://doi.org/10.1109/ICSSIT53264.2022.9716414

Riefle, L., Eisold, M., & Benz, C. (2021). Industrial Corporation's Transformation into a Digital Platform Provider: A Case Study on Enablers. *2021 IEEE 23rd Conference on Business Informatics (CBI)*, *01*, 131–140. https://doi.org/10.1109/CBI52690.2021.00024

Ruparel, N., Dhir, A., Tandon, A., Kaur, P., & Islam, J. U. (2020). The influence of online professional social media in human resource management: A systematic literature review. *Technology in Society*, *63*, 101335. https://doi.org/10.1016/j.techsoc.2020.101335

Skinner, W. (1981, September 1). Managing Human Resources. *Harvard Business Review*. https://hbr.org/1981/09/managing-human-resources

Sukawati, N. N., Gunawan, I., Ubaidillah, E., Maulina, S., & Santoso, F. B. (2020). *Human Resources Management in Basic Education Schools*. 292–299. https://doi.org/10.2991/assehr.k.201112.052

Appendix I: List of publications used for text-mining

1. Li Z., Le Kernec J., Abbasi Q., Fioranelli F., Yang S., Romain O.; Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms; 2023; Scientific Reports
2. Dong W.; A Study on the Construction of Human Resources Audit Management Platform Based on Big Data; 2022; ACM International Conference Proceeding Series
3. Ota K., Baba E.; Kyushu Promotion Plan for the Platform of Human Resource Development for Cancer; 2022; Gan to kagaku ryoho. Cancer & chemotherapy
4. Mat Saad M.F., Listyo Nugro A.W., Thinakaran R., Baijed M.; A Review of Artificial Intelligence Based Platform in Human Resource Recruitment Process; 2022; 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2021
5. Zhao X., Shi L., Ruan S., Bi W., Chen Y., Chen L., Liu Y., Li M., Qiao J., Mao F.; CircleBase: An integrated resource and analysis platform for human eccDNAs; 2022; Nucleic Acids Research
6. Cheng Y.; Robustness Analysis of Strategic Human Resource Management Information Platform based on SPSS Big Data Intelligent Debugging Algorithm; 2022; International Conference on Edge Computing and Applications, ICECAA 2022 - Proceedings
7. Lu Y.; JQuery Implementation of Online Platform for University Human Resource Management Mode Based on Multi-Centralized Networking Algorithm; 2022; 5th International Conference on Inventive Computation Technologies, ICICT 2022 - Proceedings
8. Liu L., Sun B., Xu Q.; Mobile Edge Computing Application in Enterprise Human Resource Management Platform Based on Task Scheduling Algorithm; 2022; Mobile Information Systems
9. Heinrich D., Schramm F.; New perspective on platform work - internal crowdworking as an instrument to facilitate strategic human resources management; 2022; International Journal of Work Innovation
10. Chen W., Du C.; Human Resource Decision-Making and Recommendation Based on Hadoop Distributed Big Data Platform; 2022; Mathematical Problems in Engineering
11. Xu Y.; Data Analysis of Human Resource Performance Appraisal Based on Intelligent Attendance Web Platform; 2022; Lecture Notes on Data Engineering and Communications Technologies
12. Chai M.; Design of Rural Human Resource Management Platform Integrating IoT and Cloud Computing; 2022; Computational Intelligence and Neuroscience
13. Yang H.; Human Resource Big Data Analysis and Decision Making of Group Enterprises Based on Cloud Platform; 2022; Proceedings - 2022 14th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2022
14. Peng Q.; Design of Intelligent Information Platform for Human Resource Allocation Based on Fuzzy Data Mining Algorithm; 2022; Proceedings - 4th International Conference on Smart Systems and Inventive Technology, ICSSIT 2022
15. Luo L., Zhu X.; Management Analysis of Human Resources Sharing Economy Platform Under Big Data Technology; 2022; Lecture Notes on Data Engineering and Communications Technologies
16. Hu Y.; Data Analysis of Human Resource Performance Appraisal Based on Intelligent AttendanceWeb Platform; 2021; ACM International Conference Proceeding Series
17. Yan R., Wang L.-J., Liu L., Li X.-F., Zhou B.-Y., Jiang N., Liu H.; A preliminary study on the mixed teaching of human parasitology based on mooc resources and the experimental teaching digital platform; 2021; Chinese Journal of Schistosomiasis Control
18. Guyonnet S., Rolland Y., Takeda C., Ousset P.-J., Ader I., Davezac N., Dray C., Fazilleau N., Gourdy P., Liblau R., Parini A., Payoux P., Pénicaud L., Rampon C., Valet P., Vergnolle N., Andrieu S., De Souto Barreto P., Casteilla L., Vellas B., INSPIRE Platform Group; The INSPIRE Bio-Resource Research Platform for Healthy Aging and Geroscience: Focus on the Human Translational Research Cohort (The INSPIRE-T Cohort); 2021; Journal of Frailty and Aging
19. Messinger A., Sirmpilatze N., Heuer K., Loh K.K., Mars R.B., Sein J., Xu T., Glen D., Jung B., Seidlitz J., Taylor P., Toro R., Garza-Villarreal E.A., Sponheim C., Wang X., Benn R.A., Cagna B., Dadarwal R., Evrard H.C., Garcia-Saldivar P., Giavasis S., Hartig R., Lepage C., Liu C., Majka P., Merchant H., Milham M.P., Rosa M.G.P., Tasserie J., Uhrig L., Margulies D.S., Klink P.C.; A collaborative resource platform for non-human primate neuroimaging; 2021; NeuroImage
20. Liu P., Qingqing W., Liu W.; Enterprise human resource management platform based on FPGA and data mining; 2021; Microprocessors and Microsystems
21. Kim Y.H., Cha H.R., Lee J.E., Cha S.E., Choi Y.J.; Introduction to the human disease resource search and distribution platform through the Korea Biobank Network portal; 2021; Journal of the Korean Medical Association
22. Keegan A., Meijerink J.; Online labour platforms, human resource management and platform ecosystem tensions: An institutional perspective; 2021; Platform Economy Puzzles: A Multidisciplinary Perspective on Gig Work

23. Yan Z.; Optimization of Relevance Weighting Algorithm Based on Hadoop Platform in Human Resource Information System; 2021; Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021
24. Zhu J., Wang Y., Yu X.; Hybrid Resource Modeling and Scheduling Platform Based on Multisource Cooperation for Cyber-Physical-Human System; 2021; Mobile Information Systems
25. Zhouhuo W.; Parallel Classification Algorithm Design of Human Resource Big Data Based on Spark Platform; 2021; Security and Communication Networks
26. Yang X., Yang Q.; Research on Human Resource Decision-Making Technology Based on Hadoop Platform; 2021; Lecture Notes in Electrical Engineering
27. Zhang C.-M., Liu L.; Big Data Based Human Resources Professional Distance Training Platform; 2021; Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST
28. Kuhn K.M., Meijerink J., Keegan A.; Human resource management and the gig economy: Challenges and opportunities at the intersection between organizational hr decision-makers and digital labor platforms; 2021; Research in Personnel and Human Resources Management
29. Wang Q.; Enterprise human resource management system monitoring based on embedded system and 5G big data platform; 2021; Wireless Networks
30. Boudjedar S., Bouhenniche S., Mokeddem H., Benachour H.; Automatic Human Resources Ontology Generation from the Data of an E-Recruitment Platform; 2021; Communications in Computer and Information Science
31. Meijerink J., Keegan A., Bondarouk T.; Having their cake and eating it too? Online labor platforms and human resource management as a case of institutional complexity; 2021; International Journal of Human Resource Management
32. Liu T.; Informatization construction of human resource management in retail industry based on SaaS platform; 2020; ACM International Conference Proceeding Series
33. Lyu B., Chen J., Wang N.; Live Teaching Platform for Human Resource Development and Management Based on Big Data Recommendation Technology; 2020; Proceedings - 2020 5th International Conference on Smart Grid and Electrical Automation, ICSGEA 2020
34. Khan M.H., Muktar S.N.; A bibliometric analysis of green human resource management based on scopus platform; 2020; Cogent Business and Management
35. Wang H., Yang Y., Zhang Y.; A Macro Human Resource Management Platform Enabled by Big Data Technology; 2020; Advances in Intelligent Systems and Computing
36. Telyatnikova V.S., Shokhnekh A.V., Bespyatykh V.I., Azarskaya M.A., Kolyshev O.Y.; Approaches to Formation of an Innovative Platform of University Complexes in the Strategy of Human Resources Development as Small Business Entrepreneurs; 2020; Lecture Notes in Networks and Systems
37. Sutaryono, Coredikris D., Batubara S.F.A.; The effect of platform and sensor weapon and command (Sewaco) maintenance and the quality of human resources on the operational readiness of the ship; 2019; Journal of Maritime Research
38. Meijerink J., Keegan A.; Conceptualizing human resource management in the gig economy: Toward a platform ecosystem perspective; 2019; Journal of Managerial Psychology
39. Lv Z., Tan Z., Wang Q., Yang Y.; Cloud Computing Management Platform of Human Resource Based on Mobile Communication Technology; 2018; Wireless Personal Communications
40. Koster F.; Human resource management in the platform economy [Personeelsbeleid in de platformeconomie]; 2018; Mens en Maatschappij
41. Kim Y.; A design of human cloud platform framework for human resources distribution of e-learning instructional designer; 2018; Journal of Distribution Science
42. Msiska B.; Pooling human resources needed to leverage open source health information software platforms in developing countries; 2017; 2017 IST-Africa Week Conference, IST-Africa 2017
43. Fonseca A.M., Fernandes F.H., Lima V.J., Souza V.B.; Good practices in the use of human and material resources for optimization in the planning and execution of programmable controller platform migration processes; 2017; Rio Pipeline Conference and Exposition, Technical Papers
44. Silva C., Silva S., Martins D.; The LinkedIn platform in human resources recruitment; 2017; Proceedings of the 29th International Business Information Management Association Conference - Education Excellence and Innovation Management through Vision 2020: From Regional Development Sustainability to Global Economic Growth
45. Sun R., Xie L., Liu J., Lian H.; Research on hospital human resource management based on cloud platform; 2016; RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao
46. Mittas N., Kakarontzas G., Bohlouli M., Angelis L., Stamelos I., Fathi M.; ComProFITS: A web-based platform for human resources competence assessment; 2016; IISA 2015 - 6th International Conference on Information, Intelligence, Systems and Applications

47. Okano Y., Ichikawa Y., Endo I.; Remote lecture for cancer professional training on academic authentication federation-promotion plan for the platform of human resource development for cancer at Yokohama city university; 2016; Yokohama Medical Journal
48. Okano Y., Ichikawa Y., Endo I.; Sustainable development aim-promotion plan for the platform of human resource development for cancer at Yokohama City University; 2016; Yokohama Medical Journal
49. Cabanillas C.; Exploring human resource management in crowdsourcing platforms; 2016; Lecture Notes in Business Information Processing
50. Han Y., Lian H.; Dynamic assessment of human resource management on organizational effectiveness based on network platform; 2016; International Journal of Smart Home
51. Hofman E., Meijerink J.; Platform thinking for services: the case of human resources; 2015; Service Industries Journal
52. Wang Y.Z., Zhang L.Y.; Enterprise human resources management system platform optimization design in the information technology era; 2015; Information, Computer and Application Engineering - Proceedings of the International Conference on Information Technology and Computer Application Engineering, ITCAE 2014
53. Monma N., Gojobori T., Ikeo K.; Human genome network platform: A resource for TFRN analysis; 2014; Methods in Molecular Biology
54. Frade R.F.M., Coelho J.A.S., Simeonov S.P., Afonso C.A.M.; An emerging platform from renewable resources: Selection guidelines for human exposure of furfural-related compounds; 2014; Toxicology Research
55. Jia Y.L.; Design and implementation of university human resources management platform based on YII; 2014; Applied Mechanics and Materials
56. Guo Q., Zhou D.; Information service platform of China human resources in science and technology based on WEBGIS; 2013; Proceedings - 2013 International Conference on Computer Sciences and Applications, CSA 2013
57. Yao W.; Network human resources platform strategies research with two-sided market theory; 2012; Proceedings - 2012 International Conference on Management of e-Commerce and e-Government, ICMeCG 2012
58. Zhao D., Wu J., Zhou Y., Gong W., Xiao J., Yu J.; WikiCell: A unified resource platform for human transcriptomics research; 2012; OMICS A Journal of Integrative Biology
59. Al-Kazi L.; Women and non-governmental organizations in Kuwait: A platform for human resource development and social change; 2011; Human Resource Development International
60. Oshima A.; The development of a "platform for self-directed human resource development" in corporations; 2008; MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of ICT, Society and Human Beings 2008
61. Zhang J., Lu .X.-T., Bu S.-H., Hu S.-H., Yang C.-M., Li F.; Enhancement of efficiency of human resource through the construction of platform of performance management; 2008; Pharmaceutical Care and Research
62. Luo X., Xu C., Tang G.; Study on the building of the information platform aimed at upgrading human resource ability in institutes of national defense industry; 2007; 2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007
63. Chin T.L.; Intranets emerging as new human resources platform.; 1998; Health data management
64. Silvestru, C; Colibaba, DS; Silvestru, DS; The development of the e-learning platforms in the new economy for a better training of the human resources; 2007; ECONOMIC COMPUTATION AND ECONOMIC CYBERNETICS STUDIES AND RESEARCH
65. Bel, N; Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language Technologies. PANACEA; 2010; PROCESAMIENTO DEL LENGUAJE NATURAL
66. Yun, F; Zhu, WH; Huang, S; Zhou, W; A Case Study on the Security Policies for RESTful Open API Platform of Human Resource Management; 2012; CEIT 2012: 2012 INTERNATIONAL CONFERENCE ON CIVIL ENGINEERING AND INFORMATION TECHNOLOGY
67. Haung, S; Zhou, W; Research and Implementation of a RESTful Open API Platform for Human Resource Management; 2012; CEI 2012: 2012 international conference on civil engineering and information technology
68. Tawadrous, MI; Antiado, DF; Castillo, FG; Mobile Platform in the Workplace: The Next Generation Practice in Human Resource; 2016; 3rd global conference on business and social sciences (gcbss-2016) on contemporary issues in management and social sciences research
69. Chen, MN; Wang, H; Construction of Sharing Human Resources Platform for Entrepreneurial Enterprises Based on Internet Plus; 2018; proceedings of the 2018 8th international conference on management, education and information (MEICI 2)

Toward the utilization of chatbots in the banking sector

Marija Đukić, Milica Škembarević, Olga Jejić, Ana Pajić Simović
University of Belgrade
Faculty of Organizational Sciences
Jove Ilića 154, 11010 Belgrade, Serbia
{marija.djukic, milica.skembarevic, olga.jejic, ana.pajic.simovic}@fon.bg.ac.rs

Abstract. *The increasing interest in artificial intelligence led many industry sectors to incorporate this technology in order to remain competitive. The integration of artificial intelligence chatbots in the banking sector has brought significant changes. While chatbots offer efficiency and convenience, they also have limitations in handling complex situations. This research examines the impact of chatbot technology in the banking sector, specifically its application for client communication, and assistance in banking activities. To analyze chatbot capabilities, two use case scenarios are defined, representing real banking services of varying complexity and human involvement. SWOT analysis was used to explore the benefits and limitations of the chatbot application.*

Keywords. artificial intelligence, banking sector, chatbot, SWOT analysis

1 Introduction

The concept of artificial intelligence has fascinated the scientific community for centuries, even predating the invention of computer systems. Philosophers throughout history have pondered and explored the possibilities of developing intelligence in forms other than humans. The term "artificial intelligence" was first recorded by (McCarthy et al., 1955), and has since been used to encompass a wide range of intelligent behaviours exhibited by computers.

The primary goal of artificial intelligence has always been to enable machines to reason and draw conclusions in a manner similar to human beings. Over the years, there have been both successful and unsuccessful attempts to achieve this goal, with different implementations and intentions. One commonly used benchmark for assessing machine intelligence is the Turing test, originally proposed by Alan Turing in the 1950s. The test involves a scenario where an interrogator engages in conversation with either a human or a machine in a separate room. The interrogator's objective is to determine whether they are communicating with a real person or a machine (French, 2000). If the interrogator is unable to distinguish between the responses of the human and the machine, the machine is considered to have passed the Turing test. The Turing test is particularly useful for evaluating programs designed for human interaction, as well as digital avatars and more recently, digital humans.

The COVID-19 pandemic has significantly influenced consumer behaviour, with more people embracing digital solutions and conducting business from the comfort of their homes (Iancu & Iancu, 2023). This shift in preferences and habits has created a demand for online services, prompting companies to adapt and meet the evolving needs of their customers. One notable response has been the implementation of chatbots that can provide customer support and answer inquiries in a manner similar to human employees. The implementation of chatbots has proven to be an effective solution for meeting the growing demand for online services during the pandemic and beyond (Güler et al., 2022).

Artificial intelligence has found applications in numerous domains, including healthcare, finance, transportation, and entertainment. From virtual assistants to autonomous vehicles, artificial intelligence technologies have become an integral part of daily lives (Luo et al., 2022). This paper delves into the application of artificial intelligence and chatbot technology within the banking sector. The research explores how banks are utilizing these technologies to enhance communication with clients and provide assistance across various banking activities. The aim is to identify the benefits and limitations associated with the chatbot technology application in the banking sector by examining its implementation in the Serbian market. To showcase the practical application of chatbots, the research presents two real-life use case scenarios that represent common banking services. A comprehensive SWOT analysis is conducted to evaluate the overall effectiveness of chatbot technology in the banking sector, drawing insights from the specific use cases.

The paper consists of eight main sections. The introduction provides a brief overview of the research topic and the motivation behind the study. The chatbot overview section provides a comprehensive understanding of chatbot technology and its capabilities. The state-of-the-art section presents an

overview of existing studies related to the digitalization of the banking sector. The research purpose section states the objectives of the study that will be addressed. The research method section describes the approach and methods used to address the research objectives. Next, the specific use cases are defined and described in a separate section. The analysis section evaluates and discusses the findings of the research. The conclusion section summarizes the key findings and insights from the research and provides directions for potential areas of future research.

2 Chatbots: An overview

A chatbot is a computer program that utilizes natural language processing (NLP) and artificial intelligence techniques to simulate human conversation. It can understand and interpret user questions or inputs and provide automated responses in a conversational manner. Also known as “conversational agents”, chatbots can mimic human speech, written or spoken, for the purpose of simulating a conversation or interaction with a real person. Artificial intelligence enables chatbots to interact with users through text-based inputs, voice commands, or a combination of both (Ridha & Haura Maharani, 2022). They can be integrated into various platforms such as messaging apps, websites, and social media platforms, allowing real-time communication with users. Chatbots are commonly offered as web-based applications or standalone apps.

In the past, chatbots were primarily text-based and programmed to respond to a limited set of straightforward questions using pre-written responses. While they performed well within the scope of these predefined questions, they struggled with more complex questions or the ones that hadn’t been predicted by the developers. As technology progressed, more rules and natural language processing have been integrated into chatbots making it possible for users to interact with chatbots in a manner that resembles human conversation. The latest generation of chatbots (Luo et al., 2022), often referred to as "virtual assistants," are context-aware and possess the ability to learn from exposure to a growing volume of human language data.

Natural language processing is a branch of artificial intelligence that focuses on enabling computers or computer programs to understand, analyze, and generate human language. This is achieved through the use of machine learning algorithms designed to identify patterns and structures in textual data. One key aspect of NLP is natural language understanding (NLU), which is used to extract context and meaning from user inputs in natural language and provide appropriate responses in accordance with user intention. By understanding the intent behind user inputs, chatbots can provide appropriate responses and take the necessary actions (Adamopoulou & Moussiades, 2020).

Chatbots operate based on the knowledge they have been provided with. They rely on a knowledge model that contains information and responses to questions based on compiled keywords. (Ridha & Haura Maharani, 2022) Chatbots can be designed to handle different types of tasks, and classified depending on different parameters. One classification parameter is based on the amount of data a chatbot can access. Open-domain chatbots can engage in conversations about general topics, while close-domain chatbots are more focused on specific domains or industries. (Nimavat & Champaneria, 2017). Chatbots can also be classified based on the service they provide, taking into consideration the level of interaction and emotional connection they establish with the user. Additionally, they can be classified based on the goals they aim to achieve or according to the permissions provided by their development platforms (Adamopoulou & Moussiades, 2020).

Considering the input processing and response generation method, chatbots can be classified as rule-based, retrieval-based, or generative (Adamopoulou & Moussiades, 2020).

Rule-based chatbots utilize a fixed set of predefined rules to determine appropriate responses based on recognized text inputs. The knowledge used by these chatbots is manually coded by humans and organized using conversational patterns (Ramesh et al., 2017). They are not robust when it comes to handling spelling or grammatical mistakes in user inputs. Since they strictly adhere to predefined rules, any deviation from the expected input format can lead to inaccurate or inappropriate responses.

Retrieval-based chatbots operate by retrieving relevant responses from a repository or pool of pre-existing responses when presented with a user's message (Ji et al., 2014). The system compares the user's input to the available responses and selects the most suitable or reasonable response. However, they cannot generate entirely new answers for novel or unseen questions. Compared to rule-based chatbots, retrieval-based chatbots offer more flexibility by utilizing APIs to query and analyze available resources (Wu et al., 2016).

On the other hand, generative chatbots approach conversations as an input-output mapping problem and learn to generate responses based on training with a large dataset (Peng & Ma, 2019). These chatbots require more sophisticated NLP algorithms as they have the ability to generate new responses, making them more adaptable to handling novel or unseen questions.

3 State-of-the-art

It becomes challenging to pinpoint the exact moment when individuals started placing increasing reliance on

machines and developing trust in artificial intelligence. The COVID-19 pandemic has significantly impacted our society, particularly in terms of digitalization. The authors (Güler et al., 2022) provide a comprehensive description of the effects that brought this shift toward digitalization in a post-COVID-19 world. The group of authors (Negoita et al., 2022) aim to provide a comprehensive overview of the state of digitalization, shedding light on its impact and implications for businesses in terms of skill requirements and the potential for automation.

To maintain competitiveness in the market, it is crucial for companies to align with industry trends and keep up with technological advancements. The author (Stjepić, 2017) discusses determinants for the adoption of new technologies, with a particular focus on SMEs.

(Adamopoulou & Moussiades, 2020) provides an overview of chatbots, covering their definition, creation process, basic elements, and design and development aspects. It offers a concise understanding of chatbot technology, including its structure and functionality, and outlines the considerations and techniques involved in designing and developing chatbots. The paper (Luo et al, 2022) thoroughly examines the computational approaches, usability, and applications of chatbots in various business sectors. It identifies research gaps and proposes new directions to enhance both academic research and practical business applications of chatbot technology.

(Ridha & Haura Maharani, 2022) examines the implementation of artificial intelligence chatbots and assesses their influence on service quality. The authors (Wube et al., 2022) provide a systematic literature review of text-based chatbots in the financial sector. They cover their implementation, adoption intention, user attitudes and experiences, and potential strategies for their advancement.

In (Fares et al., 2022) the authors offer a comprehensive review of the literature on the application of artificial intelligence in the banking sector. The authors conducted a systematic literature review, through thematic and content analysis, the study identifies research themes related to the use of artificial intelligence in banking. The impact of chatbot technology on bank revenues is analyzed in the paper (Hwang & Kim, 2021). It highlights the suitability of chatbots for different types of transactions and demonstrates how chatbots can enhance financial soundness and reduce operational costs.

The research on chatbots in the financial sector was inspired by prior influential works that emphasized the use of this type of artificial intelligence. The objective of this paper was to investigate the implications of chatbots for banking activities and the potential substitution of human interactions with digital assistants.

4 Research purpose

The main purpose of this research is to analyze the current trends in the digitalization of the banking sector, with a particular emphasis on how banks are utilizing artificial intelligence for client communication. The study investigates the use of chatbots and artificial intelligence by banks operating in the Serbian market to enhance customer service and streamline banking processes. The research aims to explore how these technologies are being employed by banks to communicate with clients, provide guidance, and offer assistance throughout different banking activities. By analyzing the current practices and trends in the utilization of chatbots by banks in Serbia, the research seeks to identify areas where banks can further enhance their implementation of artificial intelligence. The research also aims to uncover any limitations or challenges that may arise in the implementation of these technologies in the banking sector and provide recommendations for overcoming them.

4.1 Research objectives

Two main objectives pursued in this research are:

1. To identify the strengths and weaknesses of using chatbots in banking activities in Serbia

   This objective involves evaluating the advantages and disadvantages of chatbot applications by banks operating in the Serbian market in areas such as customer service, banking service guidance, and assistance.
2. To assess the extent to which digital assistants have the capability to replace human employees in performing banking tasks within banks operating in Serbia

   This objective involves examining the capabilities of chatbots in performing various banking activities within banks operating in Serbia and assessing their potential to handle customer interactions and other tasks traditionally performed by human employees.

5 Research method

To address the research objectives, this study followed a structured approach consisting of several phases. The first phase involved gathering relevant data from the existing literature. The aim was to establish a foundation of knowledge and identify key insights from previous studies. In the second phase, specific use cases were defined. These use cases were carefully selected to represent different scenarios that could be tested against a real chatbot. The intention was to cover a range of situations, from those that could be handled entirely by the chatbot to those that would require some level of human assistance. The third phase involved

simulating the defined use case scenarios on the chatbot within a commercial bank operating in the Republic of Serbia. Finally, a SWOT analysis was conducted on the application of chatbots in the defined use cases. The objective was to gain insights into the benefits and limitations of chatbot applications and identify future possibilities and potential challenges that may arise in the banking sector.

SWOT analysis is a strategic planning technique that helps evaluate an organization's competitive position by considering both internal and external factors (Namugenyi et al., 2019). Traditionally, organizations tend to focus on internal factors when making strategic decisions. SWOT analysis addresses this limitation by incorporating both internal and external data sources.

The analysis begins by identifying the organization's strengths, which are its internal advantages and capabilities that give it a competitive edge. Conversely, weaknesses are internal factors that hinder the organization's performance or put it at a disadvantage compared to competitors. Opportunities are external factors that can be leveraged to enhance the organization's performance or achieve strategic goals. By identifying opportunities, organizations can capitalize on favorable conditions to drive growth and success. On the other hand, threats are external factors that pose risks and challenges to the organization's operations. Recognizing threats enables organizations to develop strategies to mitigate or respond to them effectively. By considering both internal and external factors, SWOT analysis helps identify areas of strength to leverage, weaknesses to address, opportunities to seize, and threats to mitigate.

Although SWOT analysis is mostly used to inform strategic planning and decision-making, it is a versatile framework that can be applied in various contexts to assess strengths, weaknesses, opportunities, and threats. In the study (Farrokhnia et al., 2023) focused on education, SWOT analysis is used to identify the specific strengths and weaknesses of chatbot applications in the educational sector.

6 Use cases

The main focus of this research is to evaluate the capabilities of chatbots within the banking sector in the Serbian market. To demonstrate the practical application of chatbot technology in banking activities, the study introduces two use case scenarios: opening a checking account and applying for a mortgage loan. These scenarios represent real banking services that vary in complexity and the level of human involvement required. In the first scenario, opening a checking account, the process is relatively straightforward and can be handled with minimal human interference. In the second scenario, applying for a mortgage loan, the process is more complex and typically involves a higher level of human involvement. Both use case scenarios are based on the actual application of a chatbot within a commercial bank operating in the Republic of Serbia that implemented a chatbot to facilitate and guide customers through their online services. Figure 1 illustrates the initial interaction with the chatbot. By analyzing the chatbot's capabilities in these scenarios, the study aims to provide insights into the strengths, and weaknesses of implementing chatbot technology in the banking sector.

The first use case scenario in the research focuses on the process of opening a checking bank account. Traditionally, opening a bank account involved visiting a physical bank branch and completing various paperwork. However, with the emergence of digital technologies and the increasing popularity of online services, clients now expect the convenience of completing their obligations remotely. Clients can provide their personal data and submit the required documents electronically to prove their identity. The process may also involve signing agreements digitally and activating the account through online channels.

The second use case scenario portrays a more complex process: applying for a mortgage loan. This process typically involves multiple stages, such as gathering financial information, submitting loan applications, providing supporting documents, and undergoing credit checks. Traditionally, this process required extensive manual paperwork and in-person interactions with bank representatives. However, with the advancement of digital technologies, banks are increasingly offering the option for clients to apply for mortgage loans online.

Figure 2 displays a BPMN diagram representing a simulation of the online opening process of a checking account with chatbot assistance. The diagram illustrates each task involved in the account opening process. The chatbot interacts with existing or prospective bank clients, providing guidance and assistance at each step. It is important to note that while

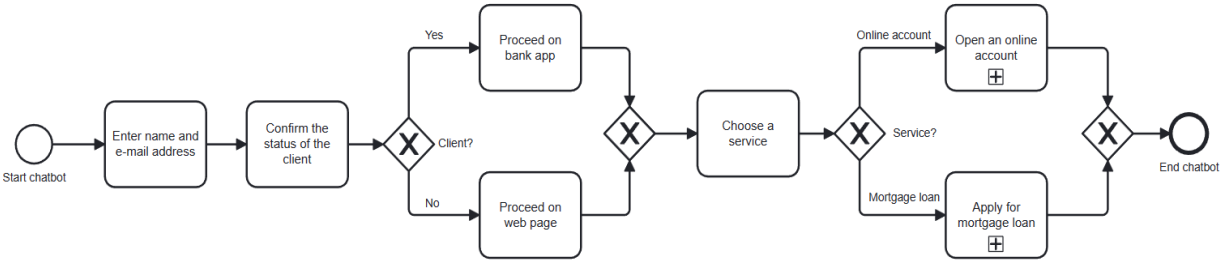


Figure 1. Chatbot application in banking services

the entire account opening process can be completed online, the chatbot's assistance alone may not be sufficient for certain activities. There are still specific tasks within the process that require human involvement.

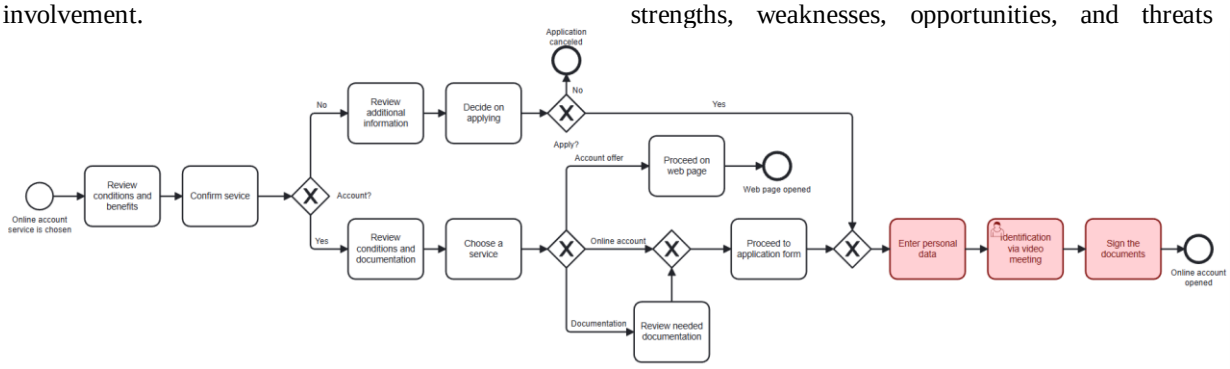


Figure 2. Use case 1: Opening a checking account online (tasks in red – chatbot assistance available only in Serbian, user task – human involvement required)

The second use case scenario involves applying for a mortgage loan, which is a more complex process compared to the first use case scenario. The focus is on the mortgage application phase where a chatbot is utilized to guide and assist clients through the process. The BPMN diagram, represented in Figure 3, represents a simulation of this phase in a previously mentioned bank. The diagram showcases each task involved in the mortgage application phase, with each task corresponding to an activity performed by existing or prospective bank clients. The chatbot plays a significant role in providing guidance and assistance to clients at each step of the process. Despite the chatbot's involvement, the mortgage application cannot be completed entirely online. The complexity of the mortgage application necessitates human involvement, particularly for tasks that cannot be digitized or automated. This means that clients will still need to visit a bank branch to fulfill certain requirements and provide the necessary documentation.

7 SWOT analysis of chatbot application in banking services

Guided by a SWOT analysis framework, this review outlines the strengths and weaknesses of a chatbot technology application in the banking sector, with a specific focus on client communication and assistance. Opportunities and threats are also discussed. The research findings are summarized in Table 1, which presents a comprehensive overview of the identified strengths, weaknesses, opportunities, and threats associated with chatbot technology in the banking sector, drawing insights from the defined use cases.

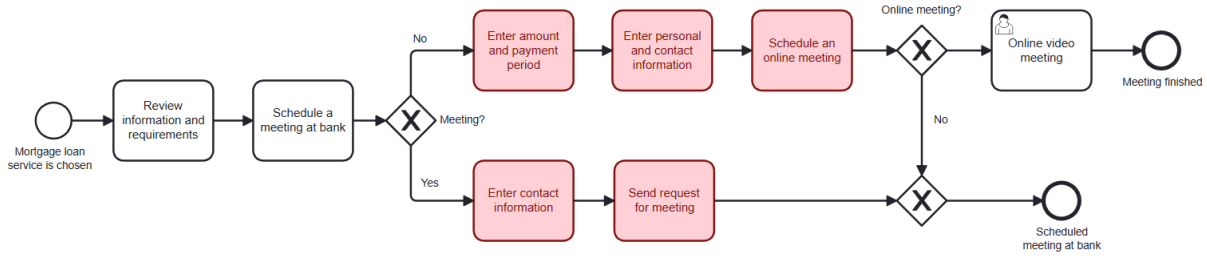


Figure 3. Use case 2: Mortgage loan application (tasks in red – chatbot assistance available only in Serbian, user task – human involvement required)

Service available constantly

One significant advantage of chatbots is their availability 24/7. Unlike human bank employees and customer support representatives who work within specific operating hours, chatbots can provide guidance and assistance at any time.

Quick response

Chatbots offer the advantage of instant response to client inquiries. The automated nature of chatbots enables them to provide prompt responses, eliminating the need for customers to wait in queues or endure long hold times on customer support phone calls.

Interaction with multiple customers at once

One of the key advantages of chatbots is their ability to handle multiple client interactions simultaneously. Unlike bank employees and customer support representatives who typically interact with one client at a time, chatbots have the capacity to engage with multiple clients simultaneously.

Web page navigation guidance

Chatbots in banking can serve as helpful guides for clients navigating banking websites. Based on the

client's query, the chatbot can provide relevant and personalized responses, directing the client to the appropriate web page or section of the banking website.

contain additional information. One potential drawback of the chatbot's functionality is that being redirected to a web page may disrupt the flow of the conversation and result in the loss of the entire previous conversation. In such cases, if users wish to continue the conversation, they are required to enter their name and email again before accessing the chatbot.

Table 1. SWOT analysis of chatbot application in banking services

Strengths	**Weaknesses**
Service available constantly	Name and e-mail required
Quick response	Previous conversation loss
Interaction with multiple clients at once	Language switch
Web page navigation guidance	Lack of empathy
Online process completion (use case 1)	Human involvement required
Process completion outside of working hours (use case 1)	Bank visit required (use case 2)
Simplified initial contact (use case 2)	Inability of understanding situation peculiarity
Opportunities	**Threats**
Service personalization	System crash
Human employees available to handle more complex tasks	Technological literacy
Client satisfaction increase	More sophisticated solutions
Clients' expectations of online process completion	Scepticism of technological assistance
	Dissatisfaction due to not meeting client requirements
	Privacy protection

Online process completion

Traditionally, banking services required clients to visit a physical bank branch and complete various paperwork. Chatbots can assist clients by providing step-by-step instructions and guiding them through various processes online. For instance, if a client intends to open an online bank account, the chatbot can provide clear instructions on the necessary steps.

Process completion outside of working hours

Unlike humans, chatbots can provide guidance and assistance to clients even outside of regular bank working hours. This 24/7 availability allows clients to complete banking processes and seek assistance at their convenience, without having to take time off from work or adjust their schedules.

Simplified initial contact

Bank processes can be fully or partially completed online, while certain initial steps, such as applying for a banking service or scheduling a meeting, can be accomplished entirely through online channels.

Name and e-mail required

In order to access the chatbot that was used in the research, users are required to enter their name and email information. While registration is not mandatory, the chatbot cannot be used if the name and email details are not provided beforehand.

Previous conversation loss

During the conversation with the chatbot, users are often provided with links to relevant web pages that

Language switch

Although the conversation with the chatbot, at the time of this research, is available in both English and Serbian languages, web pages supporting banking processes are currently available only in Serbian.

Lack of empathy

While chatbots are designed to efficiently perform repetitive tasks, there is a risk of them being perceived as impersonal if the designers do not consider the practical and emotional needs of users seeking assistance. Unlike human agents, chatbots can sometimes lack the ability to convey empathy effectively.

Human involvement required

Chatbots play a significant role in guiding and assisting clients throughout various banking processes, but it is important to recognize that their assistance may not be sufficient for certain activities. In both the online account opening and mortgage loan application scenarios, there are specific tasks within the process that still require human involvement.

Bank visit required

Although chatbots can offer valuable assistance and guidance throughout various banking processes, certain activities still require clients to visit a bank branch. For example, while applying for a mortgage loan can be initiated online, subsequent steps in the process necessitate in-person interactions at the bank.

Inability of understanding situation peculiarity

Unlike human agents, chatbots rely on pre-programmed responses and may not fully understand the unique context or emotions of a situation. Human agents can empathize with clients, understand their specific needs, and provide tailored solutions accordingly. In certain situations, such as when applying for a mortgage loan, there can be specific document conditions that require careful attention and clarification.

Service personalization

By utilizing chatbot analytics, banks can gain valuable insights into user behaviour, trends, and patterns, which can be leveraged to deliver relevant and personalized services and offers to their customers. By understanding customer behaviour and preferences, banks can tailor their services and offer to match individual customer requirements.

Human employees available to handle more complex tasks

Chatbots allow banks to address clients' inquiries at any time without the need for human intervention. By leveraging chatbots to handle routine tasks and provide basic assistance, banks can ensure that their employees can allocate their skills and knowledge to more critical and complex issues.

Client satisfaction increase

Financial institutions face significant client acquisition costs, making client retention crucial for their success. To meet changing clients' expectations, many financial institutions, banks in particular, are turning to chatbots as valuable tools. Chatbots offer a range of benefits, including quick response times, round-the-clock availability, and assistance or guidance throughout various processes. The convenience and efficiency offered by chatbots can help attract and retain clients, leading to improved client loyalty.

Clients' expectations of online process completion

The implementation of new capabilities enables banks to provide services through digital platforms, allowing clients to access and complete their obligations remotely. Chatbots serve as virtual assistants, addressing client inquiries and providing relevant information. Moreover, chatbots offer assistance in completing entire processes online, such as opening a bank account or applying for loans.

System crash

Chatbots have been designed using sophisticated technology, but it does not rule out possible disturbances in operation. For retrieval-based, there is a risk of reaching a conversational dead-end if the chatbot is unable to understand the user's intent or if there are no appropriate responses available in its knowledge base.

Technological literacy

Banks are embracing digital technologies to provide their services online, offering convenience and accessibility to their clients. However, a potential challenge lies in the technological literacy of bank clients or potential users. Not all individuals have the same level of familiarity or comfort with digital technologies.

More sophisticated solutions

The development of chatbots is a rapidly evolving field, with many developers actively working to enhance and improve the technology. More sophisticated and advanced chatbot solutions can be expected in the future and provide competition with a competitive advantage.

Scepticism of technological assistance

Despite the increasing integration of technology in various aspects of people's lives, the level of technology acceptance, especially regarding artificial intelligence, varies among individuals. While some people embrace and adopt new technologies readily, others may still hold reservations or scepticism.

Dissatisfaction due to not meeting client requirements

Satisfied clients are more likely to remain loyal and continue using the bank's services. On the other hand, failing to meet client requirements can result in dissatisfaction, leading to negative feedback and potential loss of business. In the banking industry, where client acquisition costs are high, retaining existing clients is often more cost-effective than acquiring new ones.

Privacy protection

The use of chatbots in collecting and analyzing user data can potentially raise privacy concerns. Chatbot interactions may involve the collection of personal information, and users may have concerns about how this data is stored, used, and protected.

8 Conclusion

Chatbots have emerged as valuable digital assistants across various sectors, particularly in customer support operations. The adoption of chatbots is a response to the increasing usage of digital technologies and evolving expectations of clients to complete their obligations remotely. In the banking sector in Serbia, chatbots have been deployed to address client inquiries and provide guidance for online services. Addressing the first research objective, this study utilized the

SWOT analysis framework to assess the effectiveness of a chatbot application in banking activities. The research outlines the strengths and weaknesses of chatbots in the banking sector and explores the opportunities and threats associated with their implementation.

Although chatbots excel in handling routine tasks, offering guidance, and providing assistance to clients, there are limitations to their capabilities. Certain activities within the banking sector still necessitate human involvement. Addressing the second research objective, this study examines two specific use cases to showcase the extent to which chatbots can replace human employees and handle tasks traditionally carried out by humans. However, the importance of human involvement and personalized service cannot be overlooked. Human employees bring a unique level of empathy, critical thinking, and problem-solving skills that are essential in complex financial situations and addressing the diverse needs of clients.

8.1 Future research

Possible future areas of improvement of this research topic include:

a) Quantitative questionnaire-based research approach to evaluate client satisfaction and their overall impression of communication after the introduction of chatbots,
b) Further refining and expanding the chatbot language base with domain-specific terminology for banking sector,
c) Investigating possible ways to connect chatbot functionality with a digital human to create a more human-like experience for users.

Acknowledgments

This paper is funded by the Faculty of Organizational Sciences, University of Belgrade.

References

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16 (pp. 373-383). Springer International Publishing.

Fares, O. H., Butt, I., & Lee, S. H. M. (2022). Utilization of artificial intelligence in the banking sector: a systematic literature review. Journal of Financial Services Marketing, 1-18.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 1-15.

French, R. M. (2000). The Turing Test: the first 50 years. Trends in cognitive sciences, 4(3), 115-122.

Güler, M., Mukul, E., & Büyüközkan, G. (2022). Digital Transformation Technology Evaluation in a post COVID-19 World with an Integrated Fuzzy SAW-ARAS Methodology. In Central European Conference on Information and Intelligent Systems (pp. 417-421). Faculty of Organization and Informatics Varazdin.

Hwang, S., & Kim, J. (2021). Toward a chatbot for financial sustainability. Sustainability, 13(6), 3173.

Iancu, I., & Iancu, B. (2023). Interacting with chatbots later in life: A technology acceptance perspective in COVID-19 pandemic situation. Frontiers in psychology, 13, 1111003. https://doi.org/10.3389/fpsyg.2022.1111003

Ji, Z., Lu, Z., & Li, H. (2014). An information retrieval approach to short text conversation. arXiv preprint arXiv:1408.6988.

Luo, B., Lau, R. Y., Li, C., & Si, Y. W. (2022). A critical review of state-of-the-art chatbot designs and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(1), e1434.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine, 27(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

Namugenyi, C., Nimmagadda, S. L., & Reiners, T. (2019). Design of a SWOT analysis model and its evaluation in diverse digital business ecosystem contexts. Procedia Computer Science, 159, 1145-1154.

Negoita, O. D., Popescu, M. A. M., Antonova, A., & Bork, D. (2022). A comparison of business platforms used by SMEs to digitalize management activities. In Central European Conference on Information and Intelligent Systems (pp. 195-203). Faculty of Organization and Informatics Varazdin.

Nimavat, K., & Champaneria, T. (2017). Chatbots: An overview types, architecture, tools and future possibilities. Int. J. Sci. Res. Dev, 5(7), 1019-1024.

Peng, Z., Ma, X. A survey on construction and enhancement methods in service chatbots design. CCF Trans. Pervasive Comp. Interact. 1, 204–223 (2019). https://doi.org/10.1007/s42486-019-00012-3

Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017). A survey of design techniques for conversational agents. In Information, Communication and Computing Technology: Second International Conference, ICICCT 2017, New Delhi, India, May 13, 2017, Revised Selected Papers (pp. 336-350). Singapore: Springer Singapore.

Ridha, M., & Haura Maharani, K. (2022). Implementation of Artificial Intelligence Chatbot in Optimizing Customer Service in Financial Technology Company PT. FinAccel Finance Indonesia. Multidisciplinary Digital Publishing Institute Proceedings, 83(1), 21.

Stjepić, A. M. (2017). Survey of the determinations of business intelligence systems adoption in SMEs. In Central European Conference on Information and Intelligent Systems (pp. 177-185). Faculty of Organization and Informatics Varazdin.

Wube, H. D., Esubalew, S. Z., Weldesellasie, F. F., & Debelee, T. G. (2022). Text-based chatbot in financial sector: A systematic literature review. Data Science in Finance and Economics, 2(3), 232-259.

Wu, Y., Wu, W., Xing, C., Zhou, M., Li, Z. (2016) Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-based Chatbots.

Inspection of the road surface condition on black spots – Croatia case study

Darko Babić, Mario Fiolić, Dario Babić
University of Zagreb
Faculty of Transport and Traffic Sciences, Department of Traffic Signalling
Kušlanova 2, 10 000 Zagreb, Croatia
{dbabic, mfiolic, dbabic2}@fpz.unizg.hr

Abstract. *In pursuit of a safe road network, adequate design, construction, and maintenance of road surfaces are imperative. However, inevitable wear and tear over time lead to damage. The previous project identifies 23 black spots on the state road network of the Republic of Croatia by analyzing traffic accidents and influential factors. The road surfaces of these black spots were meticulously examined using a specially equipped vehicle, assessing parameters such as international roughness index (IRI), macrotexture, rutting, and cracking. By employing this road surface inspection system, effective road maintenance strategies can be proposed to minimize the risk of traffic accidents.*

Keywords. Black spot, traffic safety, road maintenance, road surface inspection, Hawkeye 2000

1 Introduction

About 1.3 million people die in road accidents every year (Road Traffic Injuries, 2022), and the most critical factors affecting road safety are the driver, the vehicle, and the road and its surroundings. The road surface is crucial for smooth traffic flow and safety but inevitably sustains damage over time. Proper design, construction, and maintenance are essential for a safe road network. The road surface condition directly affects ride quality and indirectly influences driver distraction, vehicle operation, and traffic accidents. Consequently, it is essential to inspect the condition of the road surface, identify areas with damaged road surface, and promptly repair them to prevent potential negative impact on traffic safety (Vinayakamurthy et al., 2017).

Based on data from the Ministry of Internal Affairs of the Republic of Croatia, an average of 35,386 traffic accidents occurred on Croatian roads in the past decade, with an annual fatality rate of 351 individuals. The analysis further indicates that specific locations, especially shorter road segments, experience a higher frequency of accidents than the rest of the state road network. Potentially dangerous locations on the road network are commonly referred to as black spots. These can be defined as specific or micro-locations on the road that exhibit a higher risk of accidents than other roads in the network with similar traffic and technical characteristics (Road Traffic Safety Bulletin 2021, 2022).

Efforts can be made to treat the road infrastructure to reduce the likelihood and severity of traffic accidents. In line with this objective, the Croatian Ministry of Sea, Transport, and Infrastructure has signed a contract valued at over 9 million euros, which includes more than 7 million euros in grants for a project aimed at rehabilitating dangerous locations and eliminating black spots on state roads. Within this contract and the "Rehabilitation of 23 Dangerous Locations" project, Croatian Roads Ltd., the main road authority in Croatia, utilize these funds to improve road safety on a total of 23 black spots that were determined on state roads (secondary roads) in 12 counties across Croatia. One of the specific objectives is related to the inspection of road surface conditions using a special vehicle equipped with the Hawkeye 2000 ACD system, and the Department of Traffic Signaling, Faculty of Transport and Traffic Sciences has taken on the responsibility of recording, processing data, and creating a report analyzing the condition of pavement structures at the specified locations. This paper aims to present the results of examining the fundamental indicators of road surface conditions at locations identified as black spots by applying the methodology for determining dangerous locations.

2 Applied methodology for detecting black spots in Croatia

There are several derivatives of the term "black spot". Black spots on the road are also referred to as dangerous locations, high-risk locations, accident-prone locations and hotspots. They are defined as points where the number and characteristics of traffic accidents are prominent or where there are potential safety risks compared with other locations over a certain period due to the influence of the road

conditions, traffic conditions, climate, environment, etc. Three types of definitions of a dangerous location are generally accepted: numerical definitions, statistical definitions, and definitions based on the prediction of traffic accidents. Identifying black spots, i.e., dangerous locations in road traffic, is one of the most effective ways to reduce and prevent traffic accidents. The most critical activities in the identification phase include recording data on traffic accidents, developing a database on accidents, and defining criteria for identifying black spots (Zovak & Brčić, 2014).

In 2004, Croatian Roads Ltd. collaborated with the Institute of Civil Engineering of Croatia p.l.c. and published the methodology for identifying dangerous locations on roads in the Republic of Croatia titled "Security Approach Methodology of Traffic". It defined three criteria to classify a road section as dangerous: 1) 12 or more injury-related traffic accidents within the past three years, 2) 15 or more traffic accidents, regardless of consequences, within the previous three years, and 3) three or more traffic accidents of the same type involving the same participant groups and movement directions on the same conflict surface within the last three years. However, this methodology solely relied on the number of traffic accidents as a parameter, lacked statistical analysis of accidents at specific locations, and was not aligned with European standards and regulations. Consequently, a new methodology, titled "Methodology for the identification of dangerous locations in the road transport network" was developed in 2016. This revised methodology adhered to European Union laws and regulations and employed the Rate Quality Control (RQC) method, and involved two critical parameters for assessing the occurrence of traffic accidents: the number of traffic accidents and the average annual daily traffic volume (AADT) at the observed location. If the accident rate surpasses the critical level, it indicates that accidents are not random but rather identified black spots. Detailed analysis of accidents at specific locations aims to identify common patterns and contributing factors (Methodology for identification of dangerous locations in road transport network, 2016).

To determine dangerous spots or road stretches, the road was segmented into sections with similar characteristics, such as horizontal curves, bridges, and tunnels, within the same road category. The Croatian methodology employed the three-year period and a "sliding window" method, defining a 300-meter frame around each accident location. If additional accidents occurred within or adjacent to the frame, it could be extended up to a maximum of 1000 meters to encompass those locations. It is essential to note that this method only considered locations with at least one traffic accident, disregarding accident-free areas (Methodology for identification of dangerous locations in road transport network, 2016). The process of black spot identification involved five main steps, as depicted in Figure 1. Following the statistical investigation, identification, and inspection of a potentially dangerous location, a final determination is made to classify it as a genuine dangerous site or a non-genuine one. If a specific road defect or infrastructure issue is identified as a cause or a contributing factor to traffic accidents with severe consequences, the identified dangerous site is confirmed as genuine, leading to the subsequent phase of managing dangerous locations (Methodology for identification of dangerous locations in road transport network, 2016).

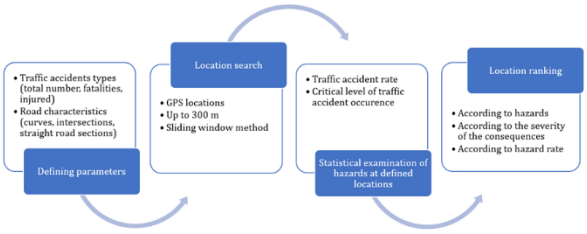


Figure 1. The process of black spot identification Source: (Methodology for identification of dangerous locations in road transport network, 2016)

3 Road surface inspection methodology

Recording of the road surface condition using a special vehicle equipped with the Hawkeye 2000 ACD system was carried out on 23 black spots on Croatian state roads. In the continuation are listed the inspection locations, i.e., identified black spots, and the essential observed parameters, testing procedure, and equipment used are described.

3.1 Observed parameters

The distress parameters collected during the road inspection using a special vehicle equipped with the Hawkeye 2000 ACD system include the International Roughness Index (IRI), macrotexture, surface cracking, and ruts on the road. These parameters measure the roughness, size, and depth of surface irregularities, the presence and severity of cracks, and the depth and severity of wheel ruts. The proposed values of indicators for the "Inspection Work Threshold" (IWT) and "Maximum Acceptable Value" (MAV) levels are provided in Table 1. Proposed values of road surface condition indicators, under the valid Regulations (Regulation on Amendments and Supplements to the Regulation on Road Maintenance, 2021). Defining the intervals within the IWT was done in accordance with recommended criteria within the framework of the COST 354 Action program: Performance Indicators for Road Pavements. Measurement values at the IWT level indicate a road surface condition close to the "Acceptable" state. At this point, more detailed investigative work should be

undertaken to determine the need for maintenance work to limit further deterioration. Values at the MAV level represent a road surface in a "Very Poor" or unsatisfactory condition, where extraordinary maintenance will likely be required.

Table 1. Proposed values of road surface condition indicators

Indicator	Unit	Inspection work threshold (acceptable)	Max. acceptable value (very poor)
IRI	m/km	2,2	3,2
Rut depth	mm	15	20
Macrotexture	mm	0,6	0,4
Block cracking	%	1%	20%
Alligator cracking	%	25%	25%

Source: (Regulation on Amendments and Supplements to the Regulation on Road Maintenance, 2021)

The evenness of the road surface is represented by the International Roughness Index (IRI), which is calculated by summing the vertical oscillations of the vehicle base over a road section (Šimun & Rukavina, 2009). The roughness index represents a measure of driving comfort on a specific road section, and the technical parameters of longitudinal road evenness are measured in the right and left wheel paths (Šimun & Rukavina, 2009).

The macrotexture directly affects the skid resistance of the road surface. It is defined as the deviation of the road surface from an ideally flat surface. The result of macrotexture measurements using non-contact methods is a technical parameter called Mean Profile Depth (MPD).

In mechanical terms, rutting refers to permanent deformations on the road surface that occur at wheel path locations, resulting from the action of traffic loads. Rutting harms driving comfort, safety, and, ultimately, the durability of the pavement. It can be determined in various ways (typically measured in the right and left wheel paths), and in Croatia, it is defined by the technical parameter of rut depth.

Cracking can be divided into four main types: fatigue cracking (alligator cracking), block cracking, longitudinal cracking, and transverse cracking. Fatigue cracks are small irregular areas caused by a single or a series of interconnected cracks. They occur in areas subjected to repeated traffic loads and depend on road surface temperature and vehicle speed. Block cracks are defined as interconnected cracks in an untraveled area that divide the road surface into rectangular blocks of varying sizes (from about 0.1 to 10 m2). They occur across the entire road surface width, not just in wheel lanes. Longitudinal cracks are long cracks that run parallel to the road's centerline. They can occur anywhere between the centerline and the outer edge of the outside wheel lane. Transverse cracks are cracks that run perpendicular to the centerline of the road. They can occur anywhere on the surface, are regularly spaced, and begin as thin hairline cracks that later widen (Ali et al., 2018). The Automatic Crack Detection (ACD) system utilizes two-line lasers to collect data with a crosswise width of up to 5.0 meters and a lengthwise width of up to 1.0 meters. The left side of Figure 3 illustrates the scanning area (ACD laser scan) captured by the two-line lasers. The software processes the data (

Figure **2** – right) based on detected longitudinal road markings. If road markings are absent, the processing range is defined by the width of the road or the width of a single lane. Based on this information, the processing range is established manually or automatically and is used to process detected cracks automatically.

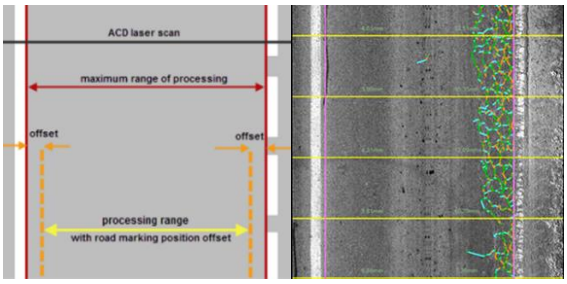


Figure 2. Illustration of ACD scanning area (left) and ACD frame with automatic detection of cracks and rut depth (right)

3.2 Equipment

The road surface condition was recorded using the Hawkeye 2000 ACD system, installed in a Croatian Roads' Ltd. specialized vehicle. This system on the vehicle consists of the following components: a laser profiler, an LCMS system for automatic crack detection, video cameras, GPS (GNSS positioning system), GPSI-TRAC 2 (a road geometry examination system), an odometer, an accelerometer, and computer hardware for data processing and storage. The laser profilometer has two lasers that measure IRI, longitudinal profile, and macrotexture. These measurements were executed using a laser bar mounted on a vehicle moving at average speed, ensuring minimal traffic flow disruption during data collection. The LCMS system consists of two 3D lasers mounted on a vehicle perpendicular to the road, emitting infrared radiation from an opening 2.2 m above the ground. Both of these systems are shown in
Figure **3**. Other components of the system include a control console, a DMI (Distance Measuring Instrument), an accelerometer, the HEARTBEAT central module, a computer, the Onlooker Live software for data collection, and the Processing toolkit for data processing.

Figure 3. Special vehicle with laser profiler (left) and LCMS system (right)

3.3 Inspection locations

The recording and data processing included 23 dangerous locations on 14 different state roads, covering 70 kilometers of road that were inspected. The characteristics of each location are presented in Table 1. Out of the 23 of observed locations, 12 were intersections, and the remaining 11 were road stretches.

Table 2. Characteristics of 23 inspected locations

Nr.	State road	Location name	Length (km)	Road section
1	D8	Rijeka, intersection DC8-DC304	2,08	intersection
2	D8	Bakar	1,54	intersection
3	D8	Rijeka 1	1,28	intersection
4	D8	Rijeka, Krešimirova	1,61	intersection
5	D8	Jasenova	6,47	road stretch
6	D8	Rijeka, intersection J. P. Kamova – DC 8	0,80	intersection
7	D8	Jadranovo - Klanfari	9,92	road stretch
8	D5	Filipovac	2,25	road stretch
9	D46	Ivankovo, Ulica Bošnjaci 180-190	1,74	road stretch
10	D38	Dervišaga Kuzmica	3,01	road stretch
11	D55	Vinkovci, Boškovićeva	1,95	intersection
12	D410	Split - B. Bušića	3,66	intersection
13	D51	Baćin Dol	3,26	road stretch
14	D8	Vodice, Šibenik	1,17	intersection
15	D410	Split, Spinčićeva	2,36	intersection
16	D3	Varaždin	3,27	intersection
17	D45	Garešnica	1,23	road stretch
18	D27	Intersection, DC 27-DC 59	1,90	intersection
19	D8	Zadar, intersection DC 8 - DC 424	3,06	intersection
20	D106	Zrće	1,40	road stretch
21	D3	Dubravci	5,70	road stretch
22	D1	Rakov Potok	3,34	road stretch
23	D6	Brajakovo Brdo	6,73	road stretch

3.4 Inspection procedure

The field inspection typically involves two individuals: a driver, and an operator. The inspection for the road stretch was conducted in two steps: first, starting with the measurement of one direction of the road stretch and second, the measurement of the opposite direction. If the black spot was an intersection, measurements were performed at all approaches to that intersection. After the field inspection, data processing and analysis were carried out using the Hawkeye Processing Toolkit software. Teams of 1 to 4 individuals were involved in the post-field inspection data processing.

4 Results

The processing of the obtained data at 23 dangerous locations on state roads in the Republic of Croatia involved measuring the longitudinal profile expressed by the International Roughness Index (IRI), measuring macrotexture (Mean Profile Depth – MPD), measuring transverse road evenness (RD), detecting road surface crack quantities, and creating georeferenced videos to mark road surfaces and road signaling. The subsequent sections of this article are an overview of each technical parameter, indicating the percentage by which each recorded technical parameter meets, fails to meet, or is acceptable for each dangerous location within the network of state roads in the Republic of Croatia.

4.1 International roughness index (IRI)

After road surface inspection and processing of the obtained data, it was determined that 46% of longitudinal profile falls under an unsatisfactory road surface condition (from 2.6 to 3.5 m/km), 22% is considered acceptable (ranging from 1.2 to 2.6 m/km) but bordering on IWT. In comparison, 32% represents satisfactory results (from 0.5 to 1.2 m/km). **Error! Reference source not found.** shows average IRI values for each location and the percentage that is satisfactory, acceptable, or unsatisfactory in terms of the IRI values.

Table 3. IRI overview on 23 dangerous locations

Nr.	International roughness index – IRI			
	Av. (m/km)	Satisfactory	Acceptable	Unsatisfactory
1	2,4	13%	46%	41%
2	2,7	29%	24%	47%
3	3,1	25%	19%	56%
4	4,1	22%	0%	78%
5	1,4	87%	8%	5%
6	4,8	0%	20%	80%
7	1,6	85%	14%	1%
8	3,8	8%	25%	67%
9	2,8	0%	33%	67%
10	2,1	31%	53%	16%
11	1,5	81%	14%	5%
12	3,7	21%	7%	72%
13	1,3	94%	6%	0%
14	3,6	0%	42%	58%
15	4,2	18%	0%	82%
16	2,4	39%	25%	36%
17	2,4	38%	31%	31%

18	2,7	17%	35%	48%
19	2,9	21%	18%	61%
20	3,2	27%	13%	60%
21	3,2	12%	12%	76%
22	1,5	62%	26%	12%
23	2,7	10%	38%	52%

* **Av. – average value**

4.2 Macrotexture

After processing the obtained data for macrotexture on the inspected locations, alarming data has been obtained. Specifically, overall results show that 80% of the macrotexture (ranging from 0.6 to 0.3mm) falls into an unsatisfactory range of road surface texture, indicating poor skid resistance of the road surface. Further, 12% of the macrotexture is considered acceptable (ranging from 0.9 to 0.6mm), while only 8% represents a satisfactory level of road surface grip (ranging from 1.1 to 0.9mm). **Error! Reference source not found.** shows the average macrotexture parameter of each inspected location and the proportions of surface area that are satisfactory, acceptable, or unsatisfactory in terms of macrotexture.

Table 4. Macrotexture overview on 23 dangerous locations

Nr.	Macrotexture – MPD (mm)			
	Av. (mm)	Satisfactory	Acceptable	Unsatisfactory
1	0,57	0%	13%	87%
2	0,73	18%	29%	53%
3	0,64	6%	31%	63%
4	0,52	0%	0%	100%
5	0,65	15%	32%	53%
6	0,54	0%	10%	90%
7	0,66	0%	33%	67%
8	0,55	0%	0%	100%
9	0,43	0%	0%	100%
10	0,59	0%	19%	81%
11	0,44	0%	5%	95%
12	0,53	0%	7%	93%
13	0,68	27%	3%	70%
14	0,44	0%	0%	100%
15	0,50	4%	7%	89%
16	0,52	8%	6%	86%
17	0,57	0%	23%	77%
18	0,49	0%	0%	100%
19	0,43	3%	0%	97%
20	0,74	26%	27%	47%
21	0,68	19%	19%	62%
22	0,23	0%	0%	100%
23	0,94	63%	9%	28%

* **Av. – average value**

4.3 Surface cracking

The inspection results show that 44% of the road surface at the 23 dangerous locations exhibits an unsatisfactory road surface condition, representing a crack ratio exceeding 20% within a measurement interval of 100 meters. 19% of the road surface falls within an acceptable range, specifically at the IWT (cracking coverage ranging from 15% to 20% within a measurement interval of 100 meters). In comparison, 37% of the surface has a cracking level below 10% and is considered satisfactory. The level of acceptability of the cracking percentage of each inspected location is shown in **Error! Reference source not found.**.

Table 5. Cracking percentage overview on 23 dangerous locations

Location number	Cracking – Cracking Percentage (%)		
	Satisfactory	Acceptable	Unsatisfactory
1	58%	21%	21%
2	6%	29%	65%
3	38%	25%	37%
4	17%	17%	66%
5	49%	28%	23%
6	60%	20%	20%
7	67%	28%	5%
8	8%	8%	84%
9	6%	6%	88%
10	28%	25%	47%
11	85%	10%	5%
12	31%	33%	36%
13	68%	3%	29%
14	61%	8%	31%
15	18%	32%	50%
16	0%	0%	100%
17	15%	23%	62%
18	70%	26%	4%
19	49%	30%	21%
20	46%	27%	27%
21	9%	7%	84%
22	79%	15%	6%
23	0%	4%	96%

4.4 Ruts on the road

After the road inspection, the data obtained indicate an excellent condition of transverse irregularities, i.e., road rutting. Specifically, 97% of the inspected road surface contains minimal irregularities (up to 4.5mm), while 3% falls within an acceptable range, with irregularities up to a depth of 10mm. The average rut depth of all inspected locations is shown in **Error! Reference source not found.**, as well as the level of acceptability for each location.

Table 6. Rutting overview on 23 dangerous locations

Nr.	Transverse irregularities – Rutting (RD)			
	Av. (mm)	Satisfactory	Acceptable	Unsatisfactory
1	2,3	100%	0%	0%
2	4,8	100%	0%	0%
3	5,4	75%	25%	0%
4	5,2	83%	17%	0%
5	5,3	100%	0%	0%
6	2,4	100%	0%	0%
7	3,2	100%	0%	0%
8	4,5	100%	0%	0%

9	5,2	100%	0%	0%
10	2,7	100%	0%	0%
11	2,0	100%	0%	0%
12	3,8	100%	0%	0%
13	1,6	100%	0%	0%
14	3,9	100%	0%	0%
15	4,4	96%	4%	0%
16	4,9	89%	8%	3%
17	3,6	100%	0%	0%
18	3,1	100%	0%	0%
19	4,9	100%	0%	0%
20	5,3	93%	7%	0%
21	4,7	95%	5%	0%
22	5,1	97%	3%	0%
23	3,7	100%	0%	0%

*** Av. – average value**

5 Discussion

The results of the road surface inspection at 23 dangerous locations on state roads in the Republic of Croatia provide valuable insights into the current condition of the road. The measurements of various technical parameters, including the International Roughness Index (IRI), macrotexture (Mean Profile Depth – MPD), transverse road evenness (RD), and surface cracking, offer a comprehensive overview of the road surface condition.

The analysis of the International Roughness Index (IRI) reveals that the majority of the inspected locations exhibit unsatisfactory road surface conditions, with 46% falling within this category. The acceptable and satisfactory percentages of the IRI are 22% and 32%, respectively, suggesting that some locations show better road surface conditions. However, these numbers are not significant enough to ignore the need for road maintenance and improvement. Among the inspected locations, "D410 Split, Spinčićeva" and "D8 Rijeka, intersection J.P. Kamova - D8" show the highest percentage of unsatisfactory road surface conditions, above 80%. Roughness is closely related to vehicle operating costs, vehicle dynamics, and drainage i.e., ride quality. Cairney and Bennett (2008) conducted a study in Victoria, Australia in which authors measured roadway properties using a multi-laser profilometer and linked them to accident data using the GPS. They found that there was a significant correlation between roughness and accident rate that followed a polynomial relationship (Cairney P, 2008). Another study found that higher road surface roughness is associated with increased accident rates, road surface with rut depths exceeding 23.5 mm and IRI values greater than 3.2 m/km were considered poor and more susceptible to accidents. On the other hand, in their study Tsubota et al. discovered counter-intuitive result showed that higher IRI values were associated with decreased accident risk, likely due to drivers being more cautious and slowing down on rougher sections (Tsubota, Fernando, Yoshii, & Shirayanagi, 2018).

The analysis of macrotexture, expressed by the Mean Profile Depth (MPD), reveals alarming data regarding the skid resistance of the road surfaces. An overwhelming 80% of the inspected locations exhibit unsatisfactory road surface texture, indicating a high risk of reduced friction and increased accident potential. The acceptable level of macrotexture is only present in 12% of locations, while only 8% of locations show satisfactory road surface grip. Of the 23 examined locations, five showed 100% unsatisfactory conditions related to the macrotexture. These results call for immediate attention to improve the skid resistance of the road network. Several studies have examined the relationship between road surface friction and crash occurrence rate, and they share a common finding: There is a higher risk of traffic accidents on road surfaces with lower skid resistance, specifically lower macrotexture values. In other words, the weaker the skid resistance and the lower the macrotexture values of the road surface, the greater the risk of accidents. A study conducted by (Austroads technical report: Road surface characteristics and crash occurrence: A literature review, 2008) suggests that when macrotexture falls below a critical value, typically around 1 mm (MPD), the crash rate increases significantly, sometimes by more than 80%. Another study revealed moderate but consistent increase in crashes when sensor-measured texture depth fell below 0.4 mm (Cairney P. , 2000). Roe et al. (Roe, Webster, & West, 1991) and Gothie (Gothie, 2001) conducted studies that revealed increased accident rates on roads with low friction. Additionally, Roe et al. (Roe, Webster, & West, 1991) and Cairney (Cairney P. , 2005) found increased accident rates at junctions on roads with low macrotexture. However, Cairney (Cairney P. , 2005) did not find any relationship between macrotexture and accidents on wet roads, while Roe et al. (Roe, Webster, & West, 1991) found no difference in effects between dry and wet roads or between different types of accidents. Furthermore, Buddhavarapu et al. (Buddhavarapu, Banerjee, & Prozzi, 2013) suggested a correlation between pavement condition and crash injury severity on two-lane undivided horizontal curves, emphasizing the importance of lateral friction.

The evaluation of surface cracking on the inspected locations indicates that 44% of the road surfaces exhibit unsatisfactory conditions, with a high cracking ratio exceeding 20% within a measurement interval of 100 meters. An additional 19% of locations show acceptable crack percentages, nearing the threshold for inspection work. Only 37% of the surface has a cracking level below 10% and is considered satisfactory. In as many as five inspected locations, the percentage indicating unsatisfactory conditions exceeds 80%. These findings indicate the necessity of proper cracking repair and maintenance to prevent further road deterioration and potential hazards.

The analysis of transverse irregularities, i.e., road rutting, indicates that 97% of the inspected locations have minimal irregularities, suggesting excellent road surface conditions in terms of rut depth. Only 3% of locations fall within an acceptable range, with rut depths up to 10mm. These results demonstrate that the majority of the inspected roads maintain satisfactory transverse evenness. Among the inspected locations, all but one ("D3 Varaždin, intersection") show excellent transverse evenness, with 100% of locations exhibiting satisfactory or acceptable rut depths. Given that it is about black spots, our results somehow contradict the findings of previous studies which revealed positive correlation between rut depth on road and traffic accidents, suggesting that more parameters should be taken into account when relating to the risk of traffic accidents. For example, Cenek et al. conducted study to develop statistical models to predict the correlation between the rut depth and traffic accidents with fatalities and injured participants on the New Zealand state road network. The results showed that the accident rate increased when the rut depth was 10 mm or more (Cenek, Henderson, Forbes, Davies, & Tait, 2014). Tsubota et al. demonstrated the relationships between the road surface conditions and accident risk through statistical analysis based on the Poisson regression model (Tsubota, Fernando, Yoshii, & Shirayanagi, 2018). The model estimation results suggested that the rut depth has significant impact on increasing the accident risk, particularly under rain conditions. Accumulation of water on the rut surfaces reduces the skid resistance and increase the effect of hydroplaning. Other authors suggested the rutting on highways should be monitored more carefully than the IRI and cracking because rutting which is not maintained can lead to cracking and disintegration from the road surface structure.

Certainly, traffic accidents are caused by more than one factor such as the accident location characteristics, the road condition inclusive of road types and surface condition, driver's behavior and also weather condition. While traffic accidents are influenced by multiple factors, the condition of the road and its surface plays minor but very important role in their occurrence. Pembuain et al. found in their study that using some RISM tools such as Road Safety Audit (RSA) conducted at the planning, design, construction and early stage of road operation and Road Safety Inspection (RSI) conducted on existing roads has been proven to reduce the rate of traffic accidents by 10-25% and 1-20% respectively (Pembuain A, 2019). The road surfaces wear and tear due to various factors, such as traffic, weather, and ground conditions. Ruts, cracking, and unevenness on the road surface compromise driving comfort and pose potential traffic hazards. Water accumulation in ruts increases the risk of aquaplaning, while cracks and ruts may make it challenging to maintain steady vehicle control. Additionally, large holes in the road surface can cause vehicle damage and potentially lead to the loss of vehicle control. However, the precise impact of road surface conditions on traffic accidents remains to be seen, as well as the extent to which the road surface standard influences accidents. On the other hand, enhancing road surface friction has proven effective in reducing the number of accidents, particularly on wet roads, in sharp bends, and in areas with low initial friction levels. Studies indicate that road surface friction is more significant in accident rates than unevenness. Therefore, prioritizing measures to improve road surface friction can significantly enhance road safety and minimize accident risks in critical driving conditions (Elvik, Høye, Vaa, & Sørensen, 2009). In addition, one study found that resurfacing and a new asphalt layer reduce injuries by 20% in unpopulated areas and 10% in populated areas, and that better skid resistance reduced traffic accidents on wet roads by 40% (Cairney P, 2008).

In summary, based on the obtained results, it is evident that a clear link between the occurrence of traffic accidents and road surface deterioration cannot be established. However, based on the measurements and previous research, it can be concluded that it is possible that on the black spots that were analyzed, there was an influence of bad macrotexture on the occurrence of traffic accidents. Further in-depth analysis is necessary to develop statistical models that establish this connection conclusively. By linking the results to specific black spots, targeted interventions can be planned and implemented to improve road safety on road network.

Future research includes linking the measurements to other factors, such as vehicle speed and weight and the accident type, further to understand the impact of road conditions on road safety. Further, the scientific perspective lies in the continuous improvement of road maintenance and safety through data-driven interventions, i.e., the development of various statistical models that can suggest the correlation between road surface conditions or predict factors that influence road surface deterioration. One of the advantages of regular road surface inspection is the possibility of practical application by road authorities. Continued efforts to monitor road surface conditions and prioritize roads or road parts for maintenance are crucial for ensuring smooth traffic flow and reducing traffic accidents. Limitations and shortcomings of the inspection should also be noted, such as potential measurement errors and the subjective nature of some distress parameters.

6 Conclusion

Based on the road surface inspection within the "Rehabilitation of 23 Dangerous Locations" project using a special vehicle equipped with the Hawkeye 2000 ACD system, the Department of Traffic Signaling of the Faculty of Transport and Traffic Sciences in Zagreb, Croatia, inspected and processed

data for 23 dangerous locations on the state road network in Croatia, including a total length of 70 kilometers. The results of the road surface inspection provide valuable information for prioritizing road maintenance efforts and improving road safety on state roads in the Republic of Croatia, particularly on the inspected locations that had previously been identified as black spots. Most inspected locations require immediate attention to address the unsatisfactory road surface conditions, particularly regarding the macrotexture. The data also indicate successful maintenance efforts in maintaining satisfactory transverse evenness in most locations. Addressing the identified issues is essential to ensure the road network safety and efficiency and prevent accidents and potential road hazards. Analyzing collected data on road surface conditions helps optimize strategies and systems for maintenance activities to enhance road safety and minimize the risk of traffic accidents.

References

Ali, A., Hossain, K., Hasmana, H., Safiuddin, M., Bazan, C., & Hussein, A. (2018). Field inspection and classification of pavement distresses of St. John's city in Newfoundland Canada. *In the 7 th International Materials Specialty Conference, Canadian Society for Civil Engineering.* Retrieved from https://www.researchgate.net/publication/326096894_Field_Inspection_and_Classification_of_Pavement_Distress_of_St_John's_City_in_Newfoundland_Canada

Austroads technical report: *Road surface characteristics and crash occurrence: A literature review.* (2008). Sydney. Retrieved fromhttps://www.academia.edu/25181046/Road_Surface_Characteristics_and_Crash_Occurrence_A_Literature_Review

Buddhavarap, P., Banerjee, A., & Prozzi, J. A. (2013). Influence of pavement condition on horizontal curve safety. *Accident Analysis & Prevention,* 52, 9-18. doi:https://doi.org/10.1016/j.aap.2012.12.010

Cairney, P., & Bennett, P. (2008). Relationship between road surface characteristics and crashes on Victorian rural roads. *ARRB Conference.* Adelaide, South Australia, Australia. Retrieved from https://www.scribbr.com/citation/generator/folders/XMYaENk6ItCWbFbOZJMih/lists/5jfCfddfSammvRCUPeQuRr/

Cairney, P. (2000). Road Surfacing Revisited – a new look at an old countermeasure. Vermont South. Retrieved from http://acrs.org.au/files/arsrpe/RS050013.pdf

Cairney, P. (2005). Road surfacing revisited: a new look at an old countermeasure. *ARRB Conference*.

Cenek, P. D., Henderson, R. J., Forbes, M., Davies, R., & Tait, A. (2014). *The relationship between crash rates and rutting.* Wellington, New Zealand. Retrieved from https://www.nzta.govt.nz/assets/resources/research/reports/545/docs/545.pdf

Elvik, R., Høyen, A., Vaa T., & Sørensen M. (2009). The Handbook of Road Safety Measures. U *Road Maintenance* (pp. 335-395). Bingley: Emerald Group Publishing Limited. doi: https://doi.org/10.1108/9781848552517-008

Gothie, M. (2001). Could adherence and road geometry be used to identify the areas of. *Proceedings of the conference Traffic Safety on Three Continents: International conference in Moscow,* (pp. 313-324). Moscow.

Methodology for identification of dangerous locations in road transport network. (2016). Croatian Roads Ltd. Retrieved from https://hrvatske-ceste.hr/uploads/documents/attachment_file/file/93/Metodologija_za_identifikaciju_opasnih_mjesta.pdf

Pembuain, A., Priyanto, S., & Suparma, L. B. (2019). The effect of road infrastructure on traffic accidents. *Proceedings of the 11th Asia Pacific Transportation and the Environment Conference (APTE 2018).* Retrieved from https://www.researchgate.net/publication/336726264_The_Effect_of_Road_Infrastructure_on_Traffic_Accidents

Regulation on Amendments and Supplements to the Regulation on Road Maintenance. (January 2021). Official Gazette. Ministry of the Sea, Transport and Infrastructue. Retrieved from https://narodne-novine.nn.hr/clanci/sluzbeni/2021_01_3_62.html

Road Traffic Safety Bulletin *2021.* (2022) Zagreb: Ministry of Internal Affairs of the Republic of Croatia. Retrieved from https://mup.gov.hr/UserDocsImages/statistika/2022/Bilten_o_sigurnosti_cestovnog_prometa_2021.pdf

Roe, P., Webster, D., & West, G. (1991). *The relation between the surface texture of roads and accidents.* Transport Road Research Laboratory.

Šimun, M., & Rukavina, T. (2009). Criterion of longitudinal flatness of asphalt pavement driving surface. Građevinar, 1143-1152.

Tsubota, T., Fernando, C., Yoshii, T., & Shirayanagi, H. (2018). Effect of Road Pavement Types and Ages on Traffic Accident Risk. *Transportation Research Procedia, 34,* 211-218. doi:10.1016/j.trpro.2018.11.034

Vinayakamurthy, M., Mamlouk, M. Underwood, S. & Kaloush, K. (2017). Effects of pavement condition on accident rate. Retrieved from https://mti.umd.edu/sites/mti.umd.edu/files/documents/NTC2016-SU-R-2%20Michael%20Mamlouk.pdf

Zovak, G., & Brčić, D. (2014). Analysis of road black spots identification method in Republic of Croatia. *IX International Conference Road Safety in Local Communities.* Retrieved from https://www.croris.hr/crosbi/publikacija/prilog-skup/610572

World Health Organization. *Road Traffic Injuries* (2022) Retrieved from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Digital Transformation and Corporate Culture

Iva Gregurec, Miroslav Mandić, Maja Zeljak
Examining the Ability of Users in the Croatian Market to Distinguish Organic and Sponsored Search Results

Rok Grgec, Leo Mršić, Mladen Konecki
Automated Analysis of Short Digital Messages with the Potential of Application in Business

Tea Mijač
Student-Centered Digital Services: Assessing User Experience in Higher Education

Snježana Križanić, Neven Vrček
Business process modeling and digital twins: a literature review

Nikolina Posarić, Damjan Ćorić
Analysis of Organizational Culture Using Cameron and Quinn's Typology: A Case Study of an ICT Organization

Nikolina Žajdela Hrustek, Renata Mekovec, Valentina Kirinić
E-Schools project - Framework for Digital Competencies of School Principals: Competencies in the Area Digital infrastructure

Examining the Ability of Users in the Croatian Market to Distinguish Organic and Sponsored Search Results

Iva Gregurec
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
iva.gregurec@foi.unizg.hr

Miroslav Mandić
University of Zagreb
Faculty of Economics and Business
Trg J.F. Kennedy 6, 10000 Zagreb, Croatia
mmandic@efzg.hr

Maja Zeljak
MZ Solutions
Jarebička 5, 10000 Zagreb, Croatia
maja@ mzsolutionsagency.com

Abstract. *Search engines generate most of their income through sponsored (paid) search results that appear as answers to user queries. Due to the goal of increasing profits, search engines try to reduce the ability of users to distinguish between sponsored and organic (unpaid) results. This is primarily visible in the design and structure of sponsored search results in the form of ads, which appear very similar to organic results. Although there are studies on other markets, on the Croatian market is insufficiently covered by existing research. Therefore, the aim of this paper is to investigate how much are users in the Croatian market aware of the difference between sponsored and organic search results, and how this knowledge influences their behavior based on search results.*

Keywords. search engines, search engine optimization, search engine advertising, search engine results page (SERP), organic search results, sponsored (paid) search results

1 Introduction

Search engines play an important role in today's society (Schultheiß & Lewandowski, 2020) they represent the most widespread services on the Internet (Purcell et al., 2012) and the most used way to locate websites (K. Li et al., 2014) a gateway used for the purpose of searching for information, for navigating from one website to another, for shopping online or promoting (Y. Li et al., 2018) (Gupta & Mateen, 2014). Search engines are indispensable for finding relevant content and products from the vast array of options available on the web (Burguet et al., 2015). So they represent a crucial link in connecting providers and users (Baye et al., 2016). For this reason, reliability in the search results is crucial because it is the search terms that determine the information that users access on the Internet (Ma et al., 2010). Therefore, it is extremely important for companies to ensure good visibility of search results on search engines (Kritzinger & Weideman, 2013) (Sabwa & Weideman, 2010). Companies have at their disposal different methods of search engine marketing (SEM), which is divided into two main strategies: search engine optimization and search engine advertising or paid search marketing (Krajnović et al., 2019). In order for the company's website to be positioned as high as possible in search engines, optimization (SEO) is carried out (K. Li et al., 2014). It involves the application of various techniques with the aim of increasing organic traffic to the website (Schultheiß & Lewandowski, 2020). On the other hand, paid search, as the name itself suggests, represents paid advertising, i.e. payment for displaying a website on search engines according to certain keywords (Chen et al., 2011) (Terrance et al., 2017). Both categories of increasing website visibility have a common goal, to attract as many users (potential users), i.e. to increase traffic to a certain website (Gašparić et al., 2021). In accordance with the two mentioned strategies, searching through search engines includes at least two categories of results on the search engine result page (SERP): organic and sponsored (paid) results (Höchstötter & Lewandowski, 2009) (Dumais et al., 2010) (Jansen & Resnick, 2006) (Ma et al., 2010) (Gupta & Mateen, 2014) (Yang & Ghose, 2010).

2 Organic and sponsored search results

Search Engine Results Page (SERP) represents a complete page that a search engine displays in response to a user's search query, i.e. a page that displays ads and other informational content in the form of links when the user queries the search engine (Sharma et al., 2017). It should be noted that the SERP is not identical for every user, it depends on factors such as the user's location, search language, search history, etc. (Fletcher, 2007). The display of results on the SERPs has changed over the years. The simplest SERP model displayed a list of ranked results in response to a user's query. However, when search engines started serving ads, two search results lists emerged, the organic (unpaid) search results and the sponsored (paid) search results (Lewandowski et al., 2018) (Sharma et al., 2017) that help increasing the company's visibility on SERPs (Sabwa & Weideman, 2010).

Search engines, specifically SERPs, create an environment in which the order of content presentation limits users' freedom and flexibility when browsing organic and paid content. The reason for this is that on the SERPs, the content located at the top attracts more attention of users compared to the content located at the bottom of the page (Sharma et al., 2017). For this reason, there are also changes in the SERPs appearance, which significantly affects the ability of users to recognize differences between different categories of results, that is, between sponsored and organic search results. Differentiating the mentioned types of results becomes more and more difficult over the years, considering that their appearance is becoming more and more similar (Lewandowski et al., 2018). In addition, SERPs are increasingly reducing the difference between the listed results by constant changes in the way ads are marked (Schultheiß & Lewandowski, 2021a). And that is exactly what this work is based on. The aim of the paper is to investigate the ability of users to distinguish between sponsored and organic, i.e. paid and unpaid search results.

Organic, unpaid search results appear on SERPs alone or together with sponsored search results (Terrance et al., 2017), and refer to free listings that make up the majority of SERP space (Jerkovic, 2009). They imply search results generated from the web page index of search engines using algorithms and are ranked according to a number of criteria (ranking factors) (Sharma et al., 2017) (Baye et al., 2016). Therefore, ranking in organic search is earned or achieved by adopting certain ranking factors such as keywords or phrases more relevant to a particular term than those ranked below (Sabwa & Weideman, 2010). The method of achieving as high position as possible on search engines is known as search engine optimization (SEO) (Schultheiß & Lewandowski, 2020). One of the definitions describes SEO as “the practice of optimizing web pages in a way that improves their ranking in the organic search results” (K. Li et al., 2014). SEO ranking is driven by the relevance of keywords or phrases driven by a well-designed, user-friendly website architecture and high-quality content that includes descriptive page titles, meta tags, alt attributes, anchor text, search friendly HTML tags and code, optimizing quality content, keyword optimization and link building strategies (Sabwa & Weideman, 2010). Therefore, the company optimizes its website using various methods and tools (harmonizing, editing) in order to appear as high as possible in the search results (Jain, 2013). If done correctly, SEO provides numerous benefits that are long-lasting, such as better quality SERPs, a historical trust factor and a lower cost of ownership (Jerkovic, 2009).

Sponsored search results as the main source of income for search engines (Robu et al., 2009) (Alipov et al., 2014) involve buying ad positions for certain keywords or phrases by advertisers in order to display their ads alongside organic results searches (Park & Agarwal, 2018) (Jansen et al., 2011) (Jansen & Schuster, 2011) (Ghose & Yang, 2009). Thus, sponsored search results refer to ad links that advertisers display when users enters a query (Sharma et al., 2017) (Sabwa & Weideman, 2010). The advertiser pays only when the user clicks on the sponsored link promoted by the advertiser (Strategic Direction, 2015). In sponsored search advertising, ad positions are usually sold through auctions managed by the search engine (Arnold et al., 2013). Advertiser positions are then adjusted based on their bid and quality score as assessed by the search engine's proprietary algorithms (Park & Agarwal, 2018). After acuiring an ad position, advertisers appear on the SERP in an ordered list in response to a user's search for keywords or phrases (Yang & Ghose, 2010) (Simonov et al., 2018). The rank of the ad usually depends on the bid price and a quality score (Jansen & Schuster, 2011) (Jansen et al., 2011). Therefore sponsored search results are a part of online advertising (Bayer et al., 2020) which help companies to be at or near the top of any displayed list (Strategic Direction, 2015). They appear along with organic results on search engines in the quest to capture users' attention (Strategic Direction, 2015), typically above the organic search results, to the right of the organic results or below them (Jansen et al., 2011). Their advantages are that they have the high ability of direct response from users, targeting the right users, control over the placement of context in advertisement, attribution of sales of an individual user to specific advertisement, and the high individual impact of an advertisement on a specific user (Bayer et al., 2020). Also, one of the biggest advantages is that ranking improvements can occur almost instantly (Sabwa & Weideman, 2010) and can be used for seasonal marketing activities or during the whole year.

Organic and sponsored search results cooperate with each other as more search results increase market competition, reduce prices (Burguet et al., 2015) and

design creative ways to satisfy users and increase the quality of the search engine. However, it should also be taken into account that companies invest more and more in sponsored search results to get to the highest possible position on the SERPs more easily and quickly, and less in organic search results, which can lead to cannibalization of organic results (Moshary, 2021).

2.1 Overview of previous research on the ability to distinguish between organic and sponsored search results

Most research conducted on the ability to differentiate search results show similar results. Research conducted in 2006 indicates that respondents feel to some degree confident in knowing how to recognize the difference between paid and unpaid research results. Likewise, respondents are biased towards sponsored search results, that is, in more than 82% of cases, they reviewed organic results first and rated the relevance of sponsored results lower than organic ones (Jansen & Resnick, 2006). The results of the research conducted in 2012 suggest that the respondents do not consciously choose organic results because they are not aware of the differences between the results on SERPs (Kobylanski, 2012). Another research conducted in the USA revealed that the majority of users (91%) indicated that they usually find what they are looking for easily, while 56% were very confident and 37% were somewhat confident in their search abilities (Purcell et al., 2012). The research carried out in the European Union territory came to similar results. (Schultheiß & Lewandowski, 2021b) found in their research that 92% of Internet users stated that they easily find what they are looking for through search engines. Likewise, 69% of respondents declared that they are clear about which search results are sponsored. (Singer et al., 2012) found in their research that users have difficulties when conducting searches, especially when it comes to complicated tasks. The respondents estimated their success in solving complicated tasks much higher than it was shown in the research. The contradiction between users' self-assessment and their real abilities when it comes to using internet search engines was also determined in the research (Lewandowski et al., 2018). The research contained questions about Google's business model, tasks with marking sponsored results and questions about self-assessment of the respondents' search skills. It has been shown that users of search engines have little or no knowledge of how search engines generate income and are unable to reliably distinguish sponsored from organic search results. At the same time, 90.8% of the same respondents rated their search skills as "excellent" or even "perfect" (Singer et al., 2012). Another study found that Internet users have difficulty distinguishing between sponsored and organic results on complex SERPs. Respondents showed difficulties in connecting organic results with the possible impact of SEO and distinguishing them from sponsored results (Schultheiß & Lewandowski, 2021b). Another study by the same authors showed that users with a lower level of knowledge about advertising on Internet search engines are more likely to choose (click on) sponsored search results compared to respondents with a higher level of knowledge. Additionally, respondents with a lower level of knowledge showed less willingness to scroll down the SERPs to organic results. A possible reason for this is the fact that they do not understand the difference between sponsored and organic results. Respondents with less knowledge consider all results to be the same and have no need to look for further results besides those at the top of the SERPs (Schultheiß & Lewandowski, 2021a). The mentioned research confirmed the findings of another study, according which the knowledge about advertising on search engines affects the behavior of users regarding the choice of results. In the same study, it was found that users who are unable to distinguish between sponsored and organic search results select sponsored results approximately twice as often as users who are able to distinguish between search results (Lewandowski, 2017). In a report based on the research of search engine users in the US, it was determined that the majority of respondents do not understand how search engines obtain information or how they rank and prioritize results on SERPs. About half of the respondents had a weak idea of how keywords and their relevance affect the work of search engines. When it comes to distinguishing between sponsored and organic search results, only a few respondents noticed that sponsored results were marked as such, and then only when it was easily noticeable (for example, when the mark was at the top of the SERP). The same respondents stated that they avoid results marked as sponsored because they assume that they do not contain what they are looking for, although they could not always explain exactly why they have this opinion (Marable, 2003).

3 Research

3.1 Research methodology

Considering the actuality of the use of search engines in everyday life and the increase in the use of search engine advertising, as well as the insufficient research of this topic on the Croatian market, which was observed by conducted desk research for the purpose of collecting secondary data, the problem of research is obvious. Regarding the research objectives, it is primarily important to investigate the ability of users to distinguish between organic and sponsored results on search engines and to find out whether this ability affects their selection of search results. Google was used as a representative of search engines because it accounts for over 97,23% of the market share in Croatia (Statcounter GlobalStats, 2023). In order to

collect primary data, quantitative descriptive research was conducted using a structured questionnaire on a purposive sample, an avalanche sample, given that the questionnaire was shared on social networks. The questionnaire was based on research conducted by (Lewandowski, 2017) and was open for completion from June to July 2022. A total of 128 respondents took part in the questionnaire, and 110 of them fully answered all the questions in the questionnaire, and the statistical analysis presented in the paper was performed on their answers.

3.2 Research results

If the demographic characteristics of the respondents are observed (Table 1), it can be concluded that the majority of the respondents were female (65.45%), between the ages of 25 and 35 (40.91%), and respondents who live in Zagreb and its surroundings. It is precisely from this distribution of the sample that the basic limitation of the conducted research can be recognized, apart from the sample which consisted of only 110 respondents, their distribution is not even, so the results cannot be generalized to the Croatian market, but they can certainly be a good basis for future research in this area.

Table 1. Demographic characteristics of the respondents

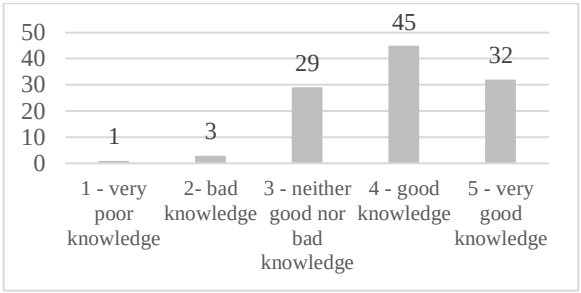
Demographic characteristics		Real number	Frequency
Gender	Male	38	34,55
	Female	72	65,45
Age	18-24	11	10,00
	25-34	45	40,91
	35-44	18	16,36
	45-54	17	15,45
	55-64	14	12,73
	65+	5	4,55
Region	Zagreb and surroundings	98	89,09
	Northern Croatia	2	1,82
	Slavonia	5	4,55
	Lika and Banovina	1	0,91
	Istria, Primorje and Gorski Kotar	4	3,64
	Dalmatia	0	0,00

Source: research

After the questions related to the demographic characteristics of the respondents, the elimination question "Have you used Google to search on the Internet in the last 3 months?" followed. This question served to distinguish users of the search engine Google from those who have not used this search engine in the last three months. Only respondents who answered affirmatively to the mentioned question continued to fill out the questionnaire. 110 of the respondents answered yes to the elimination question, therefore the research in this paper is based on their answers.

After that, the respondents were asked to rate their knowledge when it comes to searching with Google using a 5-point Likert scale. From Graph 1, it can be noted that the largest number of respondents rated their knowledge as good (45 respondents). Based on the average grade of 3.95, it can be concluded that the respondents on average rated their knowledge as good.

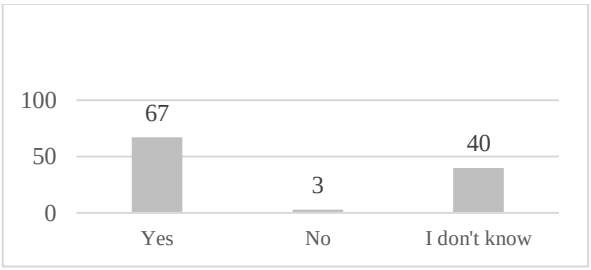
Graph 1. Respondents' knowledge when it comes to searching with Google



Source: research

Then came the question "Is it possible to pay Google to have a particular business appear at the top of the search engine results page (SERPs)?". From Graph 2, it can be seen that 67 respondents answered that it is possible for a company to pay Google to be at the top of the SERPs, 3 respondents answered that it is not possible, while 40 respondents declared that they did not know, i.e. they are not familiar with it.

Graph 2. Is it possible to pay Google to appear at the top of the search engine results page (SERPs)?

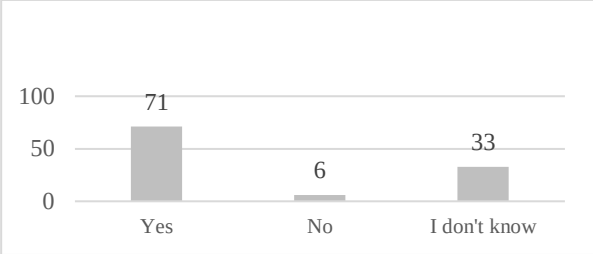


Source: research

To the question "Is it possible to distinguish sponsored results from organic search results on Google?", the largest number of respondents (71 of them) answered that it is possible, 6 respondents answered that it is not possible, while 33 respondents did not know whether it was possible or not to distinguish sponsored from organic search results (Graph 3).

When the last two questions are considered, it can be concluded that 40 users of search engines do not know that it is possible to pay Google for the preferred positioning on the SERPs, or that they do not know that it is possible on Google to distinguish between organic and sponsored search results (33 of respondents). In these results the space for user education on the one hand, and the use of their knowledge by companies on the other hand can be noted.

Graph 3. Is it possible to distinguish sponsored results from organic Google search results?



Source: research

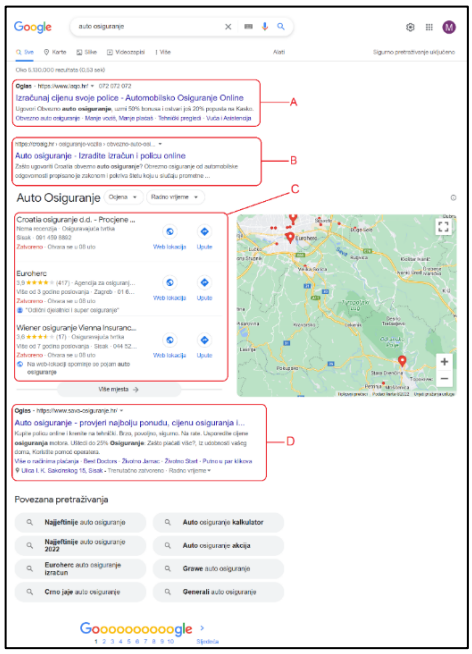
After the question about the search engine, the purpose of which was to determine how the respondents evaluate their knowledge when it comes to using the Google search engine and whether they understand the role of advertising in the Google business model, tasks related to the ability to distinguish between search results followed. The tasks prompted respondents to indicate on screenshots various of SERPs any areas where they believed that sponsored or organic search results appear. Screenshots of SERPs with different elements were used to make them as similar as possible to real search results during the research period. Also, different types of queries were used in the tasks. Search queries are classified according to Broder's classification into informational, navigational, and transactional search (Broder, 2002). Navigation search was exceptionally not used in this research. The distinction between the mentioned types of search is explicitly stated in each individual question, where the respondents were explicitly informed that the search is carried out either for the purpose of purchasing a specific product or for the purpose of informing about a specific topic.

The first question related to information on a certain topic was „Imagine that your car insurance is about to expire and you want to find out about different insurance options on the Internet. After a Google search, the following page was displayed (Figure 1). Check for sponsored search results on the page. If they exist, select each letter with which they are marked.“

The second question related to information on a specific topic assumed the situation that users on Google are looking for information about the career of fashion designer Vivienne Westwood and that they get search results that are shown to them with an image as part of the question. Respondents were asked to study existing ads, i.e. sponsored search results (Figure 2) and, if there are any, to select each letter that identifies them.

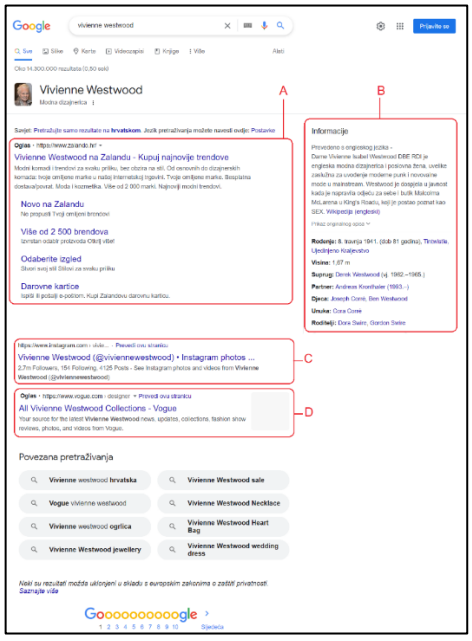
The third question was related to the search for a specific product. The picture shows respondents a page with results related to the search term „sunglasses“. Respondents were asked to select a letter next to each result for which Google was not paid, i.e. to select organic results (Figure 3).

Figure 1. Car insurance search results



Source: research

Figure 2. Vivienne Westwood search results

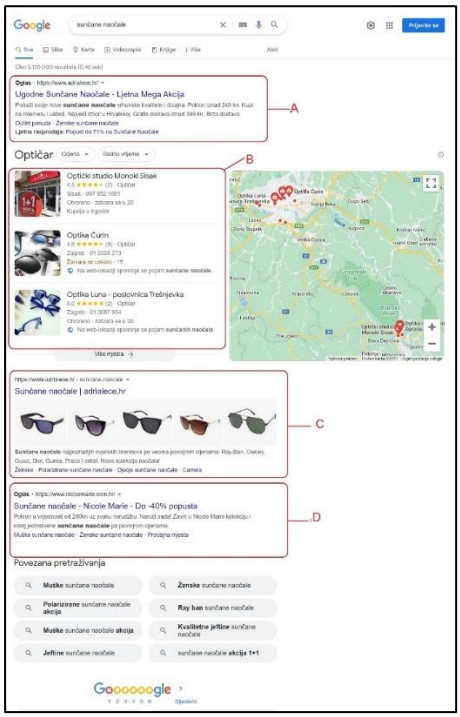


Source: research

The last question with the displayed images was related to the purchase of a certain product and the

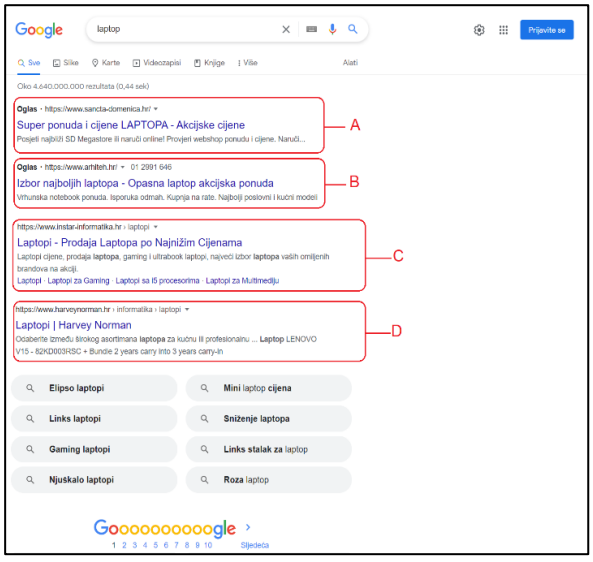
respondents were asked to imagine the following situation: „You want to buy a laptop, so you typed the term "laptop" into the Google search engine. You get the following search results (Figure 4). Please select all letters next to areas that Google did not get paid for (organic results)“.

Figure 3. Sunglasses search results



Source: research

Figure 4 Laptop search results



Source: research

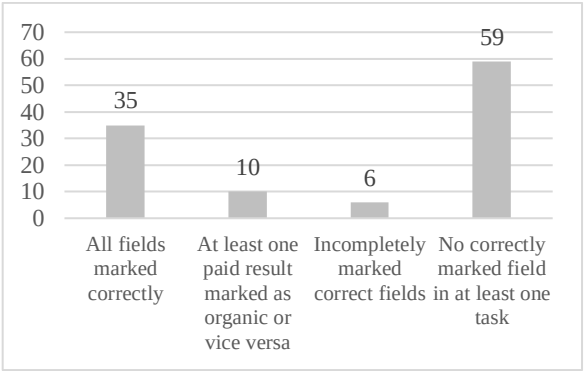
The results related to the tasks set before the respondents can be summarized as follows (Graph 4). Looking at all tasks together, only 32% of respondents correctly identified all required fields (35 respondents). The other 68% of respondents had at least one error when marking the results, and they were as follows:

- In at least one task, 10 respondents incorrectly labeled a sponsored result as organic or vice versa.
- 6 respondents correctly marked a certain type of a result that was requested of them (organic or sponsored result), but did not mark all possible correct answers of that type.
- The remaining 59 respondents were not able to correctly identify any required result in at least one task. Of the above 59 respondents, 7 of them did not correctly mark any the required fields in any task.

Taking all tasks into account, the results show that the majority of respondents (in practice) are unable to distinguish sponsored from organic Google search results.

In the last part of the research, the respondents were asked one more question: „Imagine that you want to prepare squid and you use Google to search for recipes. Your search query is „squid recipe“. Which of the following results would you click on spontaneously? Select the letter the result is labeled with“. In this question a custom SERP screenshot (Figure 5) was used that contained only organic results to provide respondents with the most relevant results. Then, the first two search results were marked as ads, and thus a modified SERP view was offered to respondents.

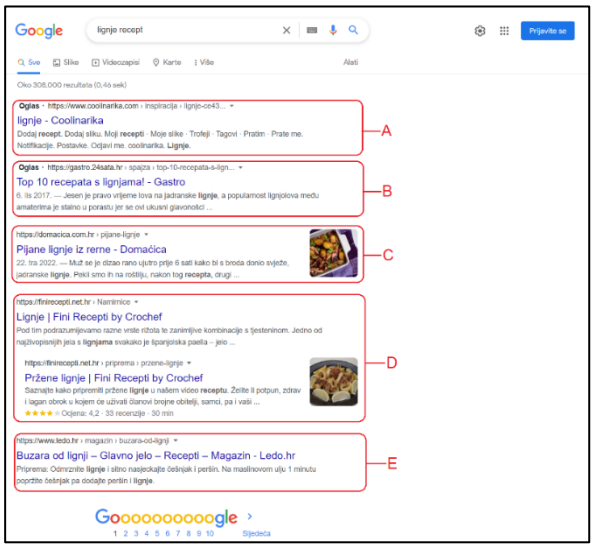
Graph 4. Distribution of correct and incorrect answers in questions with the tasks of recognizing organic and sponsored results



Source: research

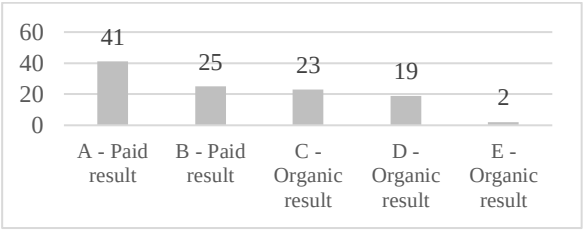
When choosing results, the majority of respondents (66 of them) chose sponsored search results (letters A or B), while 44 respondents chose organic search results (letters C or D). From the results (Graph 5) it is evident that the number of users who selected individual results decreases significantly the more the result is located at the bottom of the displayed SERP.

Figure 5. Squid recipe search results



Source: research

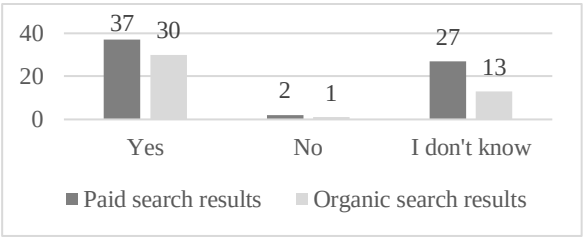
Graph 5. Respondent's choice of sponsored or organic search results



Source: research

When observing the connection between users' knowledge about the possibility of paying for a preferred position on the Google search engine and their choice of sponsored or organic results (Graph 6), it is evident that there is a significant connection between users' knowledge (or the lack of it) and their choice of sponsored search results.

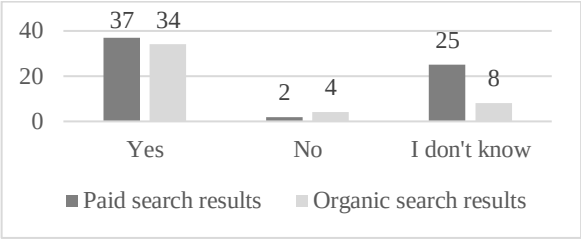
Graph 6. Comparison of user knowledge about the possibility of paying for a preferred position on Google and the choice of an organic or sponsored search result



Source: research

The relationship between users' awareness of their ability to distinguish between sponsored and organic search results and their choice between sponsored and organic search results is clearly worth exploring. In this research, a significant difference can be seen in the responses of users who answered the question positively from those who answered negatively, i.e. that they did not know the answer to the question.

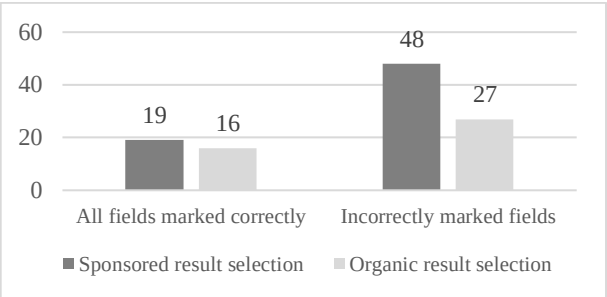
Graph 7. Comparison of user awareness about their ability to distinguish sponsored from organic Google search results and the choice of an organic or sponsored search result



Source: research

Next, in this research the task-based abilities were studied, i.e. the relationship between subjects who demonstrated their ability to consistently distinguish sponsored from organic search results in the task portion of the test and subjects who did not demonstrate this ability. A significant difference in sponsored search results selection behavior was found between users who demonstrated the ability to distinguish sponsored from organic search results versus those who did not (see Graph 8).

Graph 8. Comparison of the user's ability to distinguish between organic and sponsored search results and the choice of an organic or sponsored search result



Source: research

4.3 Research conclusion

The presented results confirm earlier findings of (Purcell et al., 2012) that users have a high level of self-confidence regarding their ability to use Internet search engines and that they are convinced that their skills are sufficient to use search engines effectively. Namely, 77 respondents out of 110 rated their ability to use Google very highly. However, the results show that most respondents are unable to distinguish clearly and with the same high confidence sponsored from organic search results. The task-based part of the questionnaire

showed that only 35 respondents were able to correctly choose all the required boxes. The part of the research based on general questions showed that there is a certain amount of misunderstanding of certain aspects of Google advertising, i.e. that 40 respondents are not aware of the fact that it is possible to pay Google for a certain position on the SERPs and that it is not possible to distinguish sponsored from organic search results (33 respondents).

The results further show that those users who are not aware of the difference between the search results choose the sponsored search results about twice as often. However, it should be emphasized that ads (sponsored search results) can be useful in satisfying the user's need to search for specific information. Therefore, not distinguishing between sponsored and organic results is not always necessarily the cause of a bad user experience, for example in cases where ads lead users to relevant content. However, while choosing a sponsored search result should not be characterized as a bad thing, choosing a sponsored search result without the knowledge of that, could result with bad experience. Some, surprising results were discovered in this research. Namely, it was discovered that users who, according to their own statements, do not know that it is possible to pay Google for a certain position on the SERPs, i.e. that it is not possible to distinguish sponsored search results from organic search results, choose sponsored search results much more often. However, it was found that even users who know about these options also choose sponsored search results more often.

Possible reasons why these respondents choose sponsored search results more often are that they do so intentionally, considering that the ads may be relevant to their search, or that they trust the way Google ranks its search results. The same results were obtained in the part of the research with tasks. The group that prove correct answers was able to recognize organic or sponsored results chose organic search results 16% more often, while the group that failed to prove the above ability chose sponsored search results even 44% more often compared to organic.

The results presented in this paper indicate the need for search engines and advertisers to establish clearer regulations and presentations of their results, i.e. more transparency of their search algorithms. The research revealed a lack of users' ability to distinguish sponsored from organic search results, given that the majority of respondents included in this research are not able to distinguish between the results. The same respondents chose sponsored search results twice as often compared to other respondents. The inability to differentiate results is thought to be caused by the high similarity in the appearance of sponsored and organic search results. Although sponsored search results are marked as such, it can be concluded that the current marking method is insufficiently efficient. The lack of ability to distinguish sponsored from organic search results is a problem when users select sponsored search results under the mistaken assumption that those are organic search results.

4.4 Research limitations and recommendations for future research

This research has certain limitations, the most obvious of which is related to the demographic characteristics of the respondents. Although the survey includes respondents from almost all groups, their distribution is not even. Also, the sample of only 110 respondents is too small to generalize the results to the entire population of Croatia. In future research, the limitation of this research should be eliminated, which can be a good foundation, as well as exploring the difference between the use of search engines on different devices.

5 Conclusion

This paper presents research on the ability to distinguish between sponsored (paid) and organic search results of users of the search engine Google in order to get insights into the awareness of users on how this is impacting their behavior in reviewing the results. Considering the small number of such researches on the Croatian market, the contribution of the work is notable. The research was conducted on the basis of the respondents' answers to the questions in the questionnaire and on the basis of their solutions to the tasks with recognition of the stated results. Additionally, a test question sought to find out whether users' ability to distinguish sponsored from organic search results affects their choice of a certain type of search result. The results of the research show that Croatian users lack certain knowledge when it comes to Google search results, i.e. that there are differences in behavior regarding the choice of sponsored or organic search results between users who were able to accurately distinguish the search results and those users who do not have this ability. Users who did not demonstrate the ability to distinguish between sponsored and organic search results selected sponsored results approximately twice as often as users who successfully recognized sponsored results. The mentioned research results show that users' trust in Google is by default, even if it may be unfounded. If users are unable to distinguish sponsored results from organic ones, and are not aware of this inability, they will consider the ads as credible content. Although the sponsored (paid) search results may be relevant to the user's query, the users should be more aware that the messages they communicate are always biased, i.e. favorable to the interests of the advertiser. From this it can be concluded that Google's income comes not only from advertising but from advertising that is largely mistaken for organic content. Therefore, the results of the conducted research clearly indicate the need for better regulation of the display of sponsored results on

the one hand and user education on the other, i.e. more transparency of their search algorithms.

References

Alipov, V., Topinsky, V., & Trofimov, I. (2014). On peculiarities of positional effects in sponsored search. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1015–1018. https://doi.org/10.1145/2600428.2609498

Arnold, M. A., Darmon, E., & Pénard, T. (2013). *To Sponsor or Not to Sponsor: Sponsored Search Auctions with Organic Links* (SSRN Scholarly Paper No. 2445921). https://doi.org/10.2139/ssrn.2445921

Baye, M. R., De los Santos, B., & Wildenbeest, M. R. (2016). Search Engine Optimization: What Drives Organic Traffic to Retail Sites? *Journal of Economics & Management Strategy*, *25*(1), 6–31. https://doi.org/10.1111/jems.12141

Bayer, E., Srinivasan, S., Riedl, E. J., & Skiera, B. (2020). The impact of online display advertising and paid search advertising relative to offline advertising on firm performance and firm value. *International Journal of Research in Marketing*, *37*(4), 789–804. https://doi.org/10.1016/j.ijresmar.2020.02.002

Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum*, *36*(2), 3–10. https://doi.org/10.1145/792550.792552

Burguet, R., Caminal, R., & Ellman, M. (2015). In Google we trust? *International Journal of Industrial Organization*, *39*, 44–55. https://doi.org/10.1016/j.ijindorg.2015.02.003

Chen, C.-Y., Shih, B.-Y., Chen, Z.-S., & Chen, T.-H. (2011). The exploration of internet marketing strategy by search engine optimization: A critical review and comparison. *African Journal of Business Management*, *5*(12), 4644–4649.

Dumais, S. T., Buscher, G., & Cutrell, E. (2010). Individual differences in gaze patterns for web search. *Proceedings of the Third Symposium on Information Interaction in Context*, 185–194. https://doi.org/10.1145/1840784.1840812

Fletcher, W. H. (2007). Concordancing the web: Promise and problems, tools and techniques. In *Corpus Linguistics and the Web* (pp. 25–45). Brill. https://doi.org/10.1163/9789401203791_004

Gašparić, K., Gregurec, I., & Dobrinić, D. (2021). PRIMJENA OPTIMIZACIJE U ONLINE TRGOVANJU - PRIMJER PODUZEĆA U SJEVERNOJ HRVATSKOJ. *Radovi Zavoda za znanstveni rad Varaždin*, *32*, 59–78. https://doi.org/10.21857/y54jofk0dm

Ghose, A., & Yang, S. (2009). An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Management Science*, *55*(10), 1605–1622.

Gupta, A., & Mateen, A. (2014). Exploring the factors affecting sponsored search ad performance. *Marketing Intelligence & Planning*, *32*(5), 586–599. https://doi.org/10.1108/MIP-05-2013-0083

Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences*, *179*(12), 1796–1812. https://doi.org/10.1016/j.ins.2009.01.028

Jain, A. (2013). The Role and Importance of Search Engine and Search Engine Optimization. Vol. 2 (3). *International Journal of Emerging Trends of Technology If Computer Science*, *2*(3).

Jansen, B. J., & Resnick, M. (2006). An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce Web searching. *Journal of the American Society for Information Science and Technology*, *57*(14), 1949–1961. https://doi.org/10.1002/asi.20425

Jansen, B. J., & Schuster, S. (2011). Bidding on the Buying Funnel for Sponsored Search and Keyword Advertising. *Journal of Electronic Commerce Research*, *12*(1), 1–18.

Jansen, B. J., Sobel, K., & Zhang, M. (2011). The Brand Effect of Key Phrases and Advertisements in Sponsored Search. *International Journal of Electronic Commerce*, *16*(1), 77–106. https://doi.org/10.2753/JEC1086-4415160103

Jerkovic, J. I. (2009). *SEO Warrior* (1st edition). O'Reilly Media.

Kobylanski, A. (2012). Search Engine Advertising (SEA) Or Organic Links: Do Customers See The Difference? *Journal of Business & Economics Research (JBER)*, *10*(3), Article 3. https://doi.org/10.19030/jber.v10i3.6858

Krajnović, A., Sikirić, D., & Hordov, M. (2019). *Digitalni marketing—Nova era tržišne komunikacije*. Sveučilište u Zadru, Odjel za ekonomiju. https://ekonomija.unizd.hr/obavijesti/view/digitalni-marketing-nova-era-trzisne-komunikacije

Kritzinger, W. T., & Weideman, M. (2013). Search Engine Optimization and Pay-per-Click Marketing Strategies. *Journal of Organizational Computing and Electronic Commerce*, *23*(3), 273–286. https://doi.org/10.1080/10919392.2013.808124

Lewandowski, D. (2017). Users' Understanding of Search Engine Advertisements. *Journal of*

Information Science Theory and Practice, *5*(4), 6–25. https://doi.org/10.1633/JISTAP.2017.5.4.1

Lewandowski, D., Kerkmann, F., Rümmele, S., & Sünkler, S. (2018). An empirical investigation on search engine ad disclosure. *Journal of the Association for Information Science and Technology*, *69*(3), 420–437. https://doi.org/10.1002/asi.23963

Li, K., Lin, M., Lin, Z., & Xing, B. (2014). Running and Chasing – The Competition between Paid Search Marketing and Search Engine Optimization. *2014 47th Hawaii International Conference on System Sciences*, 3110–3119. https://doi.org/10.1109/HICSS.2014.640

Li, Y., Yuan, Z., Li, Y., & Liu, J. (2018). Factors influencing search engine usage behavior. *Social Behavior and Personality: An International Journal*, *46*(1), 1–10. https://doi.org/10.2224/sbp.6211

Ma, Z., Pant, G., & Sheng, O. R. L. (2010). Examining Organic and Sponsored Search Results: A Vendor Reliability Perspective. *Journal of Computer Information Systems*, *50*(4), 30–38. https://doi.org/10.1080/08874417.2010.11645428

Marable, L. (2003). *Consumer Reaction to Learning the Truth About How Search Engines Work: Results of an Ethnographic Study*. Consumer WebWatch.

Moshary, S. (2021). *Sponsored Search in Equilibrium: Evidence from Two Experiments* (SSRN Scholarly Paper No. 3903602). https://doi.org/10.2139/ssrn.3903602

Park, C. H., & Agarwal, M. K. (2018). The order effect of advertisers on consumer search behavior in sponsored search markets. *Journal of Business Research*, *84*, 24–33. https://doi.org/10.1016/j.jbusres.2017.11.003

Purcell, K., Brenner, J., & Rainie, L. (2012). *Even though online Americans are more satisfied than ever with the performance of search engines, strong majorities have negative views of personalized search results and targeted ads*.

Robu, V., La Poutré, H., & Bohte, S. (2009). The Complex Dynamics of Sponsored Search Markets. In L. Cao, V. Gorodetsky, J. Liu, G. Weiss, & P. S. Yu (Eds.), *Agents and Data Mining Interaction* (pp. 183–198). Springer. https://doi.org/10.1007/978-3-642-03603-3_14

Sabwa, J., & Weideman, M. (2010). Paid search engine versus organic results: Young consumer preferences. *Proceedings of the 12th Annual Conference on World Wide Web Applications*. ANNUAL CONFERENCE ON WORLD WIDE WEB APPLICATIONS, Durban; South Africa.

Schultheiß, S., & Lewandowski, D. (2020). "Outside the industry, nobody knows what we do" SEO as seen by search engine optimizers and content providers. *Journal of Documentation*, *77*(2), 542–557. https://doi.org/10.1108/JD-07-2020-0127

Schultheiß, S., & Lewandowski, D. (2021a). How users' knowledge of advertisements influences their viewing and selection behavior in search engines. *Journal of the Association for Information Science and Technology*, *72*(3), 285–301. https://doi.org/10.1002/asi.24410

Schultheiß, S., & Lewandowski, D. (2021b). Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference. *Journal of Information Science*, 01655515211014157. https://doi.org/10.1177/01655515211014157

Sharma, D., Gupta, A., Mateen, A., & Pratap, S. (2017). Making sense of the changing face of Google's search engine results page: An advertiser's perspective. *Journal of Information, Communication and Ethics in Society*, *16*(1), 90–107. https://doi.org/10.1108/JICES-06-2017-0035

Simonov, A., Nosko, C., & Rao, J. M. (2018). Competition and Crowd-Out for Brand Keywords in Sponsored Search. *Marketing Science*, *37*(2), 200–215. https://doi.org/10.1287/mksc.2017.1065

Singer, G., Norbisrath, U., & Lewandowski, D. (2012). Ordinary search engine users assessing difficulty, effort, and outcome for simple and complex search tasks. *Proceedings of the 4th Information Interaction in Context Symposium*, 110–119. https://doi.org/10.1145/2362724.2362746

Statcounter GlobalStats. (2023, April). *Search Engine Market Share Croatia*. StatCounter Global Stats. https://gs.statcounter.com/search-engine-market-share/all/croatia

Strategic Direction. (2015). Making sponsored search ads work better: Factors that improve effectiveness. *Strategic Direction*, *31*(1), 12–14. https://doi.org/10.1108/SD-10-2014-0155

Terrance, A. R., Shrivastava, S., & Kumari, A. (2017). Importance of Search Engine Marketing in the Digital World. *Annals of Computer Science and Information Systems*, *14*, 155–158. https://annals-csis.org/proceedings/icitkm2017/drp/24.html

Yang, S., & Ghose, A. (2010). Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, *29*(4), 602–623. https://doi.org/10.1287/mksc.1090.0552

Automated Analysis of Short Digital Messages Content with the Potential of Application in Business

Automatizirana analiza sadržaja kratkih digitalnih poruka s potencijalom primjene u poslovanju

Rok Grgec, Leo Mršić
Algebra University College
Visoko učilište Algebra
Gradišćanska 24, 10000, Zagreb, Croatia
{rok.grgec, leo.mrsic}@algebra.hr

Mladen Konecki
University of Zagreb
Sveučilište u Zagrebu
Faculty of Organization and Informatics
Fakultet organizacije i informatike
Pavlinska 2, 42000 Varaždin, Croatia
mladen.konecki@foi.hr

Abstract*. In this paper the potential business value derived from applying content analysis to short digital messages is presented and elaborated. The WhatsAppNLP library, developed for research purposes, facilitates machine learning model training and demonstrates chat log analysis. By employing natural language processing methods, the library simplifies the analysis process through method calls. WhatsAppNLP leverages the WhatsApp application's data, enabling users to freely download and save their conversations in text format. The Python library processes and analyzes the exported text, generating intuitive reports in the form of graphs, tables, and text messages. The approach allows for automated, anonymized, and ethical analysis of messages, enhancing work efficiency and enabling simultaneous analysis of a large message volume.*

Keywords. WhatsAppNLP, natural language processing, application of content analysis models in business, monetization of content analysis

Sažetak. *U ovom radu predstavljena je i razrađena potencijalna poslovna vrijednost proizašla iz primjene analize sadržaja na kratke digitalne poruke. Knjižnica WhatsAppNLP, razvijena u istraživačke svrhe, olakšava obuku modela strojnog učenja i demonstrira analizu dnevnika razgovora. Koristeći metode obrade prirodnog jezika, knjižnica pojednostavljuje proces analize putem poziva metoda. WhatsAppNLP koristi podatke aplikacije WhatsApp, omogućujući korisnicima da besplatno preuzmu i pohrane svoje razgovore u tekstualnom formatu. Python biblioteka obrađuje i analizira izvezeni tekst, generirajući intuitivna izvješća u obliku grafikona, tablica i tekstualnih poruka. Pristup omogućuje automatiziranu, anonimiziranu i etičku analizu poruka, čime se povećava učinkovitost rada i omogućuje istovremena analiza velike količine poruka.*

Ključne riječi. WhatsAppNLP, obrada prirodnog jezika, primjena modela analize sadržaja u poslovanju, monetizacija analize sadržaja

1 Introduction

Millions of users use some form of communication on the Internet every day. Whether it's using Facebook Messenger, Instagram Direct Messenger, WhatsApp, email etc. With every communication, users leave a trace of their conversation somewhere on the Internet. Most apps that support communication between two or more people store that data in certain databases to allow users to see their own conversations back in time. The same data says a lot about the characteristics of the conversation and users' "internet personality"

1 Uvod

Milijuni korisnika svakodnevno koriste neki od oblika komunikacije na internetu. Bilo da se radi o korištenju Facebook Messengera, Instagram Direct Messengera, WhatsAppa, e-pošte itd. Prilikom svake komunikacije korisnici ostavljaju trag svog razgovora negdje na internetu.

Većina aplikacija koje podržavaju komunikaciju između dvije ili više osoba pohranjuju te podatke u određene baze podataka kako bi korisnicima

that is expressed by using various additions to the conversation such as the use of emojis, the use of media records, the use of voice messages, the average conversation time, the most common conversation period of the day etc.

Users are often not fully aware of these characteristics or are aware to some extent, but in reality, they can be far from what they would tentatively expect. By analyzing these data, a clear and realistic idea of patterns of behavior and conversation can be obtained, which can help in trying to change some of the bad habits in the future or by analyzing, and the main goal is to try establishing facts that represent a business value (Mukhopadhyay, 2018) (Marasović, 2021) (Dinesh et al., 2021) (Sweigart, 2015). Focus of the research is to draw attention and provide platform for commercial application of the presented methods in various industries where data analysis is a message valuable such as: market research (companies can use the system to analyze social media conversations and know/get insights about consumer preferences, mood analysis and trends.

This data can inform the level of product development and be used to manage marketing strategies and manage customer relationships), customer support and service (companies can use the system to automatically analyze customer support records, identify common issues, and generate reports on customer satisfaction, response time, and agent performance; this allows companies to optimize the support process and improve the customer experience), social media tracking (the system can be used to analyze WhatsApp conversations related to brands, products, or specific topics, providing valuable insights into public opinion, mood trends, and customer feedback; this information can guide social media strategies, reputation management, and brand tracking), HR and employee engagement (HR departments can leverage the system to analyze internal employee interviews to assess employee mood, identify areas of improvement, and track engagement levels; this can help organizations increase employee satisfaction, teamwork and productivity) or education (researchers and educational institutions can use the system to analyze WhatsApp chat records in educational activities; this can provide insight into student engagement, learning patterns, and teacher-student interactions, helping to develop effective training strategies), to name some most important ones.

Machine learning is a branch of artificial intelligence that deals with shaping algorithms that improve their efficiency based on empirical data (Walkowska, 2010). A machine learning model defines a product created by programming a computer to adjust the parameters of machine learning based on the given examples. It can define the purpose of using machine learning for predictive purposes or descriptive purposes. Predictive models predict future

omogućile da vide svoje razgovore unatrag u prošlost. Isti podaci dosta govore o karakteristikama razgovora i "internetske osobnosti" korisnika koja se izražava korištenjem raznih dodataka razgovoru kao što su korištenje emojija, korištenje medijskih zapisa, korištenje glasovnih poruka, prosječna vrijeme razgovora, najčešće razdoblje razgovora u danu itd.

Korisnici često nisu u potpunosti svjesni ovih karakteristika ili su ih svjesni u određenoj mjeri, ali u stvarnosti one mogu biti daleko od onoga što bi okvirno očekivali. Analizom ovih podataka može se dobiti jasna i realna predodžba o obrascima ponašanja i razgovora, što može pomoći u pokušaju promjene neke od loših navika u budućnosti ili analizom, a glavni cilj je pokušati utvrditi činjenice koje predstavljaju poslovna vrijednost (Mukhopadhyay, 2018) (Marasović, 2021) (Dinesh et al., 2021) (Sweigart, 2015). Fokus istraživanja je privući pozornost i osigurati platformu za komercijalnu primjenu predstavljenih metoda u raznim industrijama gdje je analiza podataka vrijedna poruka kao što su: istraživanje tržišta (tvrtke mogu koristiti sustav za analizu razgovora na društvenim mrežama i saznati/dobiti uvide o preferencije potrošača, analiza raspoloženja i trendova.

Ovi podaci mogu informirati o razini razvoja proizvoda i koristiti se za upravljanje marketinškim strategijama i upravljanje odnosima s klijentima), korisničku podršku i usluge (tvrtke mogu koristiti sustav za automatsku analizu evidencije korisničke podrške, identificiranje uobičajenih problema i generiranje izvješća o zadovoljstvu kupaca, vrijeme odziva i učinak agenata; to omogućuje tvrtkama da optimiziraju proces podrške i poboljšaju korisničko iskustvo), praćenje društvenih medija (sustav se može koristiti za analizu WhatsApp razgovora povezanih s robnim markama, proizvodima ili određenim temama, pružajući dragocjene uvide u javnost mišljenja, trendovi raspoloženja i povratne informacije kupaca; ove informacije mogu usmjeravati strategije društvenih medija, upravljanje ugledom i praćenje brenda), HR i angažman zaposlenika (HR odjeli mogu iskoristiti sustav za analizu internih razgovora sa zaposlenicima kako bi procijenili raspoloženje zaposlenika, identificirali područja poboljšanja, i pratiti razine angažmana; to može pomoći organizacijama da povećaju zadovoljstvo zaposlenika, timski rad i produktivnost) ili obrazovanje (istraživači i obrazovne ustanove mogu koristiti sustav za analizu WhatsApp chat zapisa u obrazovnim aktivnostima; ovo može pružiti uvid u angažman učenika, obrasce učenja i interakcije učitelj-učenik, pomažući u razvoju učinkovitih strategija obuke), da spomenemo neke najvažnije.

Strojno učenje je grana umjetne inteligencije koja se bavi oblikovanjem algoritama koji poboljšavaju njihovu učinkovitost na temelju empirijskih podataka (Walkowska, 2010). Model strojnog učenja definira proizvod stvoren programiranjem računala za prilagodbu parametara strojnog učenja na temelju

values based on the data they received as a learning base. Descriptive models contain knowledge of data based on a large set, hard-to-read data for people, and generate readable reports based on these invisible trends (Vetulani et al., 2008). Natural language processing is a branch of machine learning that deals with the interaction between computers and humans using natural language.

As an example of the natural language processing, the following situation can be mentioned:

1. Person gives voice command to computer
2. Computer record audio command
3. Converting an audio command to a text command
4. Process data from a text command
5. The data is reconverted to an audio command, this time user-oriented
6. The computer responds to a person with a voice response

Other known examples of using natural language processing (Wonderflow, 2018):

7. Google Translate
8. Check spelling in Word
9. Applications like OK Google, Siri, Cortana, and Alexa

The ultimate goal of natural language processing is to read, decipher and understand human language in a valuable and usable way. Most natural language processing algorithms rely on machine learning to know the meaning derived from human language (Pribisalić, 2019).

The aim of the research presented in this paper was to define robust methodology for automated message analysis, to nominate value propositions to different (most importantly business) applications and to provide framework for future researchers in this domain. Scientific contribution of this paper includes presentation of methods and models used alongside with most important research milestone results. In presented form, research provides an insights related to future development and business users' support expectations management in the domain of of data science.

1.1 Related Work

Library developed for this project fits alongside other community and market libraries like NLTK, GenSim, SpaCy, CoreNLP, TextBlob, AllenNLP, polyglot or scikit-learn to name few examples (Kumar, 2020). However, to maintain simplicity, ease of use, focused applicability and lighweight, the WhatsAppNLP has been developed aiming towards researchers who will continue to use the research results for motivation and novel purposes. There are similar research paths available (Kelemen, 2015) (Rosen et al., 2011) as research papers, but those are not suitable for fast-track application and implementation.

danih primjera. Može definirati svrhu korištenja strojnog učenja u prediktivne ili deskriptivne svrhe. Prediktivni modeli predviđaju buduće vrijednosti na temelju podataka koje su primili kao osnovu za učenje. Deskriptivni modeli sadrže znanje o podacima temeljeno na velikom skupu podataka teško čitljivih za ljude i generiraju čitljiva izvješća na temelju tih nevidljivih trendova (Vetulani et al., 2008). Obrada prirodnog jezika je grana strojnog učenja koja se bavi interakcijom između računala i ljudi koristeći prirodni jezik.

Kao primjer obrade prirodnog jezika može se navesti sljedeća situacija:

1. Osoba daje glasovnu naredbu računalu
2. Računalno snimanje audio naredbe
3. Pretvaranje audio naredbe u tekstualnu naredbu
4. Obrada podataka iz tekstualne naredbe
5. Podaci se ponovno pretvaraju u audio naredbu, ovaj put orijentiranu na korisnika
6. Računalo odgovara osobi glasovnim odgovorom

Ostali poznati primjeri korištenja obrade prirodnog jezika (Wonderflow, 2018):

7. Google prevoditelj
8. Provjerite pravopis u Wordu
9. Aplikacije poput OK Google, Siri, Cortana i Alexa

Krajnji cilj obrade prirodnog jezika je čitanje, dešifriranje i razumijevanje ljudskog jezika na vrijedan i upotrebljiv način. Većina algoritama za obradu prirodnog jezika oslanja se na strojno učenje kako bi saznali značenje izvedeno iz ljudskog jezika (Pribisalić, 2019).

Cilj istraživanja predstavljenog u ovom radu bio je definirati robusnu metodologiju za automatiziranu analizu poruka, nominirati prijedloge vrijednosti za različite (najvažnije poslovne) aplikacije i pružiti okvir budućim istraživačima u ovoj domeni. Znanstveni doprinos ovog rada uključuje prezentaciju metoda i modela koji se koriste uz najvažnije rezultate istraživačke prekretnice. U predstavljenom obliku, istraživanja pružaju uvid u budući razvoj i upravljanje očekivanjima podrške poslovnih korisnika u domeni znanosti o podacima.

1.1 Pristup i dosadašnji radovi

Knjižnica razvijena za ovaj projekt uklapa se uz druge knjižnice zajednice i tržišta kao što su NLTK, GenSim, SpaCy, CoreNLP, TextBlob, AllenNLP, polyglot ili scikit-learn da navedemo nekoliko primjera (Kumar, 2020). Međutim, kako bismo održali jednostavnost, jednostavnost korištenja, usredotočenu primjenjivost i lighweight, razvijena je WhatsAppNLP knjižnica s ciljem prema istraživačima koji će nastaviti koristiti naša istraživanja u motivacijske i drugee svrhe. Postoje slični istraživački putovi dostupni (Kelemen, 2015) (Rosen et al., 2011) u obliku istraživačkih radova, ali isti nisu prikladni za brzu primjenu i rad u poslovnom okruženju.

2 Modeling Topics

In the form of natural language processing, the theme implies a set of words that "go together". This refers to words that associate the user when thinking about a topic, for example "sport", "food", "music". Each of these topics has its own set of words that define it: with "sport" come words such as "football, team play, win, player, swimming, success", along with the theme of "food" words such as "bread, fish, wine, salty, spoiled, eat, drink, cook" and with the theme "music" come words such as "loud, concert, listeners, singing, performance" etc. (Khurana et al., 2022) (Wang et al., 2019) (Vetulani et al., 2000) (Sworna et al., 2022).

A theme modeling algorithm is an algorithm that independently defines topics based on assigned data from a set of data intended for model training (usually a large set of text documents) and groups words that describe it under each defined topic. A new, unknown set of data (words) is then forwarded to the created model, on the basis of which the model independently categorizes the transmitted data into a particular set, i.e. determines which topic these words belong to (Walkowska, 2009) (Hult et al., 2006) (Jagota, 2020).

2.1 Latent Dirichlet Allocation

One of the applicable machine learning methods is Latent Dirichlet Allocation (LDA). The LDA treats the probability p(z|d) as a distribution that depends on a parameter. In other words, the LDA defines themes as the distribution of words from a fixed data set by adding to each topic the probabilities for each word from the submitted dataset. The aim of this method is to find significant word distributions across various topics and significant distributions of topics by various textual documents (textual data sources) and therefore the number of topics is a key parameter of this method of machine learning. LDA is also the most widely used method of machine learning used for the purpose of modeling topics (Hurlow et al., 2003) (Marasović, 2021).

The LDA method consists of two parts: (1) the content/words belonging to the document, which is already known, and (2) the content/words belonging to the topic or the probability of words belonging to the topic, which need to be calculated. The finding algorithm goes through each content and randomly assigns each word in the document to one of the "k" topics ("k" is selected in advance). For each document "d", each word is passed through and the formula p(theme "t" | document "d") calculates the proportion of words in the document "d" that are assigned to the topic "t". The algorithm tends to capture how many words belong to the topic "t" for the given document "d", excluding the current word.

The LDA presents documents as a mix of topics. Similarly, the topic is a mixture of words. If there is a

2 Modeliranje pojmova

U obliku obrade prirodnog jezika, tema podrazumijeva skup riječi koje "idu zajedno". Ovo se odnosi na riječi koje asociraju korisnika kada razmišlja o nekoj temi, na primjer "sport", "hrana", "glazba". Svaka od ovih tema ima svoj skup riječi koje je definiraju: uz "sport" dolaze riječi kao što su "nogomet, timska igra, pobjeda, igrač, plivanje, uspjeh", zajedno s temom riječi "hrana" kao što je "kruh", riba, vino, slano, pokvareno, jesti, piti, kuhati" a uz temu "glazba" dolaze riječi kao što su "glasno, koncert, slušatelji, pjevanje, nastup" itd. (Khurana et al., 2022) (Wang et al., 2019) (Vetulani et al., 2000) (Sworna et al., 2022).

Algoritam za modeliranje teme je algoritam koji neovisno definira teme na temelju dodijeljenih podataka iz skupa podataka namijenjenih obuci modela (obično veliki skup tekstualnih dokumenata) i grupira riječi koje ih opisuju pod svakom definiranom temom. Novi, nepoznati skup podataka (riječi) zatim se prosljeđuje kreiranom modelu, na temelju čega model samostalno kategorizira prenesene podatke u određeni skup, odnosno određuje kojoj temi te riječi pripadaju (Walkowska, 2009) (Hult et al., 2006) (Jagota, 2020).

2.1 Latentna Dirichletova alokacija

Jedna od primjenjivih metoda strojnog učenja je Latent Dirichlet Allocation (LDA). LDA tretira vjerojatnost p(z|d) kao distribuciju koja ovisi o parametru. Drugim riječima, LDA definira teme kao distribuciju riječi iz fiksnog skupa podataka dodavanjem svakoj temi vjerojatnosti za svaku riječ iz poslanog skupa podataka. Cilj ove metode je pronaći značajne distribucije riječi po raznim temama i značajne distribucije tema po različitim tekstualnim dokumentima (izvorima tekstualnih podataka) te je stoga broj tema ključni parametar ove metode strojnog učenja. LDA je također najraširenija metoda strojnog učenja koja se koristi u svrhu modeliranja tema (Hurlow et al., 2003) (Marasović, 2021).

LDA metoda se sastoji od dva dijela: (1) sadržaj/riječi koje pripadaju dokumentu, koji je već poznat, i (2) sadržaj/riječi koje pripadaju temi ili vjerojatnost riječi koje pripadaju temi, što treba biti izračunat. Algoritam za pronalaženje prolazi kroz svaki sadržaj i svaku riječ u dokumentu nasumično dodjeljuje jednoj od "k" tema ("k" je odabrano unaprijed). Za svaki dokument "d" prolazi svaka riječ i formula p(tema "t" | dokument "d") izračunava udio riječi u dokumentu "d" koje su dodijeljene temi "t". Algoritam nastoji uhvatiti koliko riječi pripada temi "t" za dati dokument "d", isključujući trenutnu riječ.

LDA predstavlja dokumente kao mješavinu tema. Slično tome, tema je mješavina riječi. Ako postoji velika vjerojatnost da će se riječ naći u temi, svi dokumenti koji imaju odabranu riječ bit će uže

high probability that the word will be in the topic, all documents that have a selected word will be more closely related to the topic as well. Similarly, if the selected word is not very likely to belong to a topic, documents containing the selected word will also be very unlikely to be in the topic, because the rest of the words in the document will belong to another topic. Example of LDA (Kelemen, 2015) is the following (see sentences below).

1. I love eating fruits and vegetables.
2. For breakfast, I eat an apple and drink fruit juice.
3. Small cats can be disobedient.
4. My sister loves little kittens.
5. Rabbits love to eat carrots.

Sentences 1 and 2: 100% belong to topic A
Sentences 3 and 4: 100% belong to topic B
Sentence 5: 50% belong to topic A, 50% to topic B
Topic A: 30% fruit, 15% vegetables, 10% eat, 10% drink etc. It is possible to conclude that this is about the topic of "food".
Topic B: 30% cats, 30% rabbit, 10% small etc. It is possible to conclude that this is about the topic of "animals".

2.2 Collection and Processing of Data

In order for the exported messages to continue to be processed, the "excess" that directly affects the speed and quality of work of the machine learning model itself has to be removed from them. Each message is written in the form of a string consisting of the date and time of sending the message, the name of the sender of the message and the text of the message itself.

Example of a message record:

```
['6/23/20, 22:38 - Rok: Hahahah\n',
 '6/23/20, 22:38 - Rok: 🤦\u200dđ\n',
 '6/23/20, 22:39 - Matea: 😉😉\n',
 '6/23/20, 22:39 - Matea: GLEDAJ sam si prizano\n',
 '6/23/20, 22:39 - Matea: *priznao\n']
```

Figure 1. Example of an exported chat log record

In the case when the data for further processing in a trained model for modeling topics have to be prepared, only the records of the content of the sent message are needed. The whole process of "cleaning" messages consists of several key steps before further processing:

1. Remove the part of the message that contains information about the time and date of the message and the name of the sender of the message
2. Remove English and Croatian stopping words
3. Remove emojis and media files
4. Remove punctuation marks
5. Convert a word to the same lowercase initial letter

povezani i s temom. Slično tome, ako nije vrlo vjerojatno da odabrana riječ pripada temi, dokumenti koji sadrže odabranu riječ također će vrlo vjerojatno biti u temi, jer će ostatak riječi u dokumentu pripadati drugoj temi. Primjer LDA (Kelement, 2015) je sljedeći (vidi rečenice u nastavku).

1. Volim jesti voće i povrće.
2. Za doručak pojedem jabuku i popijem voćni sok.
3. Male mačke mogu biti neposlušne.
4. Moja sestra voli male mačiće.
5. Kunići vole jesti mrkvu.

Rečenice 1 i 2: 100% pripadaju temi A
Rečenice 3 i 4: 100% pripadaju temi B
Rečenica 5: 50% pripada temi A, 50% temi B
Tema A: 30% voće, 15% povrće, 10% jesti, 10% piti itd. Može se zaključiti da se radi o temi "hrana".
Tema B: 30% mačke, 30% zečevi, 10% mali itd. Može se zaključiti da se radi o temi "životinje".

2.2 Prikupljanje i obrada podataka

Kako bi se eksportirane poruke nastavile obrađivati, iz njih se mora ukloniti "višak" koji izravno utječe na brzinu i kvalitetu rada samog modela strojnog učenja. Svaka poruka je zapisana u obliku niza koji se sastoji od datuma i vremena slanja poruke, imena pošiljatelja poruke i teksta same poruke.

Primjer zapisa poruke:

```
['6/23/20, 22:38 - Rok: Hahahah\n',
 '6/23/20, 22:38 - Rok: 🤦\u200dđ\n',
 '6/23/20, 22:39 - Matea: 😉😉\n',
 '6/23/20, 22:39 - Matea: GLEDAJ sam si prizano\n',
 '6/23/20, 22:39 - Matea: *priznao\n']
```

Slika 1. Primjer izvezenog zapisa dnevnika razgovora

U slučaju kada je potrebno pripremiti podatke za daljnju obradu u obučenom modelu za teme modeliranja, potrebni su samo zapisi sadržaja poslane poruke. Cijeli proces "čišćenja" poruka sastoji se od nekoliko ključnih koraka prije daljnje obrade:

1. Uklonite dio poruke koji sadrži podatke o vremenu i datumu poruke te ime pošiljatelja poruke
2. Uklonite engleske i hrvatske zaustavne riječi
3. Uklonite emojije i medijske datoteke
4. Uklonite interpunkcijske znakove
5. Pretvorite riječ u isto malo početno slovo

The kod metode koja izvodi sve gore navedene korake. Rezultat nakon metode provedene nad prvim dnevnikom chata:

```
['gledaj', 'prizano', 'priznao', 'cudno', 'imma']
```

Slika 2. Primjer 1, Rezultat funkcije čišćenja podataka

The code of the method that performs all of the above steps. The result after the method carried out over the first chat log:

```
['gledaj', 'prizano', 'priznao', 'cudno', 'imma']
```

Figure 2. Example 1, Result of the data cleanup function

The result after the method carried out over the first chat log:

```
['moze', 'medo', 'tkalci', 'isto', 'mogli']
```

Figure 3. Example 2, The result of the data cleanup function

2.3 Translation of Processed Data

After processing, the data should be translated into English because the machine learning model for modeling topics is based on words in English, i.e. it is based on the English dataset. The translation process takes place using Google Sheets technology. In combination with the Python programming language, Google Sheets allows to automate work with tables that possess a built-in method for translating data. An example of a method of translating data from a Google Sheets table cell = GOOGLETRANSLATE (A1, "en"; "en"). The method receives the cell, the source language of the text, and the desired output language of the translated word as input parameters. The translation process takes place in three steps: loading data into Google Sheets, applying the built-in Google Sheets method "googletranslate()" over data, and retrieving translated words from Google Sheet-s tables into a previously created field in the program itself. In order to automate the process of uploading, translating and retrieving translated data, Google requires its users to register and create a Google application programming interface (Google API). All data related to the granting of rights to the user, as well as the assignment of API tokens for access to the Google Sheet application, can be found in service_account.json. The file is used as an input parameter in the pygsheets.authorize (service_account_file = 'service_account.json') method that opens a link to the Google Sheets account where the data translation process will take place. In order to bring the data to Google Sheets, it must first be stored in a .csv format file in order to use the Pandas Python library method: pandas.read_csv(file). After that, using an open link to Google Sheets, the method set_dataframe is used, which receives as input parameters a previously read pandas file, and the coordinates of the fields in the Google Sheets table on which the data from the pandas file will start to be typed. After successfully loading the data, there is a translation method in the cells of the adjacent field "=googletranslate()" which automatically translates

Rezultat nakon metode provedene nad prvim dnevnikom chata:

```
['moze', 'medo', 'tkalci', 'isto', 'mogli']
```

Slika 3. Primjer 2, rezultat funkcije čišćenja podataka

2.3 Prevođenje obrađenih podataka

Nakon obrade podatke je potrebno prevesti na engleski jer se model strojnog učenja za teme modeliranja temelji na riječima na engleskom, odnosno temelji se na engleskom setu podataka. Proces prevođenja odvija se pomoću tehnologije Google Sheets. U kombinaciji s programskim jezikom Python, Google Sheets omogućuje automatizaciju rada s tablicama koje posjeduju ugrađenu metodu za prevođenje podataka. Primjer metode prevođenja podataka iz ćelije tablice Google tablica = GOOGLETRANSLATE (A1, " en "; " en "). Metoda prima ćeliju, izvorni jezik teksta i željeni izlazni jezik prevedene riječi kao ulazne parametre. Proces prevođenja odvija se u tri koraka: učitavanje podataka u Google tablice, primjena ugrađene metode Google tablica " googletranslate ()" nad podacima i dohvaćanje prevedenih riječi iz tablica Google tablica u prethodno kreirano polje u samom programu. Kako bi automatizirao proces učitavanja, prevođenja i dohvaćanja prevedenih podataka, Google od svojih korisnika zahtijeva da se registriraju i kreiraju Google sučelje za programiranje aplikacija (Google API). Svi podaci vezani uz dodjelu prava korisniku, kao i dodjelu API tokena za pristup aplikaciji Google Sheet nalaze se u service_account.json. Datoteka se koristi kao ulazni parametar u metodi pygsheets.authorize (service_account_file = ' service_account.json ') koja otvara vezu na račun Google tablica na kojem će se odvijati proces prevođenja podataka. Kako bi se podaci prenijeli u Google tablice, prvo moraju biti pohranjeni u datoteci.csv formata kako bi se koristila metoda biblioteke Pandas Python: pandas.read_csv (datoteka). Nakon toga se otvorenom poveznicom na Google Sheets koristi metoda set_dataframe koja kao ulazne parametre prima prethodno pročitanu pandas datoteku, te koordinate polja u Google Sheets tablici na kojima će se početi unositi podaci iz pandas datoteke. biti upisan. Nakon uspješnog učitavanja podataka, u ćelijama susjednog polja postoji metoda prevođenja "= googletranslate ()" koja automatski prevodi susjedna polja s prethodno unesenim podacima (Grossz et al., 1986). S obzirom da se radi o velikoj količini podataka koje je potrebno prevesti, potrebno je napraviti pauzu s preuzimanjem podataka kako bi se obavio posao prevođenja nad svim podacima koji su u redu za prevođenje. Određena je pauza od punih pet minuta, što je dovoljno vremena da se s hrvatskog na engleski prevede stotinjak tisuća riječi. Nakon prevođenja podaci se dohvaćaju metodom – get_col () koja kao ulazni parametar dobiva redni broj stupca Google Sheet tablice iz kojeg

adjacent fields with previously entered data (Grossz et al., 1986). Given that this is a large amount of data that needs to be translated, it is necessary to take a break with the download of data in order to perform the job of translation over all data in the queue for translation. The break is set for a full five minutes, which is enough time to translate about one hundred thousand words from Croatian to English. After translation, the data is retrieved by the method – get_col(), which receives as an input parameter the ordinal number of the Google Sheet table column from which all the data contained in that column will be retrieved and stored in the previously defined field "values_list". As a final step in the translation process, the translated data needs to be processed once again. This step is necessary because the translated data is no longer separate words. The "googletranslate()" method translates most Croatian words in such a way as to create unnecessary text, which directly acts on the method of machine learning of topic recognition because it creates a false impression of the topics present. An example of creating unnecessary text is translating the word "I run" from Croatian to English using the googletranslate() method. The word is translated as "I am running", which is originally correct, but in that newly created sentence, the word "I" and the word "am" represent an excess that needs to be removed. The process of cleaning translated words is due to the method of prepare_translated_words_for_model() which receives as an input parameter a field of previously retrieved translated words. The method performs several operations on words: a "lemmatisis" operation that reduces the word to its shape in the dictionary, an operation to remove words with less than three letters, an operation to remove punctuation marks, and an operation to remove English stopping words. After processing the translated data, we get a field of translated words, which is further used as the input parameter of the method for identifying topics of conversation and as an input parameter of the method for obtaining statistical data on the number of most represented words in a conversation.

This approach of converting text documents into a field of separate words is called the "sack of words" approach.

An example of the translated data on the first chat log:

```
['lapsuz', 'free', 'friday', 'tomorrow', 'long']
```

Figure 4. Example 1. translated words

An example of translated data on another chat log:

```
['drink', 'account', 'drink', 'dump', 'team']
```

Figure 5. Example 2. translated words

će se dohvatiti svi podaci sadržani u tom stupcu i pohraniti u prethodno definirano polje " popis_vrijednosti ". Kao posljednji korak u procesu prevođenja, prevedene podatke potrebno je još jednom obraditi. Ovaj korak je neophodan jer prevedeni podaci više nisu zasebne riječi. Metoda " googletranslate ()" prevodi većinu hrvatskih riječi na način da stvara nepotreban tekst, što izravno djeluje na metodu strojnog učenja prepoznavanja tema jer stvara lažni dojam o prisutnim temama. Primjer stvaranja nepotrebnog teksta je prevođenje riječi "I run" s hrvatskog na engleski metodom googletranslate (). Riječ je prevedena kao "trčim", što je izvorno točno, ali u toj novostvorenoj rečenici riječ "ja" i riječ "jesam" predstavljaju višak koji treba ukloniti. Proces čišćenja prevedenih riječi je zahvaljujući metodi priprave_translated_words_for-_model() koja kao ulazni parametar prima polje prethodno dohvaćenih prevedenih riječi. Metoda izvodi nekoliko operacija na riječima: operaciju " lemmatisis " koja reducira riječ na njen oblik u rječniku, operaciju za uklanjanje riječi s manje od tri slova, operaciju za uklanjanje interpunkcijskih znakova i operaciju za uklanjanje engleskih zaustavnih riječi. Nakon obrade prevedenih podataka dobivamo polje prevedenih riječi koje se dalje koristi kao ulazni parametar metode za identifikaciju tema razgovora i kao ulazni parametar metode za dobivanje statističkih podataka o broju najzastupljenijih riječi u razgovor.

Ovaj pristup pretvaranja tekstualnih dokumenata u polje zasebnih riječi naziva se pristup "vreće riječi".

Primjer prevedenih podataka na prvom dnevniku chata:

```
['lapsuz', 'free', 'friday', 'tomorrow', 'long']
```

Slika 4. Primjer 1. prevedene riječi

Primjer prevedenih podataka na drugom dnevniku chata:

```
['drink', 'account', 'drink', 'dump', 'team']
```

Slika 5. Primjer 2. prevedene riječi

2.4. Model strojnog učenja

Postoji nekoliko načina za treniranje modela strojnog učenja. Svaki pristup ima svoju upotrebu, prednosti i nedostatke. Kao metoda uvježbavanja modela strojnog učenja identificiranja tema razgovora korištena je prethodno spomenuta metoda latentne Dirichletove alokacije (LDA). Kao skup tekstualnih podataka (eng. Data set) korišten je skup podataka "20 newsgroup". Skup podataka "20 newsgorup " zbirka je tekstualnih dokumenata koji se sastoje od 18.828 tekstualnih članaka, poruka i e-mailova, od kojih je svaki kategoriziran pod jednim od 20 naslova.

Ovaj skup podataka jedan je od najčešće korištenih skupova podataka u obradi prirodnog jezika, namijenjen eksperimentiranju s metodama strojnog

2.4 Machine Learning Model

There are a number of ways to train machine learning models. Each approach has its own uses, pros and cons. As a method of training machine learning models of identifying topics of conversation, the previously mentioned method of latent Dirichlet allocation (LDA) has been used. As a set of textual data (eng. Data set) the "20 newsgroup" dataset has been used. The "20 newsgorup" dataset is a collection of text documents consisting of 18,828 text articles, messages and e-mails, each of which is categorized under one of 20 headings. This dataset is one of the most commonly used datasets in natural language processing, intended to experiment with machine learning methods of categorizing and grouping textual data. In the case of training the model of identifying topics of conversation, it is one of the best possible choices because it covers a variety of topics that include politics, religion, sports, space and medicine, which directly affects the range of topics that can be recognized by the model in a given set of text messages from the WhatsApp application. The dataset as well as its processing methods and training methods of LDA models are found in the publicly available "Genism" and "Sklearn" Python libraries. The first step in training the LDA theme recognition model is to retrieve the "20newsgroup" dataset. Example:

import gensim from gensim.models import *LdaModel*
from sklearn.datasets import *fetch_20newsgroups*
newsgroups_train =
fetch_20newsgroups(subset='train', shuffle = ***True)***

Given that this is uncleaned data containing text unusable parts, the data needs to be cleaned using the preprocess() method, which receives a text document from a "20newsgroup" file as an input parameter. Pre-processing example of "20newsgroup" dataset:

processed_docs = []
for doc **in** newsgroups_train. Data:
processed_docs. append(preprocess(doc))

The "preprocess()" method performs several operations on data. An operation to remove punctuation marks, an operation to remove stopping words in English, an operation to convert a word into a lowercase initial letter, and a "leminizing" operation. Lemmatizing – converting words to dictionary form). After processing, each sentence in the data set is stored in the form of a separate word field, that is, in the form of a sentence field, each of which is a word field. Example of a field obtained using the preprocess():
The created field is additionally filtered by the extreme filtering method filter_extremes() which receives as input parameters the number of minimal repetitions of a single word, the percentage of the number of words that appear in all documents and the number of words that are retained after the filtering has been done. After učenja kategorizacije i grupiranja tekstualnih podataka. U slučaju treniranja modela identificiranja tema razgovora, to je jedan od najboljih mogućih izbora jer pokriva niz tema koje uključuju politiku, religiju, sport, svemir i medicinu, što izravno utječe na raspon tema koje se mogu koje model prepoznaje u zadanom skupu tekstualnih poruka iz aplikacije WhatsApp. Skup podataka kao i njegove metode obrade i metode obuke LDA modela nalaze se u javno dostupnim Python bibliotekama "Genism" i " Sklearn ". Prvi korak u obuci modela prepoznavanja tema LDA je dohvaćanje skupa podataka "20newsgroup". Primjer:

import gensim iz gensim.modeli import *LdaModel*
iz sklearn.datasets import *fetch_20newsgroups*
newsgroups_train =
fetch_20newsgroups(subset='train', shuffle = ***True)***

S obzirom da se radi o neočišćenim podacima koji sadrže tekst neupotrebljivih dijelova, podatke je potrebno očistiti metodom preprocess(), koja prima tekstualni dokument iz datoteke "20newsgroup" kao ulazni parametar. Primjer prethodne obrade skupa podataka "20newsgroup":

obrađeni_dokumenti = []
za doc _ newsgroups_train. Podaci:
obrađeni_dokumenti.
dodati(pretproces(doc))

Metoda "preprocess()" izvodi nekoliko operacija na podacima. Operacija uklanjanja interpunkcijskih znakova, operacija uklanjanja zaustavnih riječi na engleskom jeziku, operacija pretvaranja riječi u početno malo slovo i operacija " leminizacije ". Lematiziranje – pretvaranje riječi u oblik rječnika). Nakon obrade svaka rečenica u skupu podataka pohranjuje se u obliku zasebnog polja riječi, odnosno u obliku polja rečenice od kojih je svako polje riječi. Primjer polja dobivenog korištenjem preprocess():
Kreirano polje se dodatno filtrira metodom ekstremnog filtriranja filter_extremes () koja kao ulazne parametre prima broj minimalnih ponavljanja pojedine riječi, postotak broja riječi koje se pojavljuju u svim dokumentima i broj riječi koje se zadržavaju nakon filtriranje je obavljeno. Nakon filtriranja, pomoću metode doc2bow(), polje se pretvara u vreću riječi. Bag of words), te je spreman kao ulazni parametar LDA metode za razvoj modela strojnog učenja.

```
[['lerxst', 'thing', 'subject', 'nntp', 'post', 'host', 'organ', 'univers', 'maryland', 'colle
g', 'park', 'line', 'wonder', 'enlighten', 'door', 'sport', 'look', 'late', 'earli', 'call', 'br
icklin', 'door', 'small', 'addit', 'bumper', 'separ', 'rest', 'bodi', 'know', 'tellm', 'model',
'engin', 'spec', 'year', 'product', 'histori', 'info', 'funki', 'look', 'mail', 'thank', 'brin
g', 'neighborhood', 'lerxst'], ['guykuo', 'carson', 'washington', 'subject', 'clock', 'poll', 'f
inal', 'summari', 'final', 'clock', 'report', 'keyword', 'acceler', 'clock', 'upgrad', 'articl',
'shelley', 'qvfo', 'innc', 'organ', 'univers', 'washington', 'line', 'nntp', 'post', 'host', 'ca
rson', 'washington', 'fair', 'number', 'brave', 'soul', 'upgrad', 'clock', 'oscil', 'share', 'ex
peri', 'poll', 'send', 'brief', 'messag', 'detail', 'experi', 'procedur', 'speed', 'attain', 'ra
t', 'speed', 'card', 'adapt', 'heat', 'sink', 'hour', 'usag', 'floppi', 'disk', 'function', 'flo
ppi', 'especi', 'request', 'summar', 'day', 'network', 'knowledg', 'base', 'clock', 'upgrad', 'h
aven', 'answer', 'poll', 'thank', 'guykuo', 'washington']]
```

Slika 6. Polje dobiveno metodom preprocess().

filtering, using the doc2bow() method, the field is converted into a sack of words . Bag of words), and is ready as an input parameter of the LDA method for the development of machine learning models.

```
[['lerxst', 'thing', 'subject', 'nntp', 'post', 'host', 'organ', 'univers', 'maryland', 'colle
g', 'park', 'line', 'wonder', 'enlighten', 'door', 'sport', 'look', 'late', 'earli', 'call', 'br
icklin', 'door', 'small', 'addit', 'bumper', 'separ', 'rest', 'bodi', 'know', 'tellm', 'model',
'engin', 'spec', 'year', 'product', 'histori', 'info', 'funki', 'look', 'mail', 'thank', 'brin
g', 'neighborhood', 'lerxst'], ['guykuo', 'carson', 'washington', 'subject', 'clock', 'poll', 'f
inal', 'summari', 'final', 'clock', 'report', 'keyword', 'acceler', 'clock', 'upgrad', 'articl',
'shelley', 'qvfo', 'innc', 'organ', 'univers', 'washington', 'line', 'nntp', 'post', 'host', 'ca
rson', 'washington', 'fair', 'number', 'brave', 'soul', 'upgrad', 'clock', 'oscil', 'share', 'ex
peri', 'poll', 'send', 'brief', 'messag', 'detail', 'experi', 'procedur', 'speed', 'attain', 'ra
t', 'speed', 'card', 'adapt', 'heat', 'sink', 'hour', 'usag', 'floppi', 'disk', 'function', 'flo
ppi', 'especi', 'request', 'summar', 'day', 'network', 'knowledg', 'base', 'clock', 'upgrad', 'h
aven', 'answer', 'poll', 'thank', 'guykuo', 'washington']]
```

Figure 6. Field obtained by preprocess() method

2.5 LDA Machine Learning Method

Training models using the LDA machine learning method makes it much easier for to use the "Gensim" Python library. Model training requires the gensim.models.LdaMulticore() method. The method receives several input parameters: bow_corpus, Num_topic, Id2word, Passes and Workers. The bow_corpus parameter represents a sack of words that serves as a training data set. The num_topic parameter represents the number of topics that will be categorized by words from a given bag of words. The id2word parameter is used to map each word to its integer. This parameter was used in the later scrolling of categorized topics. The pass parameter is the number of repetitions through a set of data when training a model. It is advisable to put this parameter on a number greater than one, so that the trained model has as accurately categorized words by topic as possible. The workers parameter defines the number of additional processor cores used to train the model, with the aim of training the model as quickly as possible. An example of modelmaking using the LdaMulticore() method:

lda_model=
gensim.models.LdaMulticore(bow_corpus,
num_topics = 8, id2word = dictionary, passes = 10,
workers = 2)

After completing the model training process, it is possible to get an insight into the result of the model by scrolling through categorized words, i.e. scrolling through defined topics.

for idx, topic **in** lda_model. print_topics(-1):
print("Topic: **{}** **n**Words: **{}**". format(idx, topic)) print("**n**")

This loop shows a categorized topic and its words that describe it. Next to each word in a particular topic is a numerical record of the weight of a word that represents the most common word in that category. The more represented a word in a particular categorized topic, when processing a bag of unknown words (words that come as an input parameter for

2.5 LDA metoda strojnog učenja

Modeli obuke pomoću metode strojnog učenja LDA znatno olakšavaju korištenje biblioteke " Gensim " Python. Obuka modela zahtijeva metodu gensim.models.LdaMulticore (). Metoda prima nekoliko ulaznih parametara: bow_corpus, Num_topic, Id2word, Passes i Workers. Parametar bow_corpus predstavlja vreću riječi koja služi kao skup podataka za obuku. Parametar num_topic predstavlja broj tema koje će biti kategorizirane prema riječima iz zadane vrećice riječi. Parametar id2word koristi se za mapiranje svake riječi u njezin cijeli broj. Ovaj je parametar korišten u kasnijem listanju kategoriziranih tema. Parametar prolaza je broj ponavljanja kroz skup podataka prilikom treniranja modela. Preporučljivo je ovaj parametar staviti na broj veći od jedan, kako bi obučeni model imao što točnije kategorizirane riječi po temi. Radni parametar definira broj dodatnih procesorskih jezgri koje se koriste za treniranje modela, s ciljem što bržeg treniranja modela. Primjer izrade modela korištenjem metode LdaMulticore ():

lda_model = gensim.models.LdaMulticore
(bow_corpus, num_topics = 8, id2word = rječnik,
propusnice = 10, radnici = 2)

Nakon završetka procesa obuke modela, moguće je dobiti uvid u rezultat modela skrolanjem kroz kategorizirane riječi, odnosno skrolanjem kroz definirane teme.

za idx, tema **u** lda_model. print_topics (-1):
print("Tema: **{}** \ **n** Riječi : **{}** ". format(idx, tema)) ispis(" **n** ")

Ova petlja prikazuje kategoriziranu temu i riječi koje je opisuju. Uz svaku riječ u pojedinoj temi nalazi se brojčani zapis težine riječi koja predstavlja najčešću riječ u toj kategoriji. Što je neka riječ zastupljenija u određenoj kategoriziranoj temi, prilikom obrade baga nepoznatih riječi (riječi koje dolaze kao ulazni parametar za identifikaciju tema, u našem slučaju WhatsApp poruka), to je veća vjerojatnost prepoznavanja tema na temelju dodijeljenog baga riječi s nekategoriziranim riječima.

Primjer kategoriziranih tema dobivenih strojnim učenjem:

Tema:1
Riječi: 0.018"window" + 0.017*"file" +*
0.011"program" + 0.007*" slika " +*
0.007"version" + 0.006*"graphic" + 0.006*"avail"*
+ 0.006" softwar " + 0,006*"boja" +*
0,005"poslužitelj"*

Slika 7. Primjer kategorizirane teme dobivene strojnim učenjem

identifying topics, in our case WhatsApp messages), the greater the probability of recognizing topics based on an assigned bag of words with uncategorized words.

Example of categorized topics obtained by machine learning:

Topic:1
Words: 0.018"window" + 0.017*"file" + 0.011*"program" + 0.007*"imag" + 0.007*"version" + 0.006*"graphic" + 0.006*"avail" + 0.006*"softwar" + 0.006*"color" + 0.005*"server"*

Figure 7. Example of categorized topic obtained by machine learning

3 Model Testing

The model has been tested on an unknown text record consisting of several related sentences (sentences of the same subject matter). Given that the trained model knows how to recognize the topic of religion, as a test text a textual record of religious themes has been used.

The text record should initially be converted into a bag of words, after which we bring it as an input parameter to the trained model. The result is presented in the form of the achieved representation of the results (engl. score) determining the topic and the words that describe it (engl. topic). An example of testing a created model for identifying topics:

testing_topic = 'One of the key points of Jesus' ministry was how the Kingdom of God is accessible to all people. This underlying theme is especially evident in parables like "The Good Samaritan", "A Lost Sheep", and "A Lost Son". These parables are used to explain how Samaritans, the lost, and sinners all can find their way back to God, and how they will be greeted and embraced by a loving Jesus. These stories illustrate one of Jesus' main teachings, that all people who have faith in Him, will be accepted and embraced by God in heaven.'

Data preprocessing step for the unseen document:
bow_vector = dictionary.
doc2bow(preprocess(testing_topic))
for index, score **in** sorted(lda_model[bow_vector], key=**lambda** tup: -1*tup[1]): print("Score: **{}\t** Topic: **{}**". format(score, lda_model.
print_topic(index, 5)))

Output parameter:
Score: 0.9686757922172546
Topic: 0.013"christian" + 0.008*"jesus" + 0.007*"exist" + 0.005*"bibl" + 0.005*"morality"*

From the above example, we were convinced that the model is capable (with a representation of 97%),

3 Testiranje modela

Model je testiran na nepoznatom tekstualnom zapisu koji se sastoji od više povezanih rečenica (rečenica iste tematike). S obzirom na to da obučeni model zna prepoznati temu vjere, kao ispitni tekst korišten je tekstualni zapis religijske tematike.

Tekstualni zapis treba inicijalno pretvoriti u torbu riječi, nakon čega ga unosimo kao ulazni parametar u obučeni model. Rezultat se prikazuje u obliku postignutog prikaza rezultata (engl. score) određujući temu i riječi koje je opisuju (engl. topic). Primjer testiranja izrađenog modela za identifikaciju tema:

testing_topic = 'Jedna od ključnih točaka Isusove službe bila je kako je Kraljevstvo Božje dostupno svim ljudima. Ova temeljna tema posebno je vidljiva u parabolama kao što su "Dobri Samaritanac", "Izgubljena ovca" i "Izgubljeni sin". Ove se prispodobe koriste da objasne kako Samarićani, izgubljeni i grešnici mogu pronaći svoj put natrag do Boga i kako će ih pozdraviti i zagrliti Isus pun ljubavi. Ove priče ilustriraju jedno od Isusovih glavnih učenja, da će svi ljudi koji imaju vjere u Njega, biti prihvaćeni i zagrljeni od Boga na nebu.'

predobrade podataka za nevidljivi dokument:
luk_vektor = rječnik.
doc2bow(pretproces(test_topic))
za indeks, rezultat **u** sorted(lda_model [bow_vector], key= **lambda** tup: -1*tup[1]): print("Rezultat: **{}\t** Tema: **{}** ". format(rezultat, lda_model.
print_topic (indeks, 5)))

Izlazni parametar:
Ocjena: 0,9686757922172546
Tema: 0,013" kršćanin " + 0,008*" isus " + 0,007*"postojati" + 0,005*" bibl " + 0,005*"moral"*

Iz gornjeg primjera smo se uvjerili da je model sposoban (s zastupljenošću od 97%), na temelju vreće riječi kao ulaznog parametra, prepoznati predmetnu temu.

3.1 Testiranje modela na stvarnim primjerima zapisa chata WhatsApp

WhatsAppNLP biblioteka ima metodu koja izvršava cijeli, prethodno spomenuti i opisani proces obrade, prevođenja i obrade prevedenih podataka, modela obuke i ispisa prepoznatih tema na temelju proslijeđenih podataka iz WhatsApp chat dnevnika.

Metoda analyize_chat_topics () radi upravo to, na temelju proslijeđene eksportirane datoteke sa zapisom svih poruka u tekstualnom formatu i brojem željenih tema na temelju kojih će se kreirati LDA model, prepoznaje teme prisutne u proslijeđenoj Dnevnik chata WhatsApp. Demonstracija na dva stvarna primjera WhatsApp dnevnika razgovora:

based on the sack of words as the input parameter, to recognize the topic in question.

3.1 Testing Models on Real Examples of WhatsApp Chat Logs

WhatsAppNLP library has a method that executes the entire, previously mentioned and described process of processing, translating and processing translated data, training models, and printing recognized topics based on forwarded data from the WhatsApp chat log.

The method analyize_chat_topics() does just that, based on the forwarded exported file with a record of all messages in text format and the number of desired topics on the basis of which the LDA model will be created, it recognizes the topics present in the forwarded WhatsApp chat log. A demonstration on two real examples of WhatsApp chat logs:

```
from data_preprocessing import *
file = 'chat_with_person1.txt'
analyize_chat_topics(file, 3)
```

The result of the function over the first chat log:

```
Score: 0.6965230107307434
Topic: 0.006*"drive" + 0.005*"game" +
0.004*"team" + 0.004*"card" + 0.004*"play"
```

From the first obtained result it is possible to conclude that the top 3 topics of conversation from the attached chat log, topics related to "team games", "religion" and "technology" with a representation of 70%, 16% and 14%.

```
from data_preprocessing import *
file = 'chat_with_person2.txt'
analyize_chat_topics(file, 3)
```

The result of the function over the second chat log:

```
Score: 0.4966925084590912
Topic: 0.012*"armenian" + 0.007*"english" +
0.007*"turkish" + 0.004*"live" + 0.004*"greek"
```

As a result of the analysis over the second chat log, it can be concluded that these are more diverse words and that the topic recognition algorithm predicts the result with less representation (more less representation implies more topics, not less accuracy). From the attached we can conclude that the top 3 topics are related to "nationalities", "universe" and "team games". As is evident from the attached, the model is not able to title categorized topics. The logical reasoning process remains with the user of the WhatsAppNLP library.

```
from data_preprocessing import *
file = 'chat_with_person1.txt'
analyize_chat_topics (datoteka, 3)
```

Rezultat funkcije u prvom dnevniku razgovora:

```
Rezultat: 0,6965230107307434
Tema: 0,006*"drive" + 0,005*"game" +
0,004*"team" + 0,004*"card" + 0,004*"play"
```

Iz prvog dobivenog rezultata moguće je zaključiti da su top 3 teme razgovora iz priloženog chat dnevnika, teme vezane uz "timske igre", "religiju" i "tehnologiju" sa zastupljenošću od 70%, 16% i 14%.

```
from data_preprocessing import *
file = 'chat_with_person2.txt'
analyize_chat_topics (datoteka, 3)
```

Rezultat funkcije preko drugog dnevnika razgovora:

```
Rezultat: 0,4966925084590912
Tema: 0,012*" armenski " + 0,007*" engleski " +
0,007*" turski " + 0,004*"uživo" + 0,004*" grčki "
```

Kao rezultat analize drugog chat dnevnika može se zaključiti da se radi o više različitih riječi i da algoritam za prepoznavanje tema predviđa rezultat s manjom zastupljenošću (više manje zastupljenosti implicira više tema, a ne manju točnost). Iz priloženog možemo zaključiti da su top 3 teme vezane uz "nacionalnosti", "svemir" i "timske igre". Kao što je vidljivo iz priloženog, model nije u mogućnosti nasloviti kategorizirane teme. Logički proces zaključivanja ostaje na korisniku biblioteke WhatsAppNLP.

3.2 Statistika odabranih poruka

Osim identificiranja tema, biblioteka WhatsAppNLP sadrži metode matematičkog izračunavanja statistike poruka koje se odnose na dijelove zapisa poruka kao što su datum i vrijeme poslane poruke, najčešće korištene riječi po sudioniku u izvezenom razgovoru, najčešća razdoblja razgovora u dan, tjedan i mjesec, najčešće korišteni emoji, oblaci riječi i slično.

Statistika se prikazuje u grafičkom, numeričkom ili tekstualnom prikazu, na jednostavan i intuitivan način. Sve metode zahtijevaju samo priloženi put do eksportiranih podataka u tekstualnom formatu kao ulazni parametar. Svaka od ovih metoda radi na sličnim principima.

Iz proslijeđenog dnevnika razgovora stvara se "Pandas DataFrame " u kojem se podaci sortiraju, filtriraju i prikazuju u obliku grafikona, matematičkih izračuna i tekstualnih izvješća pomoću "Pandas"

3.2 Selected Message Statistics

In addition to identifying topics, the WhatsAppNLP library contains methods of mathematically calculating message statistics related to parts of message records such as the date and time of message sent, the most commonly used words per participant in the exported conversation, the most common conversation periods in the day, week and month, the most commonly used emojis, word clouds and the like.

Statistics are displayed in graphical, numerical or textual representations, in a simple and intuitive way. All methods require only the attached path to the exported data in text format as an input parameter. Each of these methods works on similar principles.

From the forwarded chat log, a "Pandas DataFrame" is created in which data is sorted, filtered and displayed in the form of graphs, mathematical calculations and text reports using "Pandas" built-in methods. List and demonstration of methods on two analyzed chat logs.

Description and name of the method messages_per_user (file_path) is a method used to analyze the number of messages per participant in an exported conversation. The sum of all sent messages is displayed in the form of a text report.

```
name
Matea    22103
Rok      17896
Name: name, dtype: int64
```

Figure 8. Examples of the messages_per_user method()

From the pictures it can be see the total number of messages sent by participants of two different chat logs. From the first example, it can be concluded that Rok in the above example sends fewer messages in both exported conversations, and Antonio and Matea more. Description and name of the method: media_per_user(file_path) is a method similar to the previous method, which as an output parameter generates the sum of all media records per user in the exported chat log.

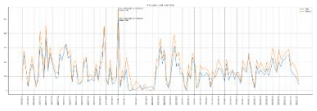


Figure 9. Example of the messages_over_time() method

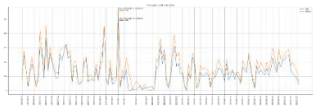
The figure shows the total number of messages sent in a given time period. The first example shows a continuous conversation through almost all dates with minor deviations in the middle of the conversation. The second example shows a graph from which can be concluded that this is an occasional conversation with deviations in regularity and the number of messages sent at the very end. The values on the x-axis represent ugrađenih metoda. Popis i demonstracija metoda na dva analizirana dnevnika razgovora.

Opis i naziv metode messages_per_user (file_path) je metoda koja se koristi za analizu broja poruka po sudioniku u izvezenom razgovoru. Zbroj svih poslanih poruka prikazuje se u obliku tekstualnog izvješća.

```
name
Matea    22103
Rok      17896
Name: name, dtype: int64
```

Slika 8. Primjeri metode messages_per_user ()

Na slikama se može vidjeti ukupan broj poruka koje su poslali sudionici dva različita chat dnevnika. Iz prvog primjera se može zaključiti da Rok u gornjem primjeru šalje manje poruka u oba izvezena razgovora, a Antonio i Matea više. Opis i naziv metode: media_per_user (file_path) je metoda slična prethodnoj metodi, koja kao izlazni parametar generira zbroj svih medijskih zapisa po korisniku u izvezenom chat logu.



Slika 9. Primjer metode messages_over_time ().

Slika prikazuje ukupan broj poruka poslanih u određenom vremenskom razdoblju. Prvi primjer prikazuje kontinuirani razgovor kroz gotovo sve datume s manjim odstupanjima u sredini razgovora. Drugi primjer prikazuje grafikon iz kojeg se može zaključiti da se radi o povremenom razgovoru s odstupanjima u redovitosti i broju poslanih poruka na samom kraju. Vrijednosti na x-osi predstavljaju kontinuirane vrijednosti, dok je konačni izgled vizualizacije posljedica fragmentacije načina komunikacije koja nije u potpunosti u realnom vremenu već ovisi o responzivnosti sudionika u razgovoru. Opis i naziv metode: emojis_per_sender (file_path) je metoda koja analizira prisutne emojije pronađene u izvezenom dnevniku chata i ispisuje broj korištenja određenog emojija s prikazom samog emojija po korisniku koji ga je koristio.

3.3 Ograničenja

Predložena ideja također ima određena ograničenja koja treba priznati. Prvo, biblioteka WhatsAppNLP oslanja se na to da korisnici ručno spremaju svoje razgovore u tekstualnom formatu, što može dovesti do potencijalnih pogrešaka ili propusta. Štoviše, analiza koju provodi knjižnica ovisi o točnosti i mogućnostima korištenih metoda obrade prirodnog jezika, što može utjecati na kvalitetu generiranih uvida. Nadalje, analiza knjižnice ograničena je na

continuous values, while the final appearance of the visualization is due to the fragmentation of the mode of communication that is not completely in real time but depends on the responsiveness of the participants in the conversation. Description and name of the method: emojis_per_sender(file_path) is a method that analyzes the emojis present found in the exported chat log and prints the number of uses of a particular emoji with the display of the emoji itself per user who used it.

3.3 Limitations

The proposed idea also has certain limitations that should be acknowledged. Firstly, the WhatsAppNLP library relies on users manually saving their conversations in text format, which may introduce potential errors or omissions. Moreover, the analysis performed by the library is dependent on the accuracy and capabilities of the natural language processing methods employed, which can impact the quality of insights generated. Additionally, the library's analysis is limited to the content of the messages and does not consider other contextual factors that may influence the interpretation of the data.

Furthermore, the automated analysis provided by the library leaves the task of data interpretation solely to the user, potentially leading to subjective or biased conclusions. Finally, while the approach enables analysis according to security protocols, privacy concerns and ethical considerations should be carefully addressed to ensure compliance with data protection regulations and user consent.

3.4 Business Value Proposition

The business value proposition from this research and the WhatsAppNLP library is to offer businesses and organizations a powerful tool for understanding and extracting meaningful information from short digital messages. It not only enhances work efficiency but also facilitates data-driven decision-making, allowing businesses to make informed choices and stay competitive in their respective industries. Additionally, the library's focus on ethical analysis and privacy protection ensures that it aligns with responsible data usage practices, increasing trust and credibility among users. Library provides a streamlined approach to analyzing short digital messages. By employing natural language processing methods and offering simple method calls, the analysis process is made more accessible to users, even those without extensive expertise in data science or NLP. This ease of use enhances the overall efficiency of analyzing the messages. It takes advantage of the data generated from the WhatsApp application. Since users can freely download and save their conversations in text format, the library can efficiently process and analyze this data, saving time and resources that would otherwise be spent on manual data collection. It also sadržaj poruka i ne uzima u obzir druge kontekstualne čimbenike koji mogu utjecati na tumačenje podataka.

Nadalje, automatizirana analiza koju pruža knjižnica prepušta zadatak tumačenja podataka isključivo korisniku, što potencijalno dovodi do subjektivnih ili pristranih zaključaka. Konačno, dok pristup omogućuje analizu u skladu sa sigurnosnim protokolima, brige o privatnosti i etička pitanja treba pažljivo razmotriti kako bi se osigurala usklađenost s propisima o zaštiti podataka i pristankom korisnika.

3.4 Poslovna vrijednost

Prijedlog poslovne vrijednosti iz ovog istraživanja i WhatsAppNLP biblioteke je ponuditi tvrtkama i organizacijama moćan alat za razumijevanje i izvlačenje smislenih informacija iz kratkih digitalnih poruka. Ne samo da poboljšava radnu učinkovitost, već i olakšava donošenje odluka na temelju podataka, omogućujući poduzećima da donesu informirane odluke i ostanu konkurentna u svojim industrijama. Osim toga, usredotočenost knjižnice na etičku analizu i zaštitu privatnosti osigurava da je u skladu s odgovornim praksama korištenja podataka, povećavajući povjerenje i vjerodostojnost među korisnicima. Knjižnica pruža pojednostavljeni pristup analizi kratkih digitalnih poruka. Korištenjem metoda obrade prirodnog jezika i ponudom jednostavnih metodskih poziva, postupak analize postaje dostupniji korisnicima, čak i onima bez opsežne stručnosti u znanosti o podacima ili NLP-u. Ova jednostavnost korištenja povećava ukupnu učinkovitost analize poruka. Iskorištava podatke generirane iz aplikacije WhatsApp. Budući da korisnici mogu slobodno preuzimati i spremati svoje razgovore u tekstualnom formatu, knjižnica može učinkovito obrađivati i analizirati te podatke, štedeći vrijeme i resurse koji bi se inače potrošili na ručno prikupljanje podataka. Također generira intuitivna izvješća u različitim formatima, kao što su grafikoni, tablice i tekstualne poruke. Ova vizualno privlačna i lako razumljiva izvješća omogućuju korisnicima da brzo shvate ključne uvide iz podataka. Ova je značajka posebno vrijedna za tvrtke i organizacije koje traže korisne informacije iz velike količine poruka. Nadalje, značajno povećava učinkovitost u obradi podataka. Ova automatizacija omogućuje tvrtkama da uštede vrijeme i resurse, omogućujući im da se usredotoče na druge kritične zadatke i strateško donošenje odluka. Konačno, privatnost i etika bitna su razmatranja kada se radi o podacima koje su izradili korisnici. WhatsAppNLP osigurava da se analiza poruka provodi anonimizirano i etički, štiteći privatnost pojedinaca, pružajući vrijedne uvide.

4. Zaključak

Stvaranje modela strojnog učenja zahtijeva zahtjevna znanja iz područja programiranja i podatkovne

generates intuitive reports in various formats, such as graphs, tables, and text messages. These visually appealing and easy-to-understand reports enable users to quickly grasp key insights from the data. This feature is especially valuable for businesses and organizations seeking actionable information from a large volume of messages. Furthermore, it significantly increases efficiency in data processing. This automation allows businesses to save time and resources, enabling them to focus on other critical tasks and strategic decision-making. Finally, privacy and ethics are essential considerations when dealing with user-generated data. WhatsAppNLP ensures that the analysis of messages is conducted in an anonymized and ethical manner, safeguarding the privacy of individuals while providing valuable insights.

4 Conclusion

Creating a machine learning model requires demanding knowledge in the field of programming and data science. The quality of the model depends on the quality of the data we use to learn the machine learning model and nowadays there are countless different sources of quality data. Although one of the most commonly used datasets was chosen, whose application is directly related to the categorization of text documents, and consists of more than 18,000 text documents with a million number of words, there are still shortcomings in the obscurity of categorization of topics on the example of a single conversation with text messages. This is best demonstrated by the complexity of creating a model that is able to recognize any topic that two or more users can talk about.

On the bright side, the WhatsAppNLP library proves that this is very possible and with the expansion of datasets with which the model will be taught to identify a larger number of topics, it gives an optimistic picture for the future of natural language processing development. Such a system can be used in real time, but the real value is in the analysis of communication files that subsequently need to be analyzed in large quantities for the needs of business research. The paper shows that using WhatsAppNLP methods of analyzing WhatsApp messages, it is possible to come to knowledge about patterns and behavior trends. WhatsAppNLP library has plenty of room for improvement and refinement like: multilingual support, real time analysis, integration with other chat applications, customizable analysis parameters and community support. By incorporating these improvements, the WhatsAppNLP library can evolve into a more powerful and versatile tool, increasing its appeal to businesses, researchers, and data analysts seeking to derive valuable insights from short digital messages.

znanosti. Kvaliteta modela ovisi o kvaliteti podataka koje koristimo za učenje modela strojnog učenja, a danas postoji bezbroj različitih izvora kvalitetnih podataka. Iako je odabran jedan od najčešće korištenih skupova podataka čija je primjena izravno povezana s kategorizacijom tekstualnih dokumenata, a sastoji se od više od 18.000 tekstualnih dokumenata s milijunskim brojem riječi, još uvijek postoje nedostaci u nejasnoći kategorizacije tema na primjer jednog razgovora s tekstualnim porukama. To najbolje pokazuje složenost izrade modela koji je u stanju prepoznati bilo koju temu o kojoj mogu razgovarati dva ili više korisnika.

Sa vedrije strane, biblioteka WhatsAppNLP dokazuje da je to vrlo moguće i uz proširenje skupova podataka s kojima će se model naučiti identificirati veći broj tema, daje optimističnu sliku za budućnost razvoja obrade prirodnog jezika. Takav sustav može se koristiti u realnom vremenu, ali prava vrijednost je u analizi komunikacijskih datoteka koje je naknadno potrebno analizirati u velikim količinama za potrebe poslovnih istraživanja. U radu se pokazuje da je pomoću WhatsAppNLP metoda analize WhatsApp poruka moguće doći do saznanja o obrascima i trendovima ponašanja. Knjižnica WhatsAppNLP ima puno prostora za poboljšanje i doradu, poput: višejezične podrške, analize u stvarnom vremenu, integracije s drugim aplikacijama za chat, prilagodljivih parametara analize i podrške zajednice. Uključivanjem ovih poboljšanja, WhatsAppNLP biblioteka može se razviti u moćniji i svestraniji alat, povećavajući svoju privlačnost tvrtkama, istraživačima i analitičarima podataka koji nastoje izvući vrijedne uvide iz kratkih digitalnih poruka.

U ovom radu predstavljena je učinkovita metoda analize sadržaja poruka kako bi se istaknuo potencijal primjene takvih i sličnih tehnika u poslovanju modernih, digitalno orijentiranih organizacija kako je opisano u rubrici prijedloga poslovne vrijednosti.

Kada je riječ o obradi prirodnog jezika, nadogradnja modela strojnog učenja znači bolje prepoznavanje tema kao i prepoznavanje šireg spektra tema.

Nadogradnja u tom smislu zahtijeva prikupljanje i kategorizaciju većeg broja tekstova u zajednički skup podataka koji će se koristiti kao izvor podataka strojnog učenja. Posljednjih godina bilježi se porast trenda implementacije sadržaja društvenih mreža, konkretno digitalnih platformi za komunikaciju i razmjenu poriva, na niz novih načina za stvaranje poslovne vrijednosti. Društveni mediji, naposljetku, sadrže mnoštvo informacija i dezinformacija o pojedinačnim korisnicima i njihovim mrežama, a više zakona ograničava što tijela za provođenje zakona mogu učiniti s podacima društvenih medija.

In this paper, an efficient method of analyzing the content of messages has been presented in order to highlight the potential of applying such and similar techniques in the business of modern, digitally oriented organizations as described in the business value proposition section.

When it comes to natural language processing, upgrading machine learning models means better recognition of topics as well as recognizing a wider range of topics.

Upgrading in this sense requires the collection and categorization of a larger number of texts into a common dataset to be used as a source of machine learning data. In recent years, there has been an increase in the trend of implementing social media content, in the specific case of digital platforms for communication and exchange of urges, in a number of new ways to create business value. Social media, after all, contains a wealth of information and misinformation about individual users and their networks, and more laws limit what law enforcement can do with social media data.

References

Mukhopadhyay, S. (2018). Advanced Data Analytics Using Python. Apress, ISBN-13 (pbk): 978-1-4842-3449-5 ISBN-13 (electronic): 978-1-4842-3450-1

Marasović, A. (2021) Latent semantic analysis, variants and applications. *Faculty of Science*, Retrieved from https://repozitorij.pmf.unizg.hr/islandora/object/pmf%3A5264/datastream/PDF

Dinesh K., Sunil A., Sorab K., & Mohanty N.M. (2021). Sentiment Study Approach Based on Chat Summarization, Retrieved from: https://www.researchgate.net/publication/334096166_Sentiment_Study_Approach_Based_on_Chat_Summarization

Sweigart A. (2015). Automate the Boring Stuff with Python, 2nd Edition: Practical Programming for Total Beginners, No Starch Press, ISBN-13: 978-1593279929,

Walkowska, J. (2010). *An NLP-Oriented Analysis of the Instant Messaging Discours.* In Sojka, P., Horák, A., Kopeček, I., & Pala, K. (Eds) (2010) *Text, Speech and Dialogue. TSD 2010. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, vol 6231. https://doi.org/10.1007/978-3-642-15760-8_73

Vetulani, Z., Marciniak, J., Konieczka, P., & Walkowska, J. (2008). An SMS-Based System Architecture (Logical Model) To Support Management of Information Exchange in Emergency Situations POLINT-112-SMS project.

Reference

Mukhopadhyay, S. (2018). Advanced Data Analytics Using Python. *Apress*, ISBN-13 (pbk): 978-1-4842-3449-5 ISBN-13 (electronic): 978-1-4842-3450-1

Marasović, A. (2021) Latent semantic analysis, variants and applications. *Faculty of Science*, Retrieved from https://repozitorij.pmf.unizg.hr/islandora/object/pmf%3A5264/datastream/PDF

Dinesh K., Sunil A., Sorab K., & Mohanty N.M. (2021). Sentiment Study Approach Based on Chat Summarization, Retrieved from: https://www.researchgate.net/publication/334096166_Sentiment_Study_Approach_Based_on_Chat_Summarization

Sweigart A. (2015). Automate the Boring Stuff with Python, 2nd Edition: Practical Programming for Total Beginners, No Starch Press, ISBN-13: 978-1593279929,

Walkowska, J. (2010). *An NLP-Oriented Analysis of the Instant Messaging Discours.* In Sojka, P., Horák, A., Kopeček, I., & Pala, K. (Eds) (2010) *Text, Speech and Dialogue. TSD 2010. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, vol 6231. https://doi.org/10.1007/978-3-642-15760-8_73

Vetulani, Z., Marciniak, J., Konieczka, P., & Walkowska, J. (2008). An SMS-Based System Architecture (Logical Model) To Support Management of Information Exchange in Emergency Situations POLINT-112-SMS project. In: *Intelligent Information Processing IV - 5th IFIP International Conference on Intelligent Information Processing* (pp. 240–253). Springer, Boston

Khurana, D., Koli, A., Khatter, K. et al. (2022). Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. https://doi.org/10.1007/s11042-022-13428-4

Wang Z., Ng P., Ma X., Nallapati R., Xiang B. (2019). Multi-passage bert: A globally normalized bert model for open-domain question answering, arXiv preprint arXiv:1908.08167

Vetulani, Z., Marciniak, J. (2000). Corpus Based Methodology in the Study and Design of Systems with Emulated Linguistic Competence, In: Christodoulakis, D.N. (Ed.) *NLP 2000*, LNCS (LNAI), vol. 1835. (pp. 346–357) Springer, Heidelberg

Sworna Z.T., Mousavi Z, Babar M.A. (2022). *NLP methods in host-based intrusion detection*

In: *5th IFIP International Conference* (pp. 240–253). Springer, Boston

Khurana, D., Koli, A., Khatter, K. et al. (2022). Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. https://doi.org/10.1007/s11042-022-13428-4

Wang Z., Ng P., Ma X., Nallapati R., Xiang B. (2019). Multi-passage bert: A globally normalized bert model for open-domain question answering, arXiv preprint arXiv:1908.08167

Vetulani, Z., Marciniak, J. (2000). Corpus Based Methodology in the Study and Design of Systems with Emulated Linguistic Competence, In: Christodoulakis, D.N. (Ed.) *NLP 2000*, LNCS (LNAI), vol. 1835. (pp. 346–357) Springer, Heidelberg

Sworna Z.T., Mousavi Z, Babar M.A. (2022). *NLP methods in host-based intrusion detection Systems: A systematic review and future directions*, arXiv preprint arXiv:2201.08066

Walkowska, J. (2009). Gathering and Analysis of a Corpus of Polish SMS Dialogues. In: *Challenging Problems of Science. Computer Science/Recent Advances in Intelligent Information Systems*, (pp. 145–157). Academic Publishing House EXIT, Warsaw

Hult, C.A., Richins, R. (2006). The Rhetoric and Discourse of Instant Messaging, In: *Computers and Composition Online*, Retrieved from: http://www.bgsu.edu/cconline/hultrichins_im/hultrichins_im.htm

Hurlow, C., Brown, A. (2003). Generation Txt? The Sociolinguistics of Young People's Text-Messaging, In: *Discourse Analysis Online*, *Department of Communication*, University of Washington, Box 353740, Seattle, WA 98195, USA, Retrieved from: http://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003-01.html

Grossz, B.J., Sidner, C.L. (1986). Attention, Intentions and the Structure of Discourse, *Computational Linguistics* 12(3), 175–204

Pribisalić, E. (2019). Obrada prirodnog jezika, *Sveučilište u Osijeku*, Retrieved from: http://www.mathos.unios.hr/~mdjumic/uploads/diplomski/PRI14.pdf

Wonderflow (2018), 2 NLP Examples: How Natural Language Processing is Used, https://www.wonderflow.ai/blog/natural-language-processing-examples

Jagota, A. (2020) Topic Modeling In NLP with a focus on Latent Dirichlet Allocation, https://towardsdatascience.com/topic-modeling-in-nlp-524b4cffeb68

Chen, E, (2011) Introduction to Latent Dirichlet Allocation, http://blog.echen.me/2011/08/22/introduction-to-latent-dirichlet-allocation/

Kelemen, Zádor Dániel. (2015). Fundamental Analysis of a Developer Support Chat Log for Identifying Process Improvement Opportunities. 10.13140/RG.2.1.1465.0407

Rosen, Devan & Miagkikh, Victor & Suthers, Dan. (2011). Social and semantic network analysis of chat logs. ACM IC Proceeding Series. 134-139. 10.1145/2090116.2090137.

Kumar T. Sandeep. (2020). 8 great Python libraries for natural language processing, Pulse, Retreived from: https://www.linkedin.com/pulse/8-great-python-libraries-natural-language-processing-t-sandeep-kumar/

https://towardsdatascience.com/topic-modeling-in-nlp-524b4cffeb68

Chen, E, (2011) Introduction to Latent Dirichlet Allocation, http://blog.echen.me/2011/08/22/introduction-to-latent-dirichlet-allocation/

Kelemen, Zádor Dániel. (2015). Fundamental Analysis of a Developer Support Chat Log for Identifying Process Improvement Opportunities. 10.13140/RG.2.1.1465.0407

Rosen, Devan & Miagkikh, Victor & Suthers, Dan. (2011). Social and semantic network analysis of chat logs. ACM IC Proceeding Series. 134-139. 10.1145/2090116.2090137.

Kumar T. Sandeep. (2020). 8 great Python libraries for natural language processing, Pulse, Retreived from: https://www.linkedin.com/pulse/8-great-python-libraries-natural-language-processing-t-sandeep-kumar/

Student-Centered Digital Services: Assessing User Experience in Higher Education

Digitalne usluge usmjerene na studenta: vrednovanje korisničkog iskustva u visokom obrazovanju

Tea Mijač
University of Split
Sveučilište u Splitu
Faculty of Economics, Business and Tourism
Ekonomski fakultet
Cvite Fiskovića 5, Split, Hrvatska
tea.mijac@efsthr

Abstract. *In today's digital age, higher education institutions (HEIs) are facing increasing demands to provide a positive user experience (UX) for students with digital services. Institutions seeking to attract and retain students must recognize the importance of students' familiarity with technology in their everyday lives. Delivering digital services that meet their needs and expectations has become essential. Defining and measuring the UX of digital services is a complex task that involves various criteria, including pragmatic and hedonic aspects. The quantitative survey was conducted with students, and two dimensional UX model was confirmed. By understanding the different aspects of UX, HEIs can enhance their digital services to meet students' expectations.*

Keywords. user experience, higher education, digital transformation, student-oriented approach.

Sažetak. *U današnjem digitalnom dobu, institucije visokog obrazovanja suočavaju se sa sve većim zahtjevima za pružanjem pozitivnog korisničkog iskustva (UX) za studente s digitalnim uslugama. Institucije koje žele privući i zadržati studente moraju prepoznati studentsku naviknutost na tehnologije u svakodnevnom životu. Pružanje digitalnih usluga koje ispunjavaju njihove potrebe i očekivanja postalo je ključno. Definiranje i mjerenje UX-a digitalnih usluga složen je zadatak koji uključuje različite kriterije, uključujući pragmatične i hedonističke aspekte. Provedeno je kvantitativno istraživanje sa studentima, te je potvrđen dvodimenzionalni UX model. Razumijevanjem različitih aspekata UX-a, visoka učilišta mogu unaprijediti svoje digitalne usluge kako bi ispunila očekivanja studenata.*

Ključne riječi. Korisničko iskustvo, visoko obrazovanje, digitalna transformacija, studentu orijentiran pristup.

1 Introduction

Digitalization of Higher Education (HE) has been rapidly grooving, and most services are switching from offline to online settings (Pasini et al., 2019). It is no surprise that the digital transformation of HE has been further accelerated by the Covid-19 crisis (Bygstad et al., 2022). In addition, the phenomenon of the internationalization of HE has resulted in intense global competition among higher education institutions (HEIs) for prospective students.

The digital transformation of HEIs involves meeting the expectations and needs of students

1 Uvod

Digitalizacija visokog obrazovanja (VO) brzo napreduje, a većina usluga prelazi s *offline* na *online* okruženje (Pasini et al., 2019). Nije iznenađenje da je digitalna transformacija VO dodatno ubrzana krizom uzrokovanom Covid-19 (Bygstad et al., 2022). Osim toga, fenomen internacionalizacije VOa rezultirao je intenzivnom globalnom konkurencijom među institucijama visokog obrazovanja (IVO) za buduće studente.

(Rodrigues, 2017), and the current generation of students, identified as Generation Z, is characterized by their familiarity and comfort with technology (Martins et al., 2019). For (future) students, digitization in HE should ensure a richer UX and provide more accurate and faster answers to inquiries that are mostly made over the phone or website (Khalid et al., 2018).

These students are accustomed to using different devices and digital services to stay connected to the world, their friends, and their work (Martins et al., 2019). Addressing these needs is critical for HEIs undergoing digital transformation to engage and retain students effectively. The new generation of students has increasingly high expectations regarding the quality of their interaction with the processes of HEIs and the level of student UX. Precisely, students expect a personalized UX, regardless of what type of digital service it is (Obaid, 2019).

For example, some of the factors that contribute to HEIs quality are (Cadena et al., 2018): (1) administrative transparency, (2) student satisfaction, (3) digitization of administrative services, and (4) share of digitized operational processes. Furthermore, providing digital services that deliver a high-quality UX can have several benefits, including increased student satisfaction. It has been confirmed that a higher level of UX leads to higher perceived benefits provided by digital services in HE (Mijač et al., 2023). Consequently, it is imperative to systematically evaluate the concept of UX in delivering digital services in HE (Al-Hunaiyyan et al., 2021).

Users' habits and interests are changing, new trends are emerging, and it should also be borne in mind that UX is highly subjective, dynamic, and contextually dependent (Halvorsrud et al., 2016). UX is crucial for the student population, especially if they are treated as consumers of higher education (Ng & Forbes, 2009).

The importance of providing high-quality UX has been highlighted in the scholarly literature (Díaz et al., 2016; Rodrigues, 2017), particularly in light of its potential to attract and retain students, which is critical in the current competitive landscape.

There have been studies examining the UX in the HE context, such as in the example of the library (Wiles et al., 2013), e-learning (Calle-Jimenez et al., 2021; Slade & Downer, 2020; van Staden et al., 2019), smart glasses (Santana et al., 2022), and Student Information Systems (Al-Hunaiyyan et al., 2021; Langgawan Putra et al., 2021). However, even though most authors consider more aspects when measuring the UX, generally speaking, there is no academic nor practical uniformly accepted model for measuring it (Hornbæk & Hertzum, 2017). This paper aims to verify the dimensions of UX and propose a multidimensional model for UX in an HE setting.

Digitalna transformacija IVO uključuje zadovoljavanje očekivanja te potreba studenata (Rodrigues, 2017), a trenutnu generaciju studenata, identificiranu kao *Generacija Z*, karakterizira njihova upoznatost i udobnost s tehnologijom (Martins et al., 2019). Za (buduće) studente, digitalizacija u VOu trebala bi osigurati bogatije korisničko iskustvo (engl. User eXperience, UX) i pružiti točnije i brže odgovore na upite koji se još uvijek najčešće postavljaju putem telefona ili web stranica (Khalid et al., 2018).

Studenti nove generacije naviknuti su na korištenje različitih uređaja i digitalne usluge, a to rade kako bi bili povezani 24/7 (Martins et al., 2019). Adresiranje njihovih potreba ključno je za institucije koji prolaze kroz digitalnu transformaciju kako bi učinkovito uključili i zadržali studente. Nova generacija studenata ima sve veća očekivanja u vezi s kvalitetom interakcije s procesima institucije, ali i veća očekivanja razine UXa. Točnije, studenti očekuju personalizirano UX iskustvo, bez obzira na vrstu digitalne usluge (Obaid, 2019).

Na primjer, neki od čimbenika koji doprinose kvaliteti institucija VO su (Cadena et al., 2018): (1) transparentnost administrativnih procesa, (2) zadovoljstvo studenata, (3) digitalizacija administrativnih usluga i (4) udio digitaliziranih operativnih procesa. Nadalje, pružanje digitalnih usluga koje pružaju visokokvalitetno UX može osigurati dodatne koristi, kao što je veća razina zadovoljstva studenata. Potvrđeno je da viša razina UXa dovodi do većih percipiranih koristi koje pružaju digitalne usluge u VO (Mijač et al., 2023). Sve navedeno implicira na važnost sustavnog vrednovanja koncepta UXa pri pružanju digitalnih usluga u VO (Al-Hunaiyyan et al., 2021).

Kako se pojavljuju novi trendovi, tako se mijenjaju navike, ali i interesi krajnjih korisnika. Treba imati na umu da je UX jako subjektivan, dinamičan i ovisan o kontekstu (Halvorsrud et al., 2016). UX je ključan za studentsku populaciju, posebno uzimajući u obzir činjenicu da se studenti često smatraju potrošačima visokog obrazovanja (Ng & Forbes, 2009).

Važnost pružanja visokokvalitetnog UXa istaknuta je u znanstvenoj literaturi (Díaz et al., 2016; Rodrigues, 2017), a posebno s obzirom na njegov potencijal da privuče i zadrži studente, što je ključno u trenutnom konkurentnom okruženju.

Provedene su studije koje istražuju UX u kontekstu VO, poput primjera u knjižnici (Wiles et al., 2013), e-učenju (Calle-Jimenez et al., 2021; Slade & Downer, 2020; van Staden et al., 2019), pametnim naočalama (Santana et al., 2022) te informacijskim sustavima studenata (engl. Student Infromation System) (Al-Hunaiyyan et al., 2021; Langgawan Putra et al., 2021). Međutim, iako većina autora uzima u obzir više aspekata pri mjerenju UXa, općenito govoreći, ne postoji jednoglasno prihvaćen model za njegovo mjerenje ni u akademiji ni u industriji (Hornbæk & Hertzum, 2017).

Based on those mentioned above, posed research question is: *How to measure UX construct in HE context?*

The paper is organized as follows: Section 2 discusses the importance of measuring the UX in the context of higher education and digital transformation. Section 3 outlines the research methodology, while Section 4 presents the empirical results of the study. Section 5 highlights the main contributions and findings of the study, along with potential areas for future research. Conclusions are presented in the last section.

2 Literature review

As mentioned in the introductory part, digital transformation in HE is happening, resulting in numerous digital services (Pasini et al., 2019). The role of technology in higher education is aimed at increasing the efficiency and effectiveness of organizational processes and meeting the needs of all stakeholders - students, parents, parents, parents, parents, government, and society (Caushi & Dika, 2018). Digital transformation encompasses all areas of work of HEIs; that is, it equally permeates the areas of research, teaching, and teaching, as well as their administrative processes (Gilch et al., 2019).

It is essential that digital services and information systems offer a positive UX and meet the needs of students simultaneously (Al-Hunaiyyan et al., 2021). All digital services and systems should have a good UX in order to minimize user confusion and get good feedback from the user (Langgawan Putra et al., 2021).

The concept of UX has garnered greater attention in both research and practice due to a recent shift in the domain towards an experiential approach that prioritizes user enjoyment and satisfaction over functionality (Sonderegger et al., 2019). According to the ISO 9241-210:2010 standard, UX encompasses the user's perceptions and response to the use or participation in the use of a particular service, product, or system.

UX is a holistic concept that includes different emotional, cognitive, or physical reactions to use, even to the very thought of using services, and is formed before, during, or after use. UX can be considered as a consequence of presentation, functionality, system performance, interactive behavior, attitudes, skills and personality, and the context of use (Hinderks et al., 2019), and at the same time, represents one of the essential aspects of applying a user-oriented approach (Richter & Fluckiger, 2014).

Researchers have made several proposals for UX models aimed at measuring the UX of different technology products. Author (Hassenzahl Marc, 2003) proposed a model which highlights the pragmatic and hedonic characteristics of a

Cilj ovog rada je istražiti dimenzije UXa te predložiti višedimenzionalni model za mjerenje UXa u kontekstu visokog obrazovanja.

Na temelju navedenog, postavljeno istraživačko pitanje glasi: *Kako mjeriti konstrukt UX-a u kontekstu VOa?*

Ovaj rad je organiziran kako slijedi: U odjeljku 2 raspravlja se o važnosti mjerenja UX-a u kontekstu visokog obrazovanja i digitalne transformacije. Odjeljak 3 prikazuje istraživačku metodologiju, dok odjeljak 4 predstavlja empirijske rezultate istraživanja. Odjeljak 5 ističe glavne doprinose i nalaze istraživanja, zajedno s potencijalnim područjima za buduća istraživanja. Zaključak rada prezentiran je u zadnjem odjeljku.

2 Pregled literature

Kao što je spomenuto u uvodnom dijelu, digitalna transformacija u visokom obrazovanju je pokret koji jako relevantan i aktualan, a rezultira brojnim digitalnim uslugama (Pasini et al., 2019). Uloga tehnologije u visokom obrazovanju ima za cilj povećanje učinkovitosti i učinkovitosti organizacijskih procesa te zadovoljavanje potreba različitih dionika (Caushi & Dika, 2018). Digitalna transformacija obuhvaća sva područja rada visokoškolskih ustanova; odnosno, jednako prožima područje istraživanja, nastave i administrativnih procesa (Gilch et al., 2019).

Važno je da digitalne usluge i informacijski sustavi pružaju pozitivno korisničko iskustvo i istovremeno zadovoljavaju potrebe studenata (Al-Hunaiyyan et al., 2021). Visoka razina UXa je bitna za sve digitalne usluge i sustave kako bi studenti imali manje problema pri korištenju, a i kako bi se ostvarila pozitivna povratna informacija (Langgawan Putra et al., 2021).

Koncept UXa privukao je veliku pažnju, kako u istraživanju, tako i u praksi zbog nedavnog pomaka fokusa prema iskustvenom pristupu koji stavlja naglasak na korisnikovo zadovoljstvo i uživanje više nego na funkcionalnost (Sonderegger et al., 2019). Prema ISO 9241-210:2010 standardu, UX obuhvaća korisnikove percepcije i odgovor na korištenje određene usluge, proizvoda ili sustava.

UX se smatra holističkim konceptom koji uključuje različite emocionalne, kognitivne ili fizičke reakcije na upotrebu, čak i na samu pomisao o korištenju usluga, a formira se prije, za vrijeme ili nakon upotrebe. UX se može smatrati posljedicom prezentacije, funkcionalnosti, performansi sustava, interaktivnog ponašanja, stavova, vještina i osobnosti te konteksta upotrebe (Hinderks et al., 2019). UX istovremeno predstavlja jedan od ključnih aspekata primjene korisniku orijentiranog pristupa (Richter & Fluckiger, 2014).

Autori su iznijeli nekoliko prijedloga za modele s ciljem mjerenja UXa različitih tehnoloških proizvoda. Primjerice, autor (Hassenzahl Marc, 2003) je

technological product that serve as the foundation for determining the resulting consequences in a given situation, including evaluations of appeal, pleasure, and satisfaction. There so, UX can be defined as a set of quality criteria encompassing various dimensions, including classic usability criteria such as efficiency and ease of use, as well as hedonic criteria such as stimulation, fun, novelty, emotions, and aesthetics (Preece et al., 2019; Wani et al., 2017). This approach has the advantage of breaking down the complex concept of UX into a series of simple quality criteria that describe different and relatively well-defined aspects of the UX, which can be measured independently (Hinderks et al., 2019).

Authors (Thüring & Mahlke, 2007) also argue for a breather perspective that regards UX. They proposed components of the user experience (CUE) model as a compound of three elements: Perception of quality and effectiveness of the system, Visual aesthetics, and Emotional response. In the HE setting, authors used three similar dimensions to measure students' UX: ease of use, appeal, and pleasure (Zardari et al., 2021).

Authors (Langgawan Putra et al., 2021) measured UX with UX Honeycomb consisting of 7 aspects: useful, usable, desirable, valuable, findable, accessible, and credible. Another research on UX with Student Information Systems used six dimensions to measure UX (Al-Hunaiyyan et al., 2021): attractiveness, efficiency, perspicuity (ease of use), dependability, stimulation, and novelty.

UX can be viewed as an expansion of the usability construct that takes a more comprehensive approach, considering not only the cognitive aspects but also the user's emotions and overall experience when interacting with an interface (Sonderegger et al., 2019).

Authors (Hornbæk & Hertzum, 2017) conducted a review paper and concluded that all UX models have four things in common:

- The majority of models distinguish between pragmatic and hedonic characteristics.
- Tthe perception of aesthetics and overall assessments of beauty are crucial in UX models.
- UX models demonstrate how perceptions of products and services evolve over time.
- Emotions play a pivotal role in UX models.

Pragmatic aspects are also known as instrumental aspects, while hedonic is known as non-instrumental aspects (Hassenzahl, 2005). The author also suggests that while individuals may have instrumental goals and functional requirements that a product can meet, it's also crucial to consider the presence of additional non-instrumental, hedonic needs.

Research so far revealed that it is not uncommon for students to have negative UX when using digital services in HE (Saplacan et al., 2018; Thoring et al., 2017). Therefore, it is particularly important to

predložio model koji ističe pragmatične i hedoničke karakteristike tehnološkog proizvoda koje služe kao temelj za određivanje posljedica u određenoj situaciji, uključujući procjene privlačnosti, užitka i zadovoljstva. UX se može definirati kao skup kriterija kvalitete koji obuhvaćaju različite dimenzije, uključujući klasične kriterije upotrebljivosti poput učinkovitosti i jednostavnosti korištenja, kao i hedoničke kriterije poput stimulacije, zabave, novosti, emocija i estetike (Preece et al., 2019; Wani et al., 2017). Ovaj pristup ima prednost razbijanja kompleksnog koncepta UXa na niz jednostavnih kriterija kvalitete koji opisuju različite i relativno dobro definirane aspekte UXa, koji se mogu neovisno mjeriti (Hinderks et al., 2019).

Autori (Thüring & Mahlke, 2007) također zagovaraju širi pogled na UX. Tako su predložili komponente modela korisničkog iskustva (CUE) kao kombinaciju tri elementa: percepciju kvalitete i učinkovitosti sustava, vizualnu estetiku i emocionalnu reakciju. U kontekstu visokog obrazovanja, autori su koristili tri slične dimenzije za mjerenje UXa studenata: jednostavnost korištenja, privlačnost i užitak (Zardari et al., 2021).

Dodatno, autori (Langgawan Putra et al., 2021) su mjerili UX s pomoću UX Honeycomb modela koji se sastoji od 7 aspekata: korisno, upotrebljivo, poželjno, vrijedno, lako za pronaći, dostupno i vjerodostojno. Istraživanja koja su se bavila mjerenjem UXa studentskih informacijskih sustava, koristila su šest dimenzija za mjerenje UXa (Al-Hunaiyyan et al., 2021): privlačnost, učinkovitost, jasnoća (jednostavnost korištenja), pouzdanost, stimulacija i novost.

UX se može promatrati kao proširenje konstrukta upotrebljivosti koje ima sveobuhvatan pristup, uzimajući u obzir ne samo kognitivne aspekte već i korisnikove emocije i cjelokupno iskustvo prilikom interakcije s sučeljem (Sonderegger et al., 2019).

Autori (Hornbæk & Hertzum, 2017) u svom preglednom radu su zaključili kako svi modeli UXa imaju četiri zajednička elementa:

- Većina modela razlikuje pragmatične i hedoničke karakteristike.
- Percepcija estetike i opća procjena ljepote ključni su u modelima UX-a.
- Modeli UXa pokazuju kako se percepcije proizvoda i usluga mijenjaju tijekom vremena.
- Emocije imaju ključnu ulogu u modelima UXa.

Pragmatični aspekti također su poznati kao instrumentalni aspekti, dok se hedonički aspekti nazivaju neinstrumentalni aspektima (Hassenzahl, 2005). Autor također sugerira da, iako pojedinci mogu imati instrumentalne ciljeve i funkcionalne zahtjeve koje proizvod može ispuniti, važno je također uzeti u obzir prisutnost dodatnih neinstrumentalnih, hedoničkih potreba.

Dosadašnja istraživanja otkrila su da nije neuobičajeno da studenti imaju negativno korisničko

examine the dimensions of UX depending on the context since context factors may influence the entire UX entirety (Sonderegger et al., 2019).

Despite the growing importance of UX in attracting and retaining students, very few papers cover these phenomena. To illustrate, a search conducted in the Scopus database on March 2023 with the keywords "user experience" or UX and "higher education" or university resulted in 57 documents. Results are illustrated in Figure. 1.

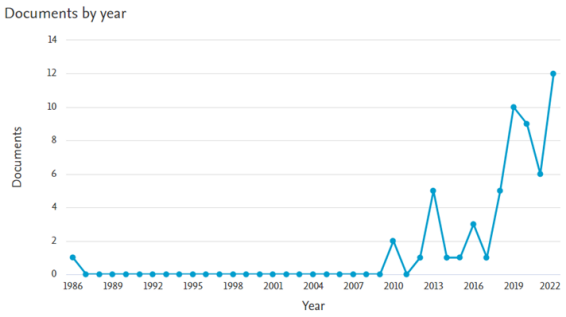


Figure 1. Literature evolution on UX in HE

Due to the fact that recent works in HE still actively investigate which dimensions make up UX and for that purpose, this piece of research tends to address identified research gap by proposing a holistic view defining UX as a multidimensional construct.

3 Methodology

A qualitative survey with student as and-users of digital services in HE was conducted to answer the posed research question. Digital services examined were not related to learning but supportive /administrative digital services. Participants were students from ten institutions of higher education in Croatia. The total number of fully-completed survey questionnaires was n=551.

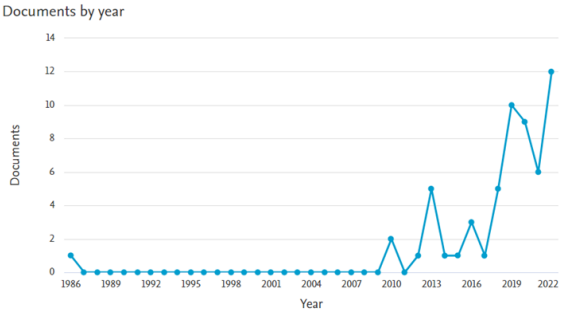
To examine and validate the most commonly used indicators for measuring UX construct, three subdimensions were taken into consideration: *usability*, *appeal*, and *satisfaction*. Items were adapted from previously conducted research.

The prepared survey was composed of 14 items related to measuring UX dimensions. The questionnaire items were rated on a five-point Likert scale, with responses ranging from 1 (strongly agree) to 5 (strongly disagree).

Once the data was gathered, statistical techniques were employed (Hair et al., 2010), specifically confirmatory factor analysis (CFA), to investigate the dimensionality of the UX construct. The data analysis was carried out using IBM SPSS Statistics and IBM SPSS Amos for structural equation modeling (SEM). Since the theory indicates the multidimensionality of UX, the proposed model was

iskustvo prilikom korištenja digitalnih usluga u visokom obrazovanju (Saplacan et al., 2018; Thoring et al., 2017). Stoga je posebno važno ispitati dimenzije UXa ovisno o kontekstu, budući da faktori konteksta mogu utjecati na cjelokupno korisničko iskustvo (Sonderegger et al., 2019).

Unatoč rastućoj važnosti UX-a u privlačenju i zadržavanju studenata, vrlo malo znanstvenih radova se bavim ovim fenomenima. Za ilustraciju, pretraživanje provedeno u bazi Scopus u ožujku 2023. s ključnim riječima "user experience" ili UX i "higher education" ili university rezultiralo je s 57 dokumenata. Rezultati pretrage vizualno su prikazani na slici 1.



Slika 1. Pregled literature o UX u VO

S obzirom da recentni radovi iz područja VO još uvijek aktivno istražuju koje dimenzije sačinjavaju konstrukt UX, ovo istraživanje teži rješavanju identificiranih praznina u istraživanju predlažući holistički pogled te definirajući UX kao višedimenzionalni konstrukt.

3 Metodologija

Da bi se pružio odgovor na postavljeno istraživačko pitanje, provedeno je kvalitativno istraživanje sa studentima kao korisnicima digitalnih usluga u visokom obrazovanju. Digitalne usluge koje su bile predmet istraživanja nisu bile direktno povezane s učenjem, već se radilo o podršci/administrativnim digitalnim uslugama. Sudionici istraživanja bili su studenti na deset institucija visokog obrazovanja u Hrvatskoj. Ukupan broj potpuno popunjenih upitnika bio je 551.

U svrhu istraživanja i potvrđivanja najčešće korištenih indikatora (dimenzija) za mjerenje konstrukta korisničkog iskustva, uzete su u obzir tri dimenzije: upotrebljivost (engl. *usability*), privlačnost (engl. *appeal*) i zadovoljstvo (engl. *satisfaction*). Čestice za svaki od indikatora su razvijene u skladu s prethodno provedenim istraživanjima te prilagođene kontekstu.

Pripremljeni anketni upitnik sastojao se od ukupno 14 čestica koje su mjerile dimenzije korisničkog iskustva. Ispitanici su svaku česticu upitnika ocijenili uz pomoć peto stupanjske Likertove ljestvice, pri

examined as a higher-order construct (HOC) (Awang et al., 2017; Hair et al., 2010).

4 Results

Descriptive statistics have been calculated for all 14 items, and the results are presented in Table 1. The average values of manifest variables are between 3.40 and 4.36. The item regarding a visual appeal has been rated with the lowest grade (*The visual appeal of this digital service keeps my attention*). In contrast, the item *I achieved what I need with a little effort* has been marked with the highest average score. Skewness (SK) and Kurtosis (KU) were calculated, and values for each manifest variable do not exceed the suggested threshold of [-3,+3] (Brown, 2006), so it can be concluded that variables are normally distributed, and further steps can be taken.

Table 1. Descriptive statistics

Items	AVG	SK	KU
U_1 I do not have any problems navigating this digital service.	4.27	-0.990	0.441
U_2 I don't need to click many times to find what I'm looking for.	4.01	-0.644	-0.219
U_3 The layout of the digital service interface elements is appropriate.	4.00	-0.625	0.023
U_4 Digital services are easy to use.	4.18	-0.687	-0.008
U_5 I achieve what I need with a little effort.	4.36	-1.081	0.915
P_1 The visual appeal of this digital service keeps my attention.	3.40	-0.300	-0.253
P_2 The appearance of this digital service is fit for purpose.	4.09	-0.995	1.603
P_3 When I use this digital service, my eyes do not feel tired.	3.69	-0.646	-0.113
P_4 I use all the functionalities of the digital service without stress.	3.88	-0.737	0.114
S_1 I want to use all the functionalities offered to me as part of this digital service.	4.09	-0.797	0.238
S_2 Using this digital service has become part of my routine.	3.93	-0.879	0.305
S_3 Using a digital service does not cause me any frustration.	3.79	-0.718	-0.175
S_4 I am satisfied with this digital service.	4.07	-0.787	0.496

čemu su odgovori varirali od 1 (potpuno se slažem) pa sve do 5 (potpuno se ne slažem).

Nakon završenog prikupljanja podataka, korištene su statističke metode (Hair et al., 2010). Konkretno, korištena je konfirmatorna faktorska analizu (CFA) kako bi se istražila dimenzionalnost konstrukta korisničkog iskustva. Analiza podataka provedena je pomoću programa IBM SPSS Statistics i IBM SPSS Amos za modeliranje strukturnim jednadžbama (SEM). Budući da teorija ukazuje na višedimenzionalnost UXa, predloženi model ispitan je kao konstrukt višeg reda (Higher Order Construct, HOC) (Awang et al., 2017; Hair et al., 2010).

4 Rezultati

Deskriptivna statistika je izračunata za svih 14 predloženih čestica (manifestnih varijabli), a rezultati su prikazani u Tablici 1. Prosječne vrijednosti manifestnih varijabli (AVG) kreću se između 3.40 i 4.36. Čestica koja se odnosi na vizualnu privlačnost dobila je najnižu ocjenu (*Vizualna privlačnost ove digitalne usluge zadržava moju pažnju*). S druge strane, stavka "*Postigao sam ono što trebam s malo napora*" ima najvišu prosječnu ocjenu. Izračunati su koeficijenti asimetrije (SK) i zaobljenosti (KU) te vrijednosti za svaku manifestnu varijablu ne prelaze predloženi prag [-3, +3] (Brown, 2006), pa se može zaključiti da su varijable normalno distribuirane i da se mogu poduzeti daljnji koraci u analizi.

Tablica 1. Deskriptivna statistika

Čestice	AVG	SK	KU
U_1 Nemam problema sa snalaženjem unutar ove digitalne usluge.	4.27	-0.990	0.441
U_2 Ne trebam puno puta kliknuti da pronađem što tražim.	4.01	-0.644	-0.219
U_3 Raspored elemenata sučelja digitalne usluge je prikladan.	4.00	-0.625	0.023
U_4 Digitalna usluga jednostavna je za korištenje.	4.18	-0.687	-0.008
U_5 Uz malo truda ostvarim ono što trebam kroz ovu digitalnu uslugu.	4.36	-1.081	0.915
P_1 Izgled ove digitalne usluge zadržava mi pažnju.	3.40	-0.300	-0.253
P_2 Izgled ove digitalne usluge primjeren je svrsi.	4.09	-0.995	1.603
P_3 Kada koristim ovu digitalnu uslugu ne osjećam zamor u očima.	3.69	-0.646	-0.113
P_4 Koristim sve funkcionalnosti digitalne usluge bez stresa.	3.88	-0.737	0.114

S_5 The satisfaction with use is in line with my expectations.	4.00	-0.743	0.413

The proposed model included three dimensions: appeal, satisfaction, and usability.

Further, CFA was conducted in IBM AMOS, and the results are shown in the Figure. 2.

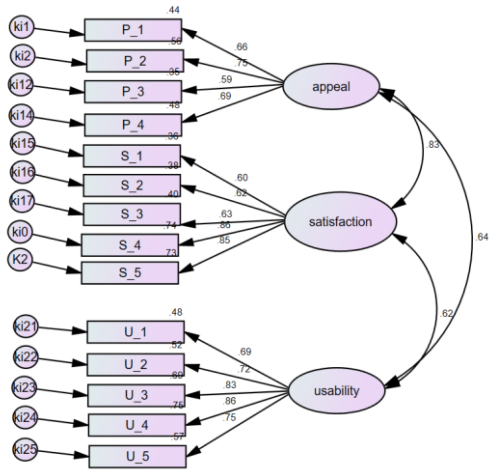


Figure 2. CFA results

To evaluate reliability and validity, factor loading (FL), average variance extracted (AVE), and composite reliability (CR) have been calculated and shown in Table 2. According to the results presented in Table 2, the scale demonstrates strong reliability as evidenced by the fact that the CR for each dimension (usability, appeal, and satisfaction) exceeds 0.70 and the AVE exceeds 0.50 (Hair et al., 2017).

Additionally, FLs have been calculated and presented to assess the reliability of individual items. It is worth noting that the smallest acceptable FL value is 0.4, but only if the AVE is greater than or equal to 0.5 (Hair et al., 2017). Based on these findings, it can be concluded that there is no need for further item deletion.

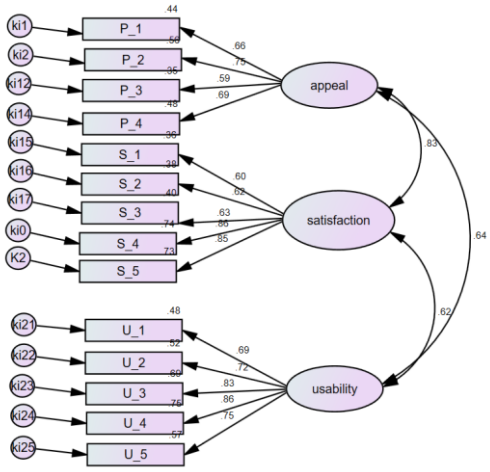
Table 2. Reliability and validity results

Model 1				
Dimension	**Items**	**FL**	**AVE**	**CR**
Usability	U_1	0.69	0.6	0.78
	U_2	0.72		
	U_3	0.83		
	U_4	0.86		
	U_5	0.75		
Appeal	P_1	0.75	0.5	0.72
	P_2	0.59		
	P_3	0.69		
	P_4	0.66		
Satisfaction	S_1	0.60	0.5	0.72
	S_2	0.62		

S_1 Želim koristit sve funkcionalnosti koje mi se nude u sklopu ove digitalne usluge.	4.09	-0.797	0.238
S_2 Korištenje ove digitalne usluge postalo mi je dio rutine.	3.93	-0.879	0.305
S_3 Korištenje digitalne usluge ne izaziva mi frustraciju.	3.79	-0.718	-0.175
S_4 Zadovoljan/a sam što sam koristim ovu uslugu.	4.07	-0.787	0.496
S_5 Zadovoljstvo korištenja u skladu je s mojim očekivanjima.	4.00	-0.743	0.413

Predloženi model sadrži tri dimenzije: privlačnost, zadovoljstvo i upotrebljivost.

Nadalje, CFA je napravljena uz pomoć IBM AMOS alata, a rezultati su prikazani na slici u nastavku (Slika 2).



Slika 2. Rezultati CFA

U svrhu procijene pouzdanost i valjanost, izračunato je faktorsko opterećenje (engl. *Factor loading*, FL), prosječna ekstrahirana varijanca (AVE) te kompozitna pouzdanost (CR). Navedeno je prikazano u Tablici 2. Prema rezultatima, instrument pokazuje snažnu pouzdanost, što potvrđuje činjenica da je CR za svaku dimenziju (upotrebljivost, privlačnost i zadovoljstvo) veći od 0,70, a AVE je za sve veći od 0,50 (Hair et al., 2017).

Dodatno, izračunata su i prikazana faktorska opterećenja (FL) kako bi se procijenila pouzdanost pojedinačnih čestica. Važno je napomenuti da najmanja prihvatljiva vrijednost FLa iznosi 0.4, ali samo ako je AVE veći ili jednak 0.5 (Hair et al., 2017). Na temelju rezultata može se zaključiti da nema potrebe za daljnjim izbacivanjem čestica.

	S_3	0.63		
	S_4	0.86		
	S_5	0.85		

The next step was to calculate discriminate validity, and for that purpose, Fornell & Larcker criterion was applied (Afthanorhan et al., 2021). Calculated results do not meet the criterion since the correlation between appeal and satisfaction exceeds square root AVE.

Table 3. Discriminant validity results

	Usability	Appeal	Satisfaction
Usability	0.78	0.64	0.62
Appeal	0.64	0.68	*0.83*
Satisfaction	0.62	*0.83*	0.72

As suggested by (Hair et al., 2010), model re-specification has to be done by eliminating manifest variables with the lowest factor loading. Two manifest variables were excluded, and the number of proposed dimensions was reduced to two (satisfaction and usability). CFA results are shown in Fig. 3.

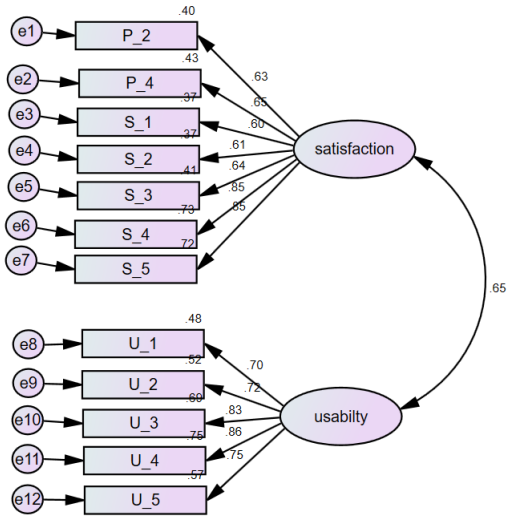


Figure 3. CFA results for the second model

Calculated FL, AVE, and CR are shown in Table 4. All values meet the suggested reference values.

Table 4. Reliability and validity results for the second model.

Model 2				
Dimension	**Items**	**FL**	**AVE**	**CR**
Usability	U_1	0.70	0.6	0.88
	U_2	0.72		
	U_3	0.83		
	U_4	0.86		
	U_5	0.75		

Tablica 2. Rezultati pouzdanosti i valjanosti

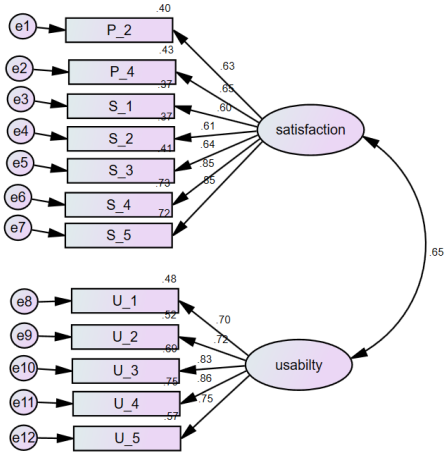
Model 1				
Dimenzija	**Čestica**	**FL**	**AVE**	**CR**
Upotrebljivost	U_1	0.69	0.6	0.78
	U_2	0.72		
	U_3	0.83		
	U_4	0.86		
	U_5	0.75		
Privlačnost	P_1	0.75	0.5	0.72
	P_2	0.59		
	P_3	0.69		
	P_4	0.66		
Zadovoljstvo	S_1	0.60	0.5	0.72
	S_2	0.62		
	S_3	0.63		
	S_4	0.86		
	S_5	0.85		

Sljedeći korak bio je izračun diskriminantne valjanosti, a u tu svrhu primijenjen je Fornell i Larckerov kriterij (Afthanorhan et al., 2021). Izračunati rezultati ne zadovoljavaju kriterij jer korelacija između privlačnosti i zadovoljstva premašuje korijen kvadrata AVEa.

Tablica 3. Rezultati diskriminantne valjanosti

	Upotrebljivost	Privlačnost	Zadovoljstvo
Upotrebljivost	**0.78**	0.64	0.62
Privlačnost	0.64	**0.68**	*0.83*
Zadovoljstvo	0.62	*0.83*	**0.72**

Kao što je sugerirano od strane autora (Hair et al., 2010), re-specifikacija modela mora biti izvedena eliminacijom čestica s najnižim faktorskim opterećenjem. Dvije čestice su isključene, te je broj predloženih dimenzija smanjen na dvije (zadovoljstvo i upotrebljivost). Rezultati CFA-a prikazani su na Slici 3.



Slika 3. Rezultati CFA za drugi model

Appeal	P_2	0.63	0.5	0.86
	P_4	0.65		
	S_1	0.60		
	S_2	0.61		
	S_3	0.64		
	S_4	0.85		
	S_5	0.85		

After checking the validity and reliability of the two proposed dimensions, Fornell & Larcker criterion was again applied, and the criterion was met. Discriminant validity was achieved, and the results are shown in Table 5.

Table 5. Discriminant validity results for the second model.

	Usability	Satisfaction
Usability	0.78	0.65
Satisfaction	0.65	0.70

According to (Hair et al., 2010), the following step was to estimate model fit measures. To evaluate the model fit, ratio of chi-square and degrees of freedom (CMIN/df), root mean square error of approximation (RMSEA), and Goodness-of-fit (GFI) have been calculated (Browne & Cudeck, 1992; Hair et al., 2010; Marsh & Hocevar, 1985). Results are shown in Table 6. And all calculated values are within the acceptable range.

Table 6. Model fit results.

	Results	**Acceptable range**
CMIN/df	3.258	[2, 5]
RMSEA	0.064	<= 0.08
GFI	0.950	>=0.9

The final results of the UX construct as HOC consisting of two dimensions are shown in the Figure. 4.

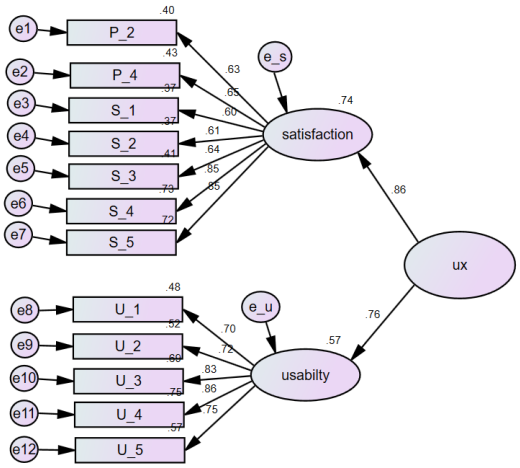


Figure 4. UX as HOC

Izračunati su FL, AVE te CR, a rezultati su prikazani u tablici 4. Svi pokazatelji zadovoljavaju predložene referentne vrijednosti.

Tablica 4. Rezultati pouzdanosti i valjanosti za drugi model

Model 2				
Dimenzija	**Čestica**	**FL**	**AVE**	**CR**
Upotrebljivost	U_1	0.70	0.6	0.88
	U_2	0.72		
	U_3	0.83		
	U_4	0.86		
	U_5	0.75		
Zadovoljstvo	P_2	0.63	0.5	0.86
	P_4	0.65		
	S_1	0.60		
	S_2	0.61		
	S_3	0.64		
	S_4	0.85		
	S_5	0.85		

Nakon provjere valjanosti i pouzdanosti dviju predloženih dimenzija, ponovno je primijenjen Fornell & Larcker kriterij, te je kriterij zadovoljen. Diskriminantna valjanost je postignuta, a rezultati su prikazani u tablici 5.

Table 5. Rezultati diskriminantne valjanosti za drugi model.

	Upotrebljivost	Zadovoljstvo
Upotrebljivost	0.78	0.65
Zadovoljstvo	0.65	0.70

Prema (Hair et al., 2010.), sljedeći korak bio je procijeniti mjere adekvatnosti modela (engl. *Model fit*). Za procjenu adekvatnosti modela izračunati su omjer hi kvadrata i stupnjeva slobode (CMIN/df), korijen srednje kvadratne pogreške aproksimacije (RMSEA) i GFI (engl. *Goodnes of fit*) (Browne & Cudeck, 1992.; Hair i sur., 2010.; Marsh i Hočevar, 1985.). Rezultati su prikazani u tablici 6. te su sve izračunate vrijednosti unutar prihvatljivog raspona.

Tablica 6. Rezultat prikladnosti modela

	Rezultati	**Raspon prihvaćanja**
CMIN/df	3.258	[2, 5]
RMSEA	0.064	<= 0.08
GFI	0.950	>=0.9

Končani rezultati za konstruktu UX, promatran kao HOC prikazan je u nastavku. Potvrđene su dvije dimenzije (Slika 4).

5 Discussion

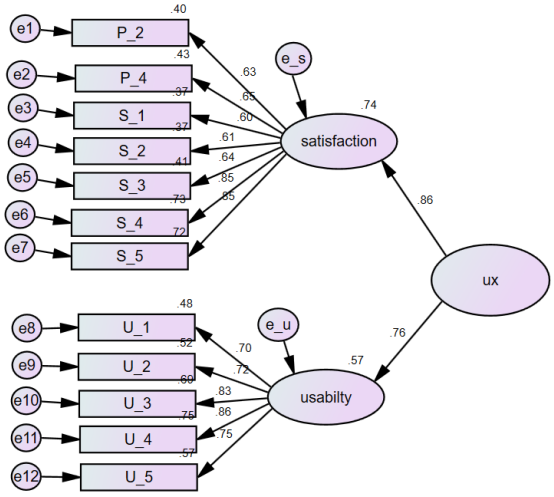
This paper aimed to examine and validate the UX model as a multidimensional construct in the HE setting. Results of a conducted survey among students confirmed the multidimensionality of UX. However, from the initially proposed three dimensions (*usability*, *appeal,* and *satisfaction*), only two dimensions were confirmed. Precisely, *usability* as one of the subdimension was confirmed in the first and the second model, but due to a strong correlation between *satisfaction* and *appeal*, the first model had to be re-specified and reduced to two dimensions. These results are inconsistent with the assumption that there is no mutual influence of the two UX aspects, made by authors who proposed measuring UX as a composite of three elements (Thüring & Mahlke, 2007).

Two items removed when the second model was proposed are *The visual appeal of this digital service keeps my attention*, and *I do not feel tired after using this digital service*. Both items (manifest variables) are related to dimension appeal. This could be explained by the fact that digital services that were the subject of this study are something that students need to use.

Precisely, those services do not provide any pleasure or additional benefits but are fully related to the process of studying (such as applying for the exam, requesting different proofs, accessing the schedule, etc.). After the model was re-specified, two remaining items related to *appeal* were assigned to dimension *satisfaction*. These results align with previous research that argued for two aspects of UX: **pragmatic** and **hedonic**. The pragmatic aspect refers to *usability*, while the hedonic dimension refers to *emotions* and *aesthetics* (Preece et al., 2019; Wani et al., 2017).

While there are studies related to satisfaction and quality of e-learning systems in Croatia (eg., Bušljeta Kardum, Jurić Vukelić, 2021; Klasnić et al, 2008), up to the author s knowledge, this is the first paper applying the HOC approach for validating comprehensive UX construct in an HE setting. The findings validate the significance of taking into account all aspects while assessing the UX. Findings are also in line with previous definitions considering the UX as an extension of usability (but not limited to), however, including also personal subjective perceptions and emotions (Sonderegger et al., 2019).

As for the results, students in Croatian HEIs rated the visual appeal of digital services with the lowest rate (average score was 3.4); however, they do agree that the visual of digital services observed fits the purpose. Similarly, Croatian students, in a previous study, evaluated satisfaction with the visual appearance of the e-learning system also with the similar average grade (3.42) (Klasnic et al, 2008). As for the usability dimension, the average score is



Slika 4. UX kao HOC

5 Diskusija

Cilj ovog rada je bio istražiti i potvrditi model korisničkog iskustva kao višedimenzionalni konstrukt u kontekstu VO. Rezultati provedene ankete među studentima potvrdili su višedimenzionalnost UX-a. Međutim, od inicijalno predložene tri dimenzije (*upotrebljivost*, *privlačnost* i *zadovoljstvo*), potvrđene su samo dvije dimenzije. Konkretno, upotrebljivost kao jedna od dimenzija potvrđena je u prvom i drugom modelu, ali zbog snažne korelacije između *zadovoljstva* i *privlačnosti*, prvi model je morao biti ponovno definiran te je sveden na dvije dimenzije. Dobiveni rezultati nisu u skladu s pretpostavkom autora koji su predložili mjerenje UXa kao kombinaciju triju elementa (Thüring & Mahlke, 2007)

Dvije čestice su uklonjene prilikom formiranja drugog modela: *Vizualna privlačnost ove digitalne usluge zadržava moju pažnju* i *Ne osjećam umor nakon korištenja ove digitalne usluge*. Obje čestice (manifestne varijable) odnose se na dimenziju privlačnosti. To bi se moglo objasniti činjenicom da su digitalne usluge koje su bile predmet ovog istraživanja nešto što studenti moraju koristiti.

Konkretno, te usluge ne pružaju nikakvo zadovoljstvo ili dodatne pogodnosti, već su potpuno povezane s procesom studiranja (kao što su prijava za ispit, traženje različitih dokaza, pristup rasporedu itd.). Nakon ponovne specifikacije modela, dvije preostale čestice povezane s *privlačnošću* dodijeljene su dimenziji *zadovoljstva*. Ovi rezultati u skladu su s prethodnim istraživanjima koja zagovaraju dva aspekta UX-a: **pragmatični** i **hedonički**. Pragmatični aspekt odnosi se na *upotrebljivost*, dok se hedonička dimenzija odnosi na *emocije* i *estetiku* (Preece et al., 2019; Wani et al., 2017).

Iako postoje istraživanja o zadovoljstvu i kvaliteti sustava e-učenja u Hrvatskoj (npr. Bušljeta Kardum, Jukić Vukelić, 2021; Klasnić i sur., 2008), prema

above 4, implying that students do not report usability issues.

There are a few limitations that should be addressed in further research. The first limitation is regarding the type of digital services in HE. As mentioned, observed digital services are the ones related to support/administrative activities. It would be interesting to measure UX with digital services that provide additional content for students (extracurricular activity, not content closely related to the studying process).

Even though participants of this study evaluated their UX higher than students in a similar study (Al-Hunaiyyan et al., 2021), there are few practical implications for HEIs that could be derived. Even though the usability dimension is important, the hedonic dimension should not be neglected. HEIs need to invest effort in designing those digital services, so they provide students with higher UX.

Providing better UX of those digital services could help improve students' satisfaction in general, which is proved to be one of the factors when measuring the quality of HEIs (Cadena et al., 2018). Results confirm the importance of measuring students' UX, especially since the number of these services will keep increasing.

Another limitation is geographical since all participants are Croatian students. It would be interesting to conduct similar research in another EU country to see whether there are any differences.

Future research should also focus on examining the impact of UX on constructs related to student satisfaction with HEIs in general.

6 Conclusion

The multidimensionality of UX in HE has been confirmed, and both hedonic and pragmatic aspects reviled as important when measuring the UX. To conclude, meaning UX in HE is particularly important for several reasons. First, the new generation of students expects high UX. Second, the number of services is increasing, and these services need to be properly designed to enable effective and good experiences. Providing optimal UX benefits not only students but also faculty and staff by streamlining their work processes and improving their efficiency. Institutions that understand the significance of employing modern technology and offering a seamless UX can potentially gain a competitive advantage. Therefore, HEIs must prioritize measuring and enhancing their UX to meet students' increasing expectations and remain competitive in the global market. Results of quantitative survey conducted on over 500 Croatian students reviled that visual appeal is graded with the lowest grade (3.4), while usability aspect regarding effort needed received the highest grade (4.36).

dosadašnjim saznanjima autora, ovo je prvi rad koji istražuje mjerenje sveobuhvatnog UXa kao višedimenzionalnog konstrukta i to u kontekstu visokog obrazovanja. Rezultati potvrđuju važnost mjerenja svih dimenzija prilikom vrednovanja UXa. Rezultati su također u skladu s prethodnim definicijama koje smatraju da je UX proširenje *upotrebljivosti* (ali ne ograničeno na to), uključujući i osobne subjektivne percepcije i emocije (Sonderegger et al., 2019).

Što se tiče rezultata, studenti na hrvatskim institucijama VO najniže su ocijenili vizualnu privlačnost digitalnih usluga (prosječna ocjena je bila 3.4); međutim, slažu se da vizualni aspekt digitalnih usluga odgovara svrsi. Slično tome, hrvatski su studenti, u prethodnom provedenom istraživanju, vrednovali zadovoljstvo vizualnim izgledom sustava za e-učenje i ocijenili ga slično (3,42) (Klasnic i sur., 2008.). Što se tiče dimenzije upotrebljivosti, prosječna ocjena je iznad 4, što implicira da studenti ne prijavljuju probleme s upotrebljivošću.

Postoji nekoliko ograničenja koja bi trebala biti adresirana u budućim istraživanjima. Prvo ograničenje odnosi se na vrstu digitalnih usluga u VOKao što je spomenuto, promatrane digitalne usluge odnose se na podršku/administrativne aktivnosti. Bilo bi zanimljivo izmjeriti UX digitalnih usluga koje pružaju dodatne sadržaje za studente (izvannastavne aktivnosti, sadržaje koji nisu usko povezani s procesom studiranja).

Iako su sudionici ovog istraživanja ocijenili UX višom ocijenom nego studenti u sličnom istraživanju (Al-Hunaiyyan et al., 2021), postoje neke praktične implikacije za institucije VO koje se mogu derivirati. Iako je dimenzija upotrebljivosti važna, hedonički aspekt ne smije biti zanemaren. Institucije trebaju uložiti napore u dizajniranje tih digitalnih usluga kako bi studentima pružile bolji UX. Osiguravanje visokre razine UXa promatranih digitalnih usluga može pomoći poboljšanju zadovoljstva studenata općenito, što se pokazalo kao jedan od čimbenika pri mjerenju kvalitete visokoškolskih ustanova (Cadena et al., 2018). dodatno, rezultati potvrđuju važnost mjerenja UXa studenata, pogotovo jer će se broj tih usluga nastaviti povećavati.

Drugo ograničenje ovog rada je geografsko, budući da su svi sudionici hrvatski studenti. Bilo bi zanimljivo provesti slično istraživanje u drugoj zemlji Europske unije kako bi se vidjele eventualne razlike.

Buduća istraživanja također bi trebala usmjeriti pažnju na ispitivanje utjecaja UXa na konstrukte koji su povezani sa zadovoljstvom studenata u visokom obrazovanju općenito.

6 Zaključak

Potvrđena je višedimenzionalnost UX-a u kontekstu visokog obrazovanja, pri čemu su i hedonički i pragmatički aspekti istaknuti kao važni pri mjerenju

The UX methodology gives priority to the needs and preferences of students (users) when creating a service experience that is easy to use and meets their goals. This approach requires a deep understanding of user behavior and careful execution to make it simple for users to accomplish what they want. Despite the growing importance of UX in attracting and retaining students, HEIs still lack to prioritize this area. The ongoing pandemic, declining revenues, rapidly evolving demographics, and a surge in the number of students accessing HE through digital devices still pose significant challenges for HEIs.

Acknowledgment

The research is supported by the Croatian Science Foundation [grant number IRP-2017-05-7625].

References

Afthanorhan, A., Ghazali, P. L., & Rashid, N. (2021). Discriminant Validity: A Comparison of CBSEM and Consistent PLS using Fornell & Larcker and HTMT Approaches. *Journal of Physics: Conference Series*, *1874*(1), 012085. https://doi.org/10.1088/1742-6596/1874/1/012085

Al-Hunaiyyan, A., Alhajri, R., Alghannam, B., & Al-Shaher, A. (2021). Student Information System: Investigating User Experience (UX). *International Journal of Advanced Computer Science and Applications*, *12*(2), 80–87. https://doi.org/10.14569/IJACSA.2021.0120210

Awang, Z., Afthanorhan, A., Mamat, M., & Aimran, N. (2017). Modeling Structural Model for Higher Order Constructs (HOC) Using Marketing Model. *World Applied Sciences Journal*, *35*(8), 1434–1444. https://doi.org/10.5829/idosi.wasj.2017.1434.1444

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Bušljeta Kardum, R. & Jurić Vukelić, D. (2021). The Challenges and Issues on the University of Zagreb during COVID-19 Crisis. Interdisciplinary *Description of Complex Systems*, 19 (3), 357-365. https://doi.org/10.7906/indecs.19.3.1

Bygstad, B., Øvrelid, E., Ludvigsen, S., & Dæhlen, M. (2022). From dual digitalization to digital

UXa. Zaključno, značaj UXa u HE-u posebno je važan iz nekoliko razloga. Prvo, nova generacija studenata očekuje visok UX. Drugo, broj usluga se povećava, a te usluge treba adekvatno dizajnirati kako bi omogućile učinkovita i dobra iskustva. Pružanje optimalnog UXa koristi ne samo studentima, već i nastavnom osoblju i osoblju podrške putem pojednostavljenja njihovih radnih procesa i poboljšanja njihove učinkovitosti. Institucije koje shvaćaju važnost korištenja modernih tehnologija i pružanja besprijekornog UXa mogu potencijalno ostvariti konkurentske prednosti. Stoga visokoškolske ustanove moraju dati prioritet mjerenju i poboljšanju svojeg UX-a kako bi zadovoljile sve veća očekivanja studenata i ostale konkurentne na globalnom tržištu. Rezultati kvantitativnog istraživanja, provedenog na više od 500 hrvatskih studenata, pokazali su da je vizualna privlačnost ocijenjena najnižom ocjenom (3,4), dok je aspekt upotrebljivosti u pogledu potrebnog truda dobio najvišu ocjenu (4,36).

Metodologija UXa daje prednost potrebama i preferencijama studenata (korisnika) prilikom stvaranja iskustva usluge koje je jednostavno za korištenje i koje zadovoljava njihove ciljeve. Ovaj pristup zahtijeva duboko razumijevanje ponašanja korisnika i pažljivu izvedbu kako bi korisnicima bilo jednostavno postići ono što žele. Unatoč sve većem značaju UX-a u privlačenju i zadržavanju studenata, institucije VO još uvijek nedovoljno tretiraju ovu područje. Nastavak pandemije, pad prihoda, brza demografska promjena i porast broja studenata koji pristupaju VO putem digitalnih uređaja i dalje predstavljaju značajne izazove za institucije.

Zahvale

Ovaj je rad financirala Hrvatska zaklada za znanost projektom UIP-2017-05-7625.

Reference

Afthanorhan, A., Ghazali, P. L., & Rashid, N. (2021). Discriminant Validity: A Comparison of CBSEM and Consistent PLS using Fornell & Larcker and HTMT Approaches. *Journal of Physics: Conference Series*, *1874*(1), 012085. https://doi.org/10.1088/1742-6596/1874/1/012085

Al-Hunaiyyan, A., Alhajri, R., Alghannam, B., & Al-Shaher, A. (2021). Student Information System: Investigating User Experience (UX). *International Journal of Advanced Computer Science and Applications*, *12*(2), 80–87. https://doi.org/10.14569/IJACSA.2021.0120210

Awang, Z., Afthanorhan, A., Mamat, M., & Aimran, N. (2017). Modeling Structural Model for Higher Order Constructs (HOC) Using Marketing Model. *World Applied Sciences Journal*, *35*(8), 1434–

learning space: Exploring the digital transformation of higher education. *Computers and Education*, *182*(August 2021). https://doi.org/10.1016/j.compedu.2022.104463

Cadena, S., García, J. L., Loza-Aguirre, E., Ortiz, J., Pérez, A., & Segura-Morales, M. (2018). Measuring Quality of Higher Education. *EDULEARN18 Proceedings*, *1*(July), 10214–10219. https://doi.org/10.21125/edulearn.2018.2484

Calle-Jimenez, T., Sanchez-Gordon, S., & Arias-Flores, H. (2021). Profiling of E-Learning Users with Accessibility Needs. In *Lecture Notes in Networks and Systems* (Vol. 275). https://doi.org/10.1007/978-3-030-80091-8_56

Caushi, B., & Dika, Z. (2018). Higher Education Information Systems : An Overview of the Latest Trends and Issues. *VIIIth Annual International Meeting of Alb-Science Institute, March 2018*.

Díaz, O. P., Ribes-Giner, G., & Perello-Marin, M. R. (2016). The impact of cocreation on the student satisfaction: Analysis through structural equation modeling. *Abstract and Applied Analysis*, *2016*. https://doi.org/10.1155/2016/3729791

Gilch, H., Beise, A. S., Krempkow, R., Muller, M., Stratmann, F., & Wannemacher, K. (2019). *Digitalisierung der Hochschulen: Ergebnisse einer Schwerpunktstudie für die Expertenkommission Forschung und Innovation*.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Pearson Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd Ed.* SAGE Publications.

Halvorsrud, R., Kvale, K., & Følstad, A. (2016). Improving service quality through customer journey analysis. *Journal of Service Theory and Practice*, *26*(6), 840–867. https://doi.org/10.1108/JSTP-05-2015-0111

Hassenzahl, M. (2005). Hedonic, emotional, and experiential perspectives on product quality. *Encyclopedia of Human Computer Interaction*, *2000*, 266–272. https://doi.org/10.4018/978-1-59140-562-7.ch042

Hassenzahl Marc. (2003). *THE THING AND I: UNDERSTANDING THE RELATIONSHIP BETWEEN USER AND PRODUCT*. 31–42. doi: 10.1007/1-4020-2967-5_4%0Ahttps://link.springer.com/chapter/10.1007%2F1-4020-2967-5_4

1444. https://doi.org/10.5829/idosi.wasj.2017.1434.1444

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Bušljeta Kardum, R. & Jurić Vukelić, D. (2021). The Challenges and Issues on the University of Zagreb during COVID-19 Crisis. Interdisciplinary *Description of Complex Systems*, 19 (3), 357-365. https://doi.org/10.7906/indecs.19.3.1

Bygstad, B., Øvrelid, E., Ludvigsen, S., & Dæhlen, M. (2022). From dual digitalization to digital learning space: Exploring the digital transformation of higher education. *Computers and Education*, *182*(August 2021). https://doi.org/10.1016/j.compedu.2022.104463

Cadena, S., García, J. L., Loza-Aguirre, E., Ortiz, J., Pérez, A., & Segura-Morales, M. (2018). Measuring Quality of Higher Education. *EDULEARN18 Proceedings*, *1*(July), 10214–10219. https://doi.org/10.21125/edulearn.2018.2484

Calle-Jimenez, T., Sanchez-Gordon, S., & Arias-Flores, H. (2021). Profiling of E-Learning Users with Accessibility Needs. In *Lecture Notes in Networks and Systems* (Vol. 275). https://doi.org/10.1007/978-3-030-80091-8_56

Caushi, B., & Dika, Z. (2018). Higher Education Information Systems : An Overview of the Latest Trends and Issues. *VIIIth Annual International Meeting of Alb-Science Institute, March 2018*.

Díaz, O. P., Ribes-Giner, G., & Perello-Marin, M. R. (2016). The impact of cocreation on the student satisfaction: Analysis through structural equation modeling. *Abstract and Applied Analysis*, *2016*. https://doi.org/10.1155/2016/3729791

Gilch, H., Beise, A. S., Krempkow, R., Muller, M., Stratmann, F., & Wannemacher, K. (2019). *Digitalisierung der Hochschulen: Ergebnisse einer Schwerpunktstudie für die Expertenkommission Forschung und Innovation*.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Pearson Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd Ed.* SAGE Publications.

Halvorsrud, R., Kvale, K., & Følstad, A. (2016). Improving service quality through customer

Hinderks, A., Schrepp, M., Domínguez Mayo, F. J., Escalona, M. J., & Thomaschewski, J. (2019). Developing a UX KPI based on the user experience questionnaire. *Computer Standards and Interfaces*, *65*(February), 38–44. https://doi.org/10.1016/j.csi.2019.01.007

Hornbæk, K., & Hertzum, M. (2017). Technology acceptance and user experience: A review of the experiential component in HCI. *ACM Transactions on Computer-Human Interaction*, *24*(5). https://doi.org/10.1145/3127358

Khalid, J., Ram, B., & Khalee, M. (2018). *Promising digital university : a pivotal need for higher education transformation Promising digital university : a pivotal need for higher education transformation Jamshed Khalid *, Braham Rahul Ram , Mohamed Soliman , Anees Janee Ali , Muhammad Khaleel a*. *12*(January), 13. https://www.researchgate.net/publication/324844227%0APromising

Klasnić, K., Seljan, S. & Stančić, H. (2008). *Quality parameters for the e-learning Omega system* in *Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces*, 519-526.

Langgawan Putra, M. G., Yogaswara, R., & Alfani Putera, M. I. (2021). Analysis Effect of User Experience on Understanding Rate of Student Using Academic Information System in Higher Education with Honeycomb Method. *2021 6th International Conference on Informatics and Computing, ICIC 2021*. https://doi.org/10.1109/ICIC54025.2021.9632997

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562–582. https://doi.org/10.1037/0033-2909.97.3.562

Martins, J., Branco, F., Gonçalves, R., Au-Yong-Oliveira, M., Oliveira, T., Naranjo-Zolotov, M., & Cruz-Jesus, F. (2019). Assessing the success behind the use of education management information systems in higher education. *Telematics and Informatics*, *38*, 182–193. https://doi.org/10.1016/j.tele.2018.10.001

Mijač, T., Jadrić, M., & Ćukušić, M. (2023). The role of user experience and co-creation in measuring the success of digital services in higher education. *Behaviour and Information Technology*, 1–10. https://doi.org/10.1080/0144929X.2023.2206924

journey analysis. *Journal of Service Theory and Practice*, *26*(6), 840–867. https://doi.org/10.1108/JSTP-05-2015-0111

Hassenzahl, M. (2005). Hedonic, emotional, and experiential perspectives on product quality. *Encyclopedia of Human Computer Interaction*, *2000*, 266–272. https://doi.org/10.4018/978-1-59140-562-7.ch042

Hassenzahl Marc. (2003). *THE THING AND I: UNDERSTANDING THE RELATIONSHIP BETWEEN USER AND PRODUCT*. 31–42. doi: 10.1007/1-4020-2967-5_4%0Ahttps://link.springer.com/chapter/10.1007%2F1-4020-2967-5_4

Hinderks, A., Schrepp, M., Domínguez Mayo, F. J., Escalona, M. J., & Thomaschewski, J. (2019). Developing a UX KPI based on the user experience questionnaire. *Computer Standards and Interfaces*, *65*(February), 38–44. https://doi.org/10.1016/j.csi.2019.01.007

Hornbæk, K., & Hertzum, M. (2017). Technology acceptance and user experience: A review of the experiential component in HCI. *ACM Transactions on Computer-Human Interaction*, *24*(5). https://doi.org/10.1145/3127358

Khalid, J., Ram, B., & Khalee, M. (2018). *Promising digital university : a pivotal need for higher education transformation Promising digital university : a pivotal need for higher education transformation Jamshed Khalid *, Braham Rahul Ram , Mohamed Soliman , Anees Janee Ali , Muhammad Khaleel a*. *12*(January), 13. https://www.researchgate.net/publication/324844227%0APromising

Klasnić, K., Seljan, S. & Stančić, H. (2008). *Quality parameters for the e-learning Omega system* in *Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces*, 519-526.

Langgawan Putra, M. G., Yogaswara, R., & Alfani Putera, M. I. (2021). Analysis Effect of User Experience on Understanding Rate of Student Using Academic Information System in Higher Education with Honeycomb Method. *2021 6th International Conference on Informatics and Computing, ICIC 2021*. https://doi.org/10.1109/ICIC54025.2021.9632997

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562–582. https://doi.org/10.1037/0033-2909.97.3.562

Martins, J., Branco, F., Gonçalves, R., Au-Yong-Oliveira, M., Oliveira, T., Naranjo-Zolotov, M., & Cruz-Jesus, F. (2019). Assessing the success

Ng, I. C. L., & Forbes, J. (2009). Education as service: The understanding of university experience through the service logic. *Journal of Marketing for Higher Education*, *19*(1), 38–64. https://doi.org/10.1080/08841240902904703

Obaid, T. (2019). *Digital Transformation in Higher Education UniSZA Case Study*. *February*.

Pasini, A., Estevez, E., & Pesado, P. (2019). Assessment Model for Digital Services provided by Higher Education Institutions. *ACM International Conference Proceeding Series*, 468–477. https://doi.org/10.1145/3325112.3325268

Preece, J., Rogers, Y., & Sharp, H. (2019). *Interaction design; Beyond human-computer interaction* (5th ed.). Wiley.

Richter, M., & Fluckiger, M. (2014). *User-Centered Engineering*. Springer.

Rodrigues, L. S. (2017). Challenges of digital transformation in higher education institutions: A brief discussion. *Proceedings of the 30th International Business Information Management Association Conference, IBIMA 2017 - Vision 2020: Sustainable Economic Development, Innovation Management, and Global Growth*, *2017-Janua*(November 2017), 4490–4493.

Santana, R., Rodríguez, A., Rybarczyk, Y., Méndez, G. G., Vera, F., & Rossi, G. (2022). A Study on User Experience of Smart Glasses for Higher Education Students. *Iberian Conference on Information Systems and Technologies, CISTI*, *2022-June*(June), 22–25. https://doi.org/10.23919/CISTI54924.2022.9820326

Saplacan, D., Herstad, J., & Pajalic, Z. (2018). Feedback from digital systems used in higher education: An inquiry into triggered emotions two universal design oriented solutions for a better user experience. *Studies in Health Technology and Informatics*, *256*, 421–430. https://doi.org/10.3233/978-1-61499-923-2-421

Slade, C., & Downer, T. (2020). Students' conceptual understanding and attitudes towards technology and user experience before and after use of an ePortfolio. *Journal of Computing in Higher Education*, *32*(3), 529–552. https://doi.org/10.1007/s12528-019-09245-8

Sonderegger, A., Uebrlbacher, A., & Sauer, J. (2019). The UX Construct – Does the Usage Context Influence the Outcome of User Experience Evaluations? In *Human-Computer Interaction – INTERACT 2019* (Vol. 11749, pp. 140–157). https://doi.org/10.1007/978-3-030-29390-1

behind the use of education management information systems in higher education. *Telematics and Informatics*, *38*, 182–193. https://doi.org/10.1016/j.tele.2018.10.001

Mijač, T., Jadrić, M., & Ćukušić, M. (2023). The role of user experience and co-creation in measuring the success of digital services in higher education. *Behaviour and Information Technology*, 1–10. https://doi.org/10.1080/0144929X.2023.2206924

Ng, I. C. L., & Forbes, J. (2009). Education as service: The understanding of university experience through the service logic. *Journal of Marketing for Higher Education*, *19*(1), 38–64. https://doi.org/10.1080/08841240902904703

Obaid, T. (2019). *Digital Transformation in Higher Education UniSZA Case Study*. *February*.

Pasini, A., Estevez, E., & Pesado, P. (2019). Assessment Model for Digital Services provided by Higher Education Institutions. *ACM International Conference Proceeding Series*, 468–477. https://doi.org/10.1145/3325112.3325268

Preece, J., Rogers, Y., & Sharp, H. (2019). *Interaction design; Beyond human-computer interaction* (5th ed.). Wiley.

Richter, M., & Fluckiger, M. (2014). *User-Centered Engineering*. Springer.

Rodrigues, L. S. (2017). Challenges of digital transformation in higher education institutions: A brief discussion. *Proceedings of the 30th International Business Information Management Association Conference, IBIMA 2017 - Vision 2020: Sustainable Economic Development, Innovation Management, and Global Growth*, *2017-Janua*(November 2017), 4490–4493.

Santana, R., Rodríguez, A., Rybarczyk, Y., Méndez, G. G., Vera, F., & Rossi, G. (2022). A Study on User Experience of Smart Glasses for Higher Education Students. *Iberian Conference on Information Systems and Technologies, CISTI*, *2022-June*(June), 22–25. https://doi.org/10.23919/CISTI54924.2022.9820326

Saplacan, D., Herstad, J., & Pajalic, Z. (2018). Feedback from digital systems used in higher education: An inquiry into triggered emotions two universal design oriented solutions for a better user experience. *Studies in Health Technology and Informatics*, *256*, 421–430. https://doi.org/10.3233/978-1-61499-923-2-421

Slade, C., & Downer, T. (2020). Students' conceptual understanding and attitudes towards technology and user experience before and after use of an ePortfolio. *Journal of Computing in Higher*

Thoring, A., Rudolph, D., & Vogl, R. (2017). Digitalization of higher education from a student's point of view. *EUNIS Paper*, *47*.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human-technology interaction. *International Journal of Psychology*, *42*(4), 253–264. https://doi.org/10.1080/00207590701396674

van Staden, C. J., Kroeze, J. H., & van Biljon, J. A. (2019). User Experience Evaluation in eModeration: The Case of Higher Education Institutions in Developing Countries. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11701 LNCS*, 621–633. https://doi.org/10.1007/978-3-030-29374-1_50

Wani, M., Raghavan, V., Abraham, D., & Kleist, V. (2017). Beyond utilitarian factors: User experience and travel company website successes. *Information Systems Frontiers*, *19*(4), 769–785. https://doi.org/10.1007/s10796-017-9747-1

Wiles, A., Roberts, S., & Abdelnour-Nocera, J. (2013). Library usability in higher education: How user experience can form library policy. *IFIP Advances in Information and Communication Technology*, *407*, 139–149. https://doi.org/10.1007/978-3-642-41145-8_12

Zardari, B. A., Hussain, Z., Arain, A. A., Rizvi, W. H., & Vighio, M. S. (2021). Development and validation of user experience-based e-learning acceptance model for sustainable higher education. *Sustainability (Switzerland)*, *13*(11), 1–17. https://doi.org/10.3390/su13116201

Thoring, A., Rudolph, D., & Vogl, R. (2017). Digitalization of higher education from a student's point of view. *EUNIS Paper*, *47*.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human-technology interaction. *International Journal of Psychology*, *42*(4), 253–264. https://doi.org/10.1080/00207590701396674

van Staden, C. J., Kroeze, J. H., & van Biljon, J. A. (2019). User Experience Evaluation in eModeration: The Case of Higher Education Institutions in Developing Countries. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11701 LNCS*, 621–633. https://doi.org/10.1007/978-3-030-29374-1_50

Wani, M., Raghavan, V., Abraham, D., & Kleist, V. (2017). Beyond utilitarian factors: User experience and travel company website successes. *Information Systems Frontiers*, *19*(4), 769–785. https://doi.org/10.1007/s10796-017-9747-1

Wiles, A., Roberts, S., & Abdelnour-Nocera, J. (2013). Library usability in higher education: How user experience can form library policy. *IFIP Advances in Information and Communication Technology*, *407*, 139–149. https://doi.org/10.1007/978-3-642-41145-8_12

Zardari, B. A., Hussain, Z., Arain, A. A., Rizvi, W. H., & Vighio, M. S. (2021). Development and validation of user experience-based e-learning acceptance model for sustainable higher education. *Sustainability (Switzerland)*, *13*(11), 1–17. https://doi.org/10.3390/su13116201

Business process modeling and digital twins: a literature review

Snježana Križanić, Neven Vrček
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin
{skrizanic, nvrcek}@foi.unizg.hr

Abstract. *Business process modeling is one of the most important steps in managing the lifecycle of a business process. Business process simulation using process simulation tools is a technique used to produce estimates of business process performance and cost structure. A digital twin of a business process is a virtual model used to replicate the behaviour of a real business process. The difference between a digital twin and a business process simulation is that the simulation does not require real-time data, while the digital twin relies on real-time data. This paper is about digital twins of business processes. The achievements of existing research in the field of business process simulation and digital twins are presented. In addition, the limitations of the existing research are investigated.*

Keywords. business process, business process modeling, digital twin, Internet of Things.

1 Introduction

This paper deals with the field of business processes and digital twins and their construction. A digital twin of a business process is a business process that is represented in terms of a virtual representation. The virtual representation of a real entity consists of a physical entity in the real world, a digital twin in program form, and data connecting the two aforementioned elements (TechTarget, 2023). According to Dumas (2021), a digital twin is a model of an object or system that, together with a set of data about events related to the object or system, allows accurate prediction of the performance of the physical object or system over time. The digital twin paradigm is specific to the importance of connecting the digital twin to real-world data sources from the environment and updating the state of the digital twin in real time. Digital twins are among the most important components of the recent Industry 4.0 revolution (Lugaresi & Matta, 2021). Creating digital twins and verifying their performance helps to understand user motivations and provides information and alternative options and suggestions in a form that supports decision-making by the various participants within the system (West et al., 2021). Examples of digital twins, among others, include a digital twin of the product, a digital twin of the production facility, a digital twin of the supply chain, Google Maps, which is a digital twin of the Earth's surface, etc. Modeling the business process and its digital twin includes all steps: from conception and engineering to service delivery, which means that real data about the business process and its activities are obtained as the business process unfolds. Unlike a simulation, a digital twin processes the collected data in real-time and intervenes in the real world through a specific actuator when needed. The advantages of a digital twin of a business process are: monitoring the state of business processes, timely response to unexpected events, shorter product development time, shorter product delivery time, faster product design optimization, product quality improvement, etc.

The purpose of this paper is to examine the results of previous research in the field of business process modeling, business process simulation, and digital twins.

This study aims to address the following research questions: Firstly, what are the identified limitations in previous research? Secondly, in light of these limitations, is there a viable approach to address them effectively?

The article is organized as follows; Chapter 2 explains the methodology used to conduct the study. Chapter 3 presents a literature review that highlights the results of existing research in the field of business process simulation and digital twins. The architecture of the process digital twin is proposed in Chapter 4. Chapter 5 presents the results of existing research in the field of business process modeling and Petri Nets. Chapter 6 identifies the limitations of the previous research. The paper ends with a conclusion in Chapter 7.

2 Methodology

For the purpose of this research, a literature review was conducted in the fields of business process simulation and digital twins. The literature search was conducted using Scopus and Web of Science databases and resulted in 25 scientific papers in total. Out of 25 papers, 12 were included in the in-depth analysis, which were directly related to the research field. The query used to search both databases was:

(("business process simulation" OR "simulation of business processes") AND "digital twin")
ALL FIELDS

The search terms were "*simulation of business processes*" and "*digital twin*". All fields in the scientific databases were searched and the entire time period was considered. In this way, the search was not limited to a specific time period or to specific fields in which the research was published. The oldest research was from 2019 and the most recent was from 2023, indicating that this research field is still quite new. To support the research with additional knowledge on business process modeling, additional literature in the fields of simulation modeling and colored Petri nets was searched. The Scopus database was searched, and the search query was:

("business process modeling" AND "graph theory" AND "Coloured Petri net").

All fields of the database and the entire time period were searched. The search resulted in 38 articles, of which 6 were selected for in-depth analysis and deemed most relevant to the research. The following is an overview of the results of previous research.

3 Simulation of business processes and digital twins

This chapter presents the results of previous research in the fields of business process simulation and digital twins. Based on the analysis of the existing literature it will be determined whether there are limitations in the previous research.

The authors Perez, Wassick & Grossmann (2022) conducted research on the application of digital twin and simulation in the supply chain. The authors propose a framework for creating a virtual replica of the business transaction process in the supply chain. The framework consists of three main components: (1) a system identification module used to model the business process in the supply chain, (2) a simulation module to create and run event simulations, and (3) an optimization module to optimize the business processes in the supply chain. The fundamental elements of the proposed digital twin were developed using the Julia programming language.

In the research (Romero et al., 2021), the authors mention the application of digital twin as one of the technologies that companies could use for effective adaptation in the integration of design, engineering, and production processes, using data collected by sensors and similar methods that can create parametric designs based on the business model process. As digital twins are created for more and more products today, advanced simulations and predictive models can be used to proactively identify and correct program and device performance.

Research (Dymitrowski & Mielcarek, 2021) has stated that digital twins can simulate any aspect of a company's operations. For example, digital twins can simulate the operation of a process or device (one or more) based on documentation to eliminate potential errors before a physical device is implemented or built.

The authors (Lugaresi & Matta, 2021) conduct a study using real-time factory data flow for online development of a simulation model of a production system. The authors investigate how model generation works online during the transition phase of establishing a production system. Within production planning and control applications, event simulation can be used as a digital twin of the production system.

Dumas (2021) deals in his work with process prediction through process interventions in real-time. More specifically, how to predict the values of the effects that occur when one or more processes are executed after an intervention in a business process. The research states that the digital twin of a business process has a potential application in the field of process optimization. The purpose of a digital twin of business processes is not so much to capture the business process as it is but to think about the effects of interventions, including interventions that have not been observed before. Research cites two ways to theoretically build a digital twin of a business process: by simulating system dynamics and by using machine learning methods.

Research (West et al., 2021) states that the digital twin adds value and support to the process of real-time operational decision making. Using the examples of 10 use cases, the authors show how digital twins help create value in the business process. Some of these use cases are: tunnel maintenance and repair, wood pattern cutting, factory operational planning and material flow, factory asset management, and server room temperature management and control. Ten use cases in which concepts of the digital twin were applied produced different outcomes. All of the digital twin development concepts were in the early stages, and the digital twins were not yet fully implemented and ready for use. Therefore, a limitation of the study is that the use cases were early concepts of digital twins and not fully developed digital twins, so they are considered only as a representation of extended use case concepts. The research is based on the assumption that in a

business context, value is created when a decision is made and a certain action is performed in the system. According to the research results, the support of the digital twin in decision making can be seen at the operational, tactical, and strategic levels.

In their work, the authors (Sumereder & Woitsch, 2022) deal with digitization in the context of digital transformation, using the possibilities of a digital twin. The authors show which digitization principles/patterns are suitable for digital twin organizations and which principles/patterns are suitable for which organizational structure. The paper mentions the difference between a digital shadow and a digital twin. A digital shadow can be a hybrid version of a digital model and a twin, while a digital twin is characterized by automated data flows.

Research (Mallek-Daclin et al., 2022) indicates that the main goal of developing digital twins for a business process is to detect deviations that may appear during runtime. When deviations are detected, the digital twin suggests solutions to the users to help them make a decision, and this decision is considered the most optimal for the product development plan. The digital twin needs to retrieve data from the real system in real time to verify what is happening in the system and to determine what was predicted in the generated product development plan.

The authors (Park & van der Aalst, 2021) propose a framework for integrating the ERP system with business process simulation. The framework consists of three components: the ERP system, the simulation engine, and the transformation engine that converts commands into executions according to which the system updates the model. A simulation model represents reality in a simplified form and creates hypothetical instances of a business process, which enables the simulation of different scenarios. Successful implementation of digital twins of business processes using business process simulation depends on how accurately the simulation model represents reality.

In the research (der Landwehr, Trott & von Viebahn, 2021), the authors propose a framework for providing structural guidance to practitioners and researchers working on evaluating the individual suitability of a computer simulation-based analysis approach and the adoption of a simulation study, particularly a digital twin. The proposed framework provides a holistic view of the benefits, opportunities, limitations, challenges, and characteristics of simulation in general, as well as simulation studies with a focus on digital twins. The authors' goal was to develop an applicable framework for simulation that represents the design artifact of the DSR (Design Science Research) method and addresses structural guidelines for the use of simulation studies and digital twins. The digital twin shows the continuous integration of physical and virtual space before and during system operation with permanent two-way data exchange. A simulation study and a digital twin require a significant amount of data and the ability to analyze the data. A digital twin can only achieve its full functionality if historical, current, and sensor data are integrated.

In research (Meierhofer et al., 2020), the concept of creating additional service value using a digital twin is elaborated. The goal is to design and project services based on the needs of customers, using the potential of the digital twin. The digital twin is based on a combination of data, analytics, and visualization insights to support decision making. Implementing a digital twin requires the development of various elements of simulation models at different levels of granularity, which are then combined. The authors propose a procedure for developing a digital twin that ranges from modeling ecosystem services to integrating technical resources.

In the research paper (Bocciarelli, D'Ambrogio & Panetti, 2023), the authors present a model-driven method for automated development of digital twins of real business processes. The proposed method uses data from IoT sensors to automatically reconfigure the simulation model so that the digital twin is continuously coherent and aligned with its real-world counterpart. The paper proposes a framework to support development and analysis based on simulation of business processes informed by the operation of IoT devices. The scientific contributions of the research lie in: (1) an IoT (Internet of Things) metamodel that provides a flexible and coherent description of the entities of the IoT system, which consists of network sensors; (2) a methodology that illustrates how the framework uses the UML language and BPMN to effectively support the development and analysis of IoT-aware business processes; (3) an architecture that describes the proposed framework in the context of its components and the capabilities it provides.

Table 1. Business features associated with the DT

Key insights into the business features associated with the digital twin (DT)	**Authors and Publication year**
Business process intervention.	Dumas, 2021.
Manufacturing system.	Lugaresi & Matta, 2021.
Decision-making, use case.	West et al., 2021.
Supply chain.	Perez, Wassick, & Grossmann, 2022.
Production management systems.	Romero et. al., 2021.
Business Model Innovation (BMI), literature analysis, empirical research conducted in companies.	Dymitrowski & Mielcarek, 2021.

Business process modeling, process discovering, product development plan.	Mallek-Daclin et al., 2022
Digital transformation.	Sumereder & Woitsch, 2022; der Landwehr, Trott & von Viebahn, 2021.
Ecosystems.	Sumereder & Woitsch, 2022; Meierhofer et al., 2020.
Simulation.	Park & van der Aalst, 2021; Landwehr, Trott & von Viebahn, 2021; Meierhofer et al., 2020; Bocciarelli,D'Ambrogio & Panetti, 2023.

Table 1 highlights the key aspects of business features that previous research has connected with the digital twin paradigm. Most of the previous research connects the usage of the digital twin to manufacturing systems, simulation, digital transformation, and ecosystems. It was also noted that authors of previous research emphasized the importance of the digital twin for operational decision making. Some authors have conducted research using use cases, which is considered a good way to demonstrate a proposed framework or methodology. Overall, the contribution of the digital twin in business can be seen in the context of business interventions, timely responses, product delivery, supply chain, and production systems.

4 The architecture of process digital twin

This chapter proposes the architecture of the process digital twin. The proposal is based on the literature (Bocciarelli, D'Ambrogio & Panetti, 2023) and Figure 1 shows the architecture of the process digital twin.

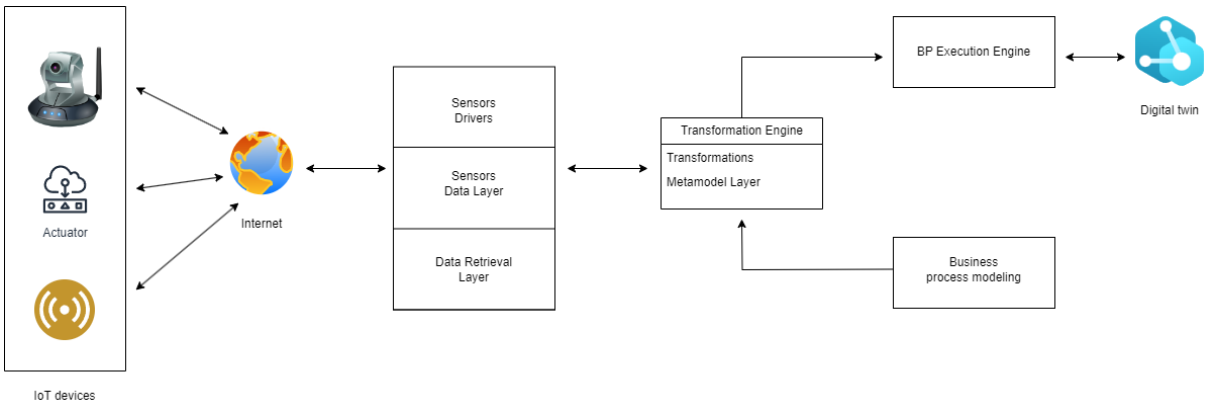


Figure 1. The architecture of process digital twin

The architecture of the process digital twin consists of IoT devices, Internet, Sensor Drivers, Sensors Data Layer, Data Retrieval Layer, Transformation Engine, Business process Execution Engine, Business process modeling, and digital twin model. Business process modeling provides a visual environment for specifying IoT-enabled business process and monitoring its execution. Business process Execution Engine executes the business process, and it is responsible for constructing the digital twin model and its maintenance. Transformation Engine implements the model-to-text transformations to produce an executable process description tailored to the adopted business process execution engine. It contains the transformation layer that provides model-to-text transformations, and the metamodel layer that provides an implementation of the IoT-BPMN metamodel. Sensors Drivers and Sensors Data Layer are the components "whose implementation is supported by the IoTBPMN-to-Sensor model-to-text transformation. Specifically, they provide the implementation of device-specific drivers and data models" (Bocciarelli, D'Ambrogio & Panetti, 2023). Data Retrieval Layer is responsible for updating the digital twin model with the data collected by the sensors that define the actual state of the real environment.

5 Business Process Modeling and Petri Nets

According to (Wang, Zhang & Shi, 2007), the goal of business process management is to improve and optimize existing business processes, and for this purpose, it works closely with other scientific

methodologies. Petri nets methodology is used to model elements in a business process in which several participants are involved. Petri nets were invented by Carl Adam Petri in 1962. They have been widely used in the field of systems analysis and design for decades. "A Petri net in its basic form is called a place/transition net, or PTN, and is a directed bi-partite graph with nodes consisting of places (drawn as ellipses) and transitions (drawn as rectangles). The state of a Petri net is called a marking and consists of a distribution of tokens (drawn as black dots) positioned on the places. The execution of a Petri net (also called the 'token game') consists of occurrences of enabled transitions removing tokens from input places and adding tokens to output places, as described by integer arc weights, thereby changing the current state (marking) of the Petri net" (Jensen & Kristensen, 2009).

Petri nets can be used to analyze the structural and dynamical properties of the system through rigorous mathematical analysis and visualize computer simulations (ElMansouri et al., 2009). The authors (Yun & Zhang, 2011) state that flexible and hierarchical business processes cannot be well modeled using an "ordinary" Petri net due to its limited modeling capabilities. Considering the limitations of "ordinary" Petri nets, colored Petri nets with a hierarchical structure are proposed for modeling business processes. Colored Petri nets are an event modeling language that combines the capabilities of Petri nets with a high-level programming language (Mukhlash et al., 2018; ElMansouri et al., 2009). In colored Petri nets, each place has a type, and each token has a value (color) corresponding to the place type. The color is used to distinguish tokens (Mukhlash et al., 2018). Tokens contain data associated with them (ElMansouri et al., 2009). Since a business process can be divided into different levels of subprocesses, the model can also be divided into corresponding levels of subnets called pages. Models of colored Petri nets can be structured into pages (Yun & Zhang, 2011).

Petri nets are widely used to simulate and evaluate the execution of different types of processes and are very effective (Cartelli, Di Modica & Tomarchio, 2015). In research (Ou-Yang & Lin, 2008), a framework is developed to analyze the feasibility of the BPMN business model under Petri net conditions. The goal is to bridge the gap between BPMN and Petri nets so that the developed business model can be transferred to the Petri net model. The result is that the constructed BPMN business model can be interpreted by business people and its feasibility can be simulated and analyzed by experts who understand Petri nets. The authors propose an approach to bridge the gap between the BPMN business model and Petri nets through the use of XML. XML is a recognized standard for integrating various technologies in a Web-based environment (Ou-Yang & Lin, 2008).

In (Mukhlash et al., 2018), the business process of production systems was analyzed using colored Petri nets. First, the researchers collected data from the event log and created a model of Petri nets. Then, they analyzed the model to check compliance, availability, and bottlenecks. After that, they converted the model into colored Petri nets. In the end, they made a conclusion about the business process based on the results of the analysis and simulation.

6 Limitations of the existing research

In their research, the authors (Bocciarelli, D'Ambrogio & Panetti, 2023) state that their methodology makes no assumptions about the availability of IoT BPMN modeling tools. There is no information in the research about data collection by network sensors. In addition, most of the previous research does not address the feedback that the digital twin sends back to the reactor to change the performance of the activity when unexpected events occur. Therefore, the question is how the digital twin communicates with external devices and whether it even sends feedback message about the performance of business processes to make a decision in real-time.

The study (Perez, Wassick & Grossmann, 2022) states that digital twin models were generated using historical data from databases and updated in real-time using live process data. The study does not explain exactly how the digital twin models are updated using "live process data", and questions arise about the data content, where it comes from, and how it is communicated to the digital twin. Also, if the digital twin model is generated based on historical data, is that data still "real-time" at the time of go-live?

In studies (Lugaresi & Matta, 2021; Park & van der Aalst, 2021), the concept of the digital twin is equated with simulation, which is not quite correct, because a digital twin is not a simulation, since its implementation requires more advanced conditions and its creation is more complex. The simulated business process model does not require real-time data, and the business process simulation is not updated in real-time but is executed in order to observe the past behavior of the business process without indicating possible changes. A digital twin uses data in real-time, updates its performance in real-time using data it reads from external devices and sensors, and is a more advanced form of digital twin when machine learning is incorporated into its implementation. The idea of machine learning is that the digital twin model is updated according to the amount of data collected during its execution and forms a dataset on which artificial intelligence (AI) can be trained, enabling it to communicate with the digital twin in terms of adjusting execution based on its behavioral patterns.

Upon comprehensive evaluation of the existing literature, it became apparent that the research conducted thus far does not sufficiently encompass the structural transition to the digital twin.

7 Conclusion

A digital twin is a tool that supports decision making by translating possible options into a business context and helping to identify the consequences of different decisions. A process digital twin is a virtual representation or replica of a physical industrial process or system. It combines real-time data, advanced analytics, and simulation models to mimic and simulate the behavior of the actual process. Process digital twin captures and models the process's behavior, dynamics, and characteristics. The concept of process digital twins is revolutionizing the way industrial processes are monitored, controlled, and optimized. Organizations can gain valuable insights, improve efficiency, and make informed decisions to drive operational excellence by creating virtual replicas that closely mirror physical processes.

It is a matter of when to create a process digital twin because limitations can arise if a process digital twin is created too early in terms of insufficient or incomplete data. It is important to enrich the business process with a digital twin and execute it in real time. This allows for more realistic modeling of reality, and the non-deterministic behavior of the system can be made more adaptive. This paper presents the results of previous research in the field of business processes, digital twins, and colored Petri nets. A process digital twin architecture is also proposed, and the limitations of previous research are identified.

Acknowledgment

This work is supported by the Croatian Science Foundation under the project IP-2019-04-4864.

References

Bocciarelli, P., D'Ambrogio, A., Panetti, T. A Model Based Framework for IoT-Aware Business Process Management. In Future Internet 15(2), 50, 2023.

Cartelli, V., Di Modica, G., Tomarchio, O. A cost-centric model for context-aware simulations of business processes. In IC3K 2015 - Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management 3, pp. 303-314, 2015.

ElMansouri, R., Chaoui, A., Kerkouche, E., Khalfaoui, K. From UML statecharts and collaboration diagrams to coloured Petri Nets models: A graph transformation based approach for modeling and analysis of business processes in Virtual Enterprises. In 4th South-East European Workshop on Formal Methods: Formal Methods for Web Services; Formal Methods for Agent-Based Systems, SEEFM 2009 5465144, pp. 73-78, 2009.

der Landwehr, M.A., Trott, M., von Viebahn, C. Waste of time and money? Constructing an applicability framework for organizational use of simulation studies and digital twins. In International Conference on Information Systems, ICIS 2020 - Making Digital Inclusive: Blending the Local and the Global, 2021.

Dumas, M. Constructing digital twins for accurate and reliable what-if business process analysis. In CEUR Workshop Proceedings 2938, pp. 23-27, 2021.

Dymitrowski, A., Mielcarek, P. Business model innovation based on new technologies and its influence on a company's competitive advantage. In Journal of Theoretical and Applied Electronic Commerce Research 16(6), pp. 2110-2128, 2021

Jensen, K. & Kristensen, L.M. (2015). Colored Petri Nets: A Graphical Language for Formal Modeling and Validation of Concurrent Systems. In COMMUNICATIONS OF THE ACM, 58 (6), pp.61-70.

Lugaresi, G., Matta, A. Automated digital twins generation for manufacturing systems: A case study. In IFAC-PapersOnLine 54(1), pp. 749-754, 2021.

Mallek-Daclin, S., Chaabaoui, L., Daclin, N., Rabah, S., Zacharewicz, G. Planification model-based process discovering. In 21st International Conference on Modeling and Applied Simulation, MAS 2022, 2022

Meierhofer, J., West, S., Rapaccini, M., Barbieri, C. The Digital Twin as a Service Enabler: From the Service Ecosystem to the Simulation Model. In Lecture Notes in Business Information Processing 377 LNBIP, pp. 347-359, 2020.

Mukhlash, I., Rumana, W.N., Adzkiya, D., Sarno, R. Business process improvement of production systems using coloured Petri nets. In Bulletin of Electrical Engineering and Informatics 7(1), pp. 102-112, 2018.

Ou-Yang, C., Lin, Y.D. BPMN-based business process model feasibility analysis: A Petri net approach. In International Journal of Production Research 46(14), pp. 3763-3781, 2008.

Park, G., van der Aalst, W.M.P. Towards Reliable Business Process Simulation: A Framework to Integrate ERP Systems. In Lecture Notes in Business Information Processing 421, pp. 112-127, 2021.

Perez, H.D., Wassick, J.M., Grossmann, I.E. A digital twin framework for online optimization of supply chain business processes. In Computers and Chemical Engineering 166,107972, 2022.

Romero, D., Von Cieminski, G., Wuest, T., (...), Alfnes, E., Rudberg, M. Advances in Production Management Systems: Issues, Trends, and Vision Towards 2030. In IFIP Advances in Information and Communication Technology 600, pp. 194-221, 2021.

Sumereder, A., Woitsch, R. DIGITIZATION PRINCIPLES FOR APPLICATION SCENARIOS TOWARDS DIGITAL TWINS OF ORGANIZATIONS. In World Congress in Computational Mechanics and ECCOMAS Congress, 2022.

TechTarget: What is a digital twin and how does it work? (2023). Retrieved from www.techtarget.com/searcherp/definition/digital-twin.

Wang, X., Zhang, Y., Shi, H. Scenario-based Petri net approach for collaborative business process modeling. In Proceedings of the 2nd IEEE Asia-Pacific Services Computing Conference, APSCC 2007 4414436, pp. 18-25, 2007.

West, S., Stoll, O., Meierhofer, J., Züst, S. Digital twin providing new opportunities for value co-creation through supporting decision-making. In Applied Sciences (Switzerland) 11(9), 3750, 2021.

Yun, G.X., Zhang, J. Process-oriented software development based on HPN business process model. In Advanced Materials Research 201-203, pp. 763-767, 2011.

Analysis of Organizational Culture Using Cameron and Quinn's Typology: A Case Study of an ICT Organization

Nikolina Posarić, Damjan Ćorić
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42 000 Varaždin
niposari@foi.unizg.hr, dcoric@student.foi.hr

Abstract. *In the last few decades, the creation and development of organizational culture has been an essential factor in the success of contemporary organizations. In order for employees to be satisfied and productive in their workplaces, management must account for various aspects of the development of organizational culture. The purpose of this paper is to analyze the role and importance of organizational culture for contemporary organizations. The concept and importance of organizational culture, as well as Cameron and Quinn's typology as one of the most famous typologies of organizational culture, are explained first. Cameron and Quinn's typology emphasizes four different types of organizational culture: clan culture, hierarchy culture, market culture, and adhocracy culture. At the core of the paper is a case study of a selected ICT organization. Using Cameron and Quinn's typology and their Organizational Culture Assessment Instrument (OCAI), the study examines which type of organizational culture is currently present and which type is desirable in the selected ICT organization. The results of the research conducted among employees of an ICT organization indicate that clan culture is currently dominant, but an adhocracy culture is desirable. These results are consistent with the theoretical assumptions of Cameron and Quinn's typology of organizational culture.*

Keywords. organizational culture, Cameron and Quinn's typology, OCAI, case study, ICT organization

1 Introduction

All organizations have an organizational culture, and employees are exposed to the influences of that culture, while the values and fundamental assumptions contained in the organizational culture shape employee behavior (Načinović Braje, 2016). Ćorić (2019) explains that the concept of organizational culture has been gaining greater importance in studies of organizational behavior. Depending on the leadership style within the organization, management generally sets the "rules of the game," *i.e.*, the values, norms, and rules that the organizational culture is recognized with. The organization and the employees should benefit if these rules and values are respected and accepted by all, or at least by most, employees. Additionally, managers must give clear instructions to employees regarding the desirable behavior within the organization while respecting the same instructions themselves. In sum, the organization's success depends on accepting and nurturing its organizational culture, the importance of which is therefore justified.

According to Žugaj *et al*. (2004, p. 14), the importance of organizational culture comes to the fore in the organization's everyday life; it "affects the design and implementation of strategy, the application of authority, the type of structure and the way of organizing." Sikavica (2011) claims that in some organizations with a strong organizational culture, the organizational culture is also the main component of the organization's effectiveness, which is why many successful organizations take care to develop and maintain an effective organizational culture. Moreover, organizational culture creates, with recognizable symbols, a positive image of the organization in the public, delivering another benefit to the organization's success.

This paper first explains the concept of organizational culture through a literature review, shows contemporary trends, and presents Cameron and Quinn's typology of organizational culture. After the theory, we present the results of an empirical study on the organizational culture of an ICT organization. The most significant conclusions and implications for future research are presented at the end of the paper.

2 Literature review

2.1 Determining the Concept of Organizational Culture

Numerous definitions of organizational culture are mentioned in the literature, and many authors have

contributed to defining the concept of organizational culture. Bellot (2011) claims that Pettigrew was the first author to explain the term organizational culture formally. According to Pettigrew (1979), organizational culture is the system of collectively and publicly accepted meanings that operate for a given group at a given time. Smircich (1983) states that several conceptions of organizational culture underlie the research of these five content areas: comparative management, corporate culture, organizational cognition, organizational symbolism, and structural and psychodynamic views. Smircich (1983, p. 343-351) introduces several definitions of organizational culture according to different perspectives:

- Comparative management perspective: "patterns of attitudes and actions of individual organization members"
- Corporate culture perspective: "social or normative glue that holds an organization together. It expresses the values or social ideals and the beliefs that organization members come to share. These values or patterns of belief are manifested by symbolic devices such as myths, rituals, stories, legends, and specialized language"
- Organizational cognition perspective: "a system of shared cognitions or a system of knowledge and beliefs"
- Organizational symbolism perspective: "systems of shared symbols and meanings … postulates or understandings, declared or implicit, tacitly approved or openly prompted, that orient and stimulate social activity"
- Structural and psychodynamic perspective: "expression of unconscious psychological processes."

One of the most cited authors, Schein (1992, p. 12), defined the culture of a group as "a pattern of shared basic assumptions that the group learned as it solved its problems of external adoption and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems." Another definition by Schein (1996) claims that organizational culture is the norms, values, and assumptions about how organizations function, which are shared by the members of the organizations. Schein (1999) also claims that organizational culture is the learned, shared, and tacit assumptions on which employees base their daily behavior, *i.e.*, organizational culture is the way we do things here.

According to Schneider, Ehrhart, and Macey (2013), organizational culture refers to the implicit values, beliefs, and assumptions that serve as a guide to employees' behavior and is based on the stories, myths, and socialization experiences they have and the behaviors they observe, especially from leaders. Kumar (2016) explains that organizational culture can be viewed as an important concept in social and organizational psychology. Organizational culture reflects the beliefs, norms, and values that characterize an organization. In other words, organizational culture reflects what is typical and general for the organization (Kumar, 2016). Načinović Braje (2016) emphasizes that in organizational theory, culture describes practices and patterns of behavior, shared values, and beliefs, or simply a way of life characteristic of a particular organization. Therefore, we can conclude that organizational culture is a set of values, norms, assumptions, and beliefs that are shared within the organization and that keep the members of the organization together.

2.2 Cameron and Quinn's Typology of Organizational Culture

Numerous organizational culture typologies are described in contemporary literature, explaining the different types and characteristics of organizational cultures. According to Ćorić (2019), Harrison was among the first authors to create a typology of organizational culture, and he proposed four types: power culture, role culture, task culture, and person culture. As noted by Žugaj *et al*. (2004), six years later, Handy took over and elaborated on Harrison's ideas, connecting each type of organizational culture with its corresponding type of organizational structure, thus creating an entrepreneurial culture (power culture), role culture (bureaucratic culture), task culture (team culture), and individual culture. Among the more well-known typologies of organizational culture mentioned in the literature include Deal and Kennedy's typology, Edwards and Kleiner's typology, Scholz's typology, Schein's typology, and Cameron and Quinn's typology.

Cameron and Quinn (2006) considered four main types of organizational culture based on the Competing Values Framework (CVF): clan culture, hierarchy culture, market culture, and adhocracy culture. According to Cameron and Quinn's typology, the four types of organizational culture are based on two basic dimensions: one speaks to the organization's stability or flexibility, while the second deals with its focus, *i.e.*, is it focused on its internal processes or to processes in its environment (Šmider & Petr Balog, 2012). Figure 1. shows the four types of organizational culture according to Cameron and Quinn and their important characteristics.

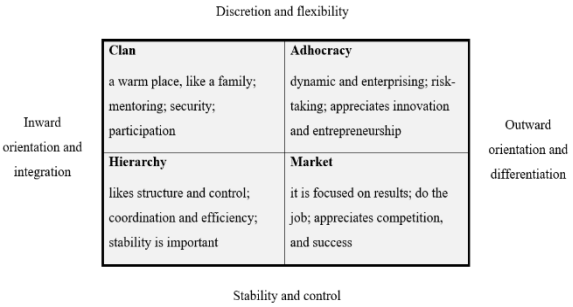


Figure 1. Four types of organizational culture according to Cameron and Quinn
(Šmider & Petr Balog, 2012, p. 76)

2.3 Contemporary Trends in the Study of Organizational Culture

Considering the important role given to organizational culture in modern business conditions, the existing literature mentions six primary characteristics that seem to capture the essence of organizational culture (Chatman *et al*., 2014; Robbins & Judge, 2019):

- **Adaptability***:* the degree to which employees are encouraged to be innovative and flexible, to take risks and experiment
- **Detail orientation***:* the degree to which employees are expected to exhibit analysis, precision, and attention to detail
- **Results/outcome orientation***:* the degree to which management focuses on results/outcomes
- **People/customer orientation***:* the degree to which management decisions consider the effect of outcomes on people within and outside the organization (i.e., customer)
- **Collaboration/team orientation***:* the degree to which work activities are organized around teams
- **Integrity***:* the degree to which people exhibit integrity and high ethical standards in work.

According to these characteristics, contemporary trends in the study of organizational culture imply three key features influencing modern organizations: creating an innovative culture, creating a customer-responsive culture, and creating a sustainability culture (Robbins & Coulter, 2018).

An innovative culture has been the focus of the academic community in recent years due to the turbulent and increasingly competitive conditions in which modern organizations operate. Robbins and Judge (2019) claim that start-up enterprises often have innovative cultures because they are focused on solving different problems to survive and grow. The results of previously conducted research in different types of organizations show that innovative organizational culture is positively associated with not only organizational performance (Karczewska, 2021; Park, Lee, & Kim, 2016; Ryu *et al*., 2021; Sofi & Devanadhen, 2015; Wei *et al*., 2013) but also employee behavior and attitude (Madueke & Emerole, 2017; Umemezia & Osifo, 2021).

In the conditions of a highly competitive market, customers are at the center of the activities of modern organizations. Madhani (2018) emphasizes that customer-focused culture exhibits attitudes concerned with developing and enhancing customer value. In other words, in customer-focused culture, employees are committed to creating superior customer value (Madhani, 2021a). Customer-focused culture is based on the idea that what is best for the customer is also best for the organization (Madhani, 2021b). Madhani (2021b) also goes on to claim that customer-focused culture leads to better customer satisfaction, and satisfied customers are more likely to keep a stable relationship with the organization. All this leads to the conclusion that the customer-focused culture seen in modern organizations greatly affects overall organizational performance and brings competitive advantages.

Robbins and Coulter (2018) believe that an organization can achieve its business goals and increase long-term value for shareholders by integrating economic, environmental, and social opportunities into its business strategies. Therefore, sustainability is embedded in the overall organizational culture of many organizations. Further, Sen (2020) emphasizes that in a sustainability culture, all aspects of an organization reflect its sustainability orientation. In other words, sustainability is integrated with the organization's mission, vision, and core values. The benefits of creating a culture of sustainability include increased levels of employee performance and commitment, as well as improved overall organizational performance (Galpin, Whittington, & Bell, 2015).

Previous research on organizational culture based on Cameron and Quinn's typology suggests that different organizations have different types of culture. The results of research conducted in Poland show that in public universities the dominant type of culture is hierarchy culture, while in non-public universities the dominant type of culture is market culture (Dębski, Cieciora, Pietrzak, & Bołkunow, 2020). Also, the results of research conducted at Riga Technical University in 2014 indicated that the current organizational culture of the University is more like a hierarchy culture, and the desirable organizational culture is a mix of clan culture and adhocracy culture (Lapiņa, Kairiša, & Aramina, 2015). The research conducted in the 5-Star Hotel's in Istanbul found that the mean values for each type of organizational culture are very close to each other and high, but the highest is for market culture (Oz, Kaya, & Ciftci, 2015). The findings of the research conducted in the Turkish construction industry show that the industry has been dominated by companies with a mixture of clan and hierarchy cultures (Oney-Yazıcı, Giritli, Topcu-Oraz, & Acar, 2007). The results of research conducted in Turkish pharmaceutical companies indicate that the perceived type of culture is a hierarchy, while the preferred type is clan culture (Demir, Unnu, & Erturk, 2011).

By searching the literature, it was determined that research on organizational culture according to Cameron and Quinn's typology in ICT organizations is not so prevalent. Cameron and Quinn (2006) claim that in highly technologically oriented companies, the organizational culture most often changes from clan and adhocracy to hierarchy and market culture, as companies grow, and they explain this important fact using the example of Apple. Bendak, Shikhli, and Abdel-Raze (2020) have determined that in one medium size IT company in the United Arab Emirates, the hierarchy type of organizational culture is currently the most prevalent, while the desired type is a mixture of the clan and hierarchy culture. Since it has been established that there needs to be more research on organizational culture concerning this typology, this

paper will contribute to the literature on organizational culture.

3 Empirical Research of Organizational Culture: A Case Study of an ICT Organization

For the purpose of analyzing the organizational culture using the case study method, a survey was conducted among the employees of an ICT organization from Bosnia and Herzegovina with headquarters in Mostar. At the time of the research, this ICT organization had 39 employees. The selected ICT organization was founded in 2012 and develops innovative software solutions and mobile interfaces for both public institutions and small and medium-sized enterprises.

3.1 Research Problem and Goals

The main objective of this research is to examine and analyze organizational culture in a selected ICT organization using Cameron and Quinn's typology of organizational culture using the case study method of analysis. A secondary goal of the research is to find out which type of organizational culture prevails in the selected ICT organization and which type of organizational culture is desirable in the future. In accordance with the set objectives, three research questions were formed:

RQ1. What type of organizational culture currently prevails in the selected ICT organization?

RQ2. What is the desired type of organizational culture in the selected ICT organization in the future?

3.2 Research Instrument, Sampling, and Data Collection

A survey questionnaire was used as a research instrument, and it is one of the most well-known organizational culture assessment instruments developed by Cameron and Quinn (2006). Named the Organizational Culture Assessment Instrument (OCAI), it is used to assess both the current and desired organizational culture. To conduct the research, an original survey questionnaire from Cameron and Quinn (2006) was translated into the Croatian language. It consists of six basic categories: (1) dominant characteristics, (2) organizational leadership, (3) management of employees, (4) strategic emphases, (5) organizational glue, and (6) criteria of success. Each category consists of four statements, each of which represents a characteristic of an organizational culture. Category A represents clan culture, category B adhocracy culture, category C market culture, and category D hierarchy culture.

According to Šmider and Petr Balog (2012), the interesting thing about the OCAI instrument is the rating scale used, which is the ipsative scale. Most researchers are familiar with the Likert scale (which belongs to the category of normative scales), where respondents choose one number from 1 to 5 or 1 to 7, representing "I do not agree at all" to "I completely agree." In contrast to the Likert scale, the ipsative scale offers respondents a total of 100 points to score answers (Šmider & Petr Balog, 2012).

The questionnaire consisted of two parts[1]. The first part of the questionnaire contained three questions related to each respondent's gender, years of experience, and the position they are currently working in, which best suits the workplace in the organization. Furthermore, respondents assessed the current and desired organizational culture in the second part of the questionnaire. Each answer was filled in by scoring each statement in the empty square. Respondents were asked to allocate 100 points to several offered answers in such a way as to give the highest number of points to the answer they agreed with the most. They were instructed to distribute the points so that, in the end, the sum of the points of all the answers added up to 100.

From the selected ICT organization in Bosnia and Herzegovina, a total of 11 out of 39 (28.20 percent) employees answered. To get a complete picture of the current and desirable type of organizational culture, all employees from the selected ICT organization should have participated in the research, which is not the case here. However, we assess that this number of respondents is sufficient for our analysis according to the case study method used in this paper. Considering this, our small sample of respondents is valid for analysis.

The research was conducted from July 7 to August 5, 2022. The results were processed in Microsoft Excel. The sample consisted of four female (36.36 percent), and seven male (63.64 percent) employees. Department heads made up 18.18 percent of the sample, and other employees made up 81.82 percent. Employees with up to two years of service, as well as those with more than six years in the organization, accounted for 18.18 percent each, while employees with between three and five years of service in the organization accounted for 63.64 percent of the sample.

The next step was analyzing and processing all survey questionnaires, which required summing up all statements for category A in column "Current organizational culture" and dividing the obtained sum by the number 6 (number of six basic categories of organizational culture) to get the mean value. After that, statements for categories B, C, and D were added up similarly, and each sum was also divided by the number 6 to get the mean value. The same procedure was applied

[1] Due to the size of the questionnaire, it is not attached at the end of the paper, but a questionnaire in Croatian and English can be obtained upon request from the authors.

to the column "Desired organizational culture". Further, when the analysis was done according to the six basic categories of organizational culture, the mean value of the scores of all respondents was calculated for the statements for category A, similarly for the statements for categories B, C, and D, for each of six basic categories, and current and desired organizational culture.

3.3 Data Analyses and Results

The results of the research shown in Figure 2 indicate that the currently dominant organizational culture in the selected ICT organization is clan culture (the mean value of the answers is 37.05). It is followed by hierarchy (the mean answer value was 24.99), adhocracy (the mean answer value was 22.95), and finally, market culture (the mean answer value was 15.01). It can also be determined that there are significant gaps between the current and desired state, *i.e.*, the most desirable type of organizational culture is adhocracy (the mean answer value was 30.68), followed by the hierarchy (the mean answer value was 25.08), and in third place is clan (the mean answer value was 24.25). The respondents declared the market culture as the least desirable type of organizational culture (the mean answer value was 19.99).

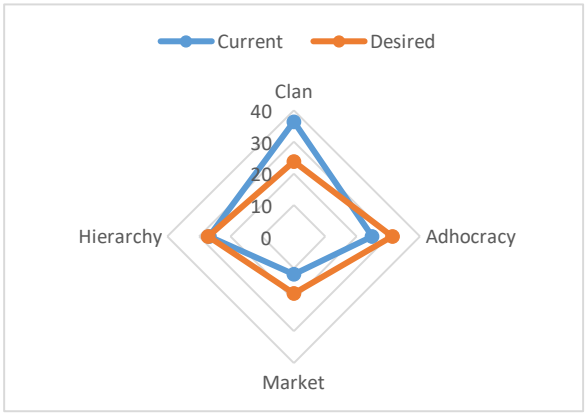


Figure 2. Perception of the current and desired organizational culture of the selected ICT organization

Table 1 presents the results of the current and desired type of culture for each of the six characteristics of organizational culture according to Cameron and Quinn (2006). It can be seen that in all characteristics, clan culture is the current type of organizational culture. In dominant characteristics, the desired organizational culture is a mixture of clan and adhocracy cultures. In organizational leadership, clan culture is also the desired type of organizational culture. In management of employees, strategic emphases, and criteria of success, the desired type of organizational culture is adhocracy culture. In organizational glue, the desired type of organizational culture is the hierarchy type.

Table 1. Perception of the current and desired organizational culture of the selected ICT organization according to the six characteristics of organizational culture

Characteristics of organizational culture	Current	Desired
Dominant characteristics	Clan (mean: 35.45)	Mixture of clan and adhocracy (mean: 29.09 for both)
Organizational leadership	Clan (mean: 40.91)	Clan (mean: 29.09)
Management of employees	Clan (mean: 35.45)	Adhocracy (mean: 34.09)
Strategic emphases	Clan (mean: 34.55)	Adhocracy (mean: 33.64)
Organizational glue	Clan (mean: 31.36)	Hierarchy (mean: 36.36)
Criteria of success	Clan (mean: 44.55)	Adhocracy (mean: 29.09)

4 Discussion, Limitation, and Implication for Future Research

As seen in the previous part of the paper, the research results in the selected ICT organization indicated that according to Cameron and Quinn's typology, the current type of organizational culture is clan culture, while the desired type is adhocracy culture. Since the selected ICT organization has been operating on the market for ten years and has 39 employees, it is understandable that the current type of culture is a clan culture characterized by the organization as a family, a warm workplace, and employee participation. An important fact is that clan culture is recognized as a dominant type of culture through all six characteristics of organizational culture.

Given that the selected ICT organization operates in an industry characterized by dynamism, constant changes, and adaptation to new market requirements, it should have the type of organizational culture that will be able to cope with these challenges. The adhocracy type is just such a type of organizational culture, and the research results show that it is a desirable type of organizational culture in the selected ICT organization in the future. If the six characteristics of the organizational culture are observed, it was determined that the employees in the selected ICT organization, regarding these six characteristics, tend to some other type of organizational culture, but that it is not necessarily an adhocracy culture. This may be because employees still want their organization to be like a family, for leaders to continue to be mentors and parent

figures, and for formal rules, hierarchy, and stability to be set in the organization.

These findings are valuable as a contribution to the literature on organizational culture based on Cameron and Quinn's typology, which has not been researched to such an extent in ICT organizations. Also, it can be noted that according to the previous results in research conducted by Bendak, Shikhli, and Abdel-Raze (2020), the results for the selected ICT organization are different, considering that it is about different sizes of ICT organizations. Research of this type is desirable in all organizations, both for organizational scientists and management. In this way, organizational scientists discover what kind of organizational culture prevails in organizations and conclude the importance of developing organizational culture. On the other hand, this research can help management to see how employees perceive the organizational culture and what steps they need to take to create a better culture.

We are aware that the 11 employees from the selected ICT organization who participated in our mini-research do not represent a sample that is large enough, and this is one of the limitations of our research. To obtain more precise results, all employees of the same organization should participate in research. However, this sample of respondents is large enough to get a picture of the organizational culture of an organization, especially if it is a small organization and if the case study method is used to analyze the organizational culture.

Future research should focus on trends that currently exist in business practice in the form of creating an innovative culture, a customer-responsive culture, and a sustainability culture, but also on the relationship of organizational culture with a competitive advantage. Also, in future research, an analysis of organizational culture using Cameron and Quinn's typology could be made for the whole sector of ICT companies in Croatia or neighboring countries, which would be an important contribution to the literature.

5 Conclusion

Organizational culture differs from organization to organization in many small ways. It is apparent that organizational culture must be nurtured and accepted by most, if not all, employees, with clear support from the organization's management, that is, the people who establish the organizational culture. The more accepted the organizational culture is by all employees, the more successful the organization will be. However, organizational culture depends on the activity that the organization is engaged in. The assumption is that values, components, and indicators are cultivated in the different activities in which organizations operate.

There are different typologies of organizational culture, depending on the authors. One of the typologies is that of Cameron and Quinn (2006), who divided the organizational culture into four types: clan, adhocracy, hierarchy, and market. Each of the mentioned cultures has its characteristics. The clan culture is characterized by an organization that is warm, like a family, providing security, mentoring, and participation. An adhocracy culture is characterized by dynamism and enterprise, where risk and innovation are especially valued. The hierarchy culture prefers structure and control and considers stability and efficiency. Market culture is focused on results, values, competition, success, and getting the job done. The same authors also created an instrument for measuring the organization of culture, OCAI, used in this paper. The research conducted in the selected ICT organization suggests that the clan type of organizational culture is currently dominant but that the adhocracy type of organizational culture is the most desirable. Given the sector of activity in which the selected ICT organization operates, these results are expected but also consistent with the theoretical assumptions of Cameron and Quinn's typology of organizational culture.

Organizations are more and more emphasizing the importance of organizational culture. Over the past few decades, organizational culture has become one of the most important factors on which the work performance of employees and the entire organization depends. Many modern organizations are successful thanks to creating, developing, and maintaining a strong organizational culture that manifests through various indicators. For this reason, organizations operating in the twenty-first century's new and turbulent business conditions should continue establishing a strong organizational culture that will enable them to achieve competitiveness and excellent business results.

References

Bellot, J. (2011). Defining and Assessing Organizational Culture. Nursing forum. 46(1), 29–37. doi: 10.1111/j.1744-6198.2010.00207.x

Cameron, K. S., & Quinn, R. E. (2006). Diagnosing and Changing Organizational Culture: Based on the Competing Values Framework. Revised Edition. San Francisco: Jossey-Bass.

Chatman, J. A., Caldwell, D. F., O'Reilly, C. A., & Doerr, B. (2014). Parsing Organizational Culture: How the Norm for Adaptability Influences the Relationship Between Culture Consensus and Financial Performance in High-technology Firms. Journal of Organizational Behavior, 35(6), 785–808. doi: 10.1002/job.1928

Ćorić, N. (2019). Korporativna komunikacija, Zagreb–Sarajevo–Mostar: Synopsis.

Dębski, M., Cieciora, M., Pietrzak, P., & Bołkunow, W. (2020). Organizational culture in public and non-public higher education institutions in Poland:

A study based on Cameron and Quinn's model. Human Systems Managemen, 39(3), 345–355. doi: 10.3233/HSM-190831

Demir, C., Ayyildiz Unnu, N. A., & Erturk, E. (2011). Diagnosing the organizational culture of a Turkish pharmaceutical company based on the competing values framework. Journal of business economics and management, 12(1), 197–217. doi: 10.3846/16111699.2011.555451

Galpin, T., Whitttington, J. L., & Bell, G. (2015). Is your sustainability strategy sustainable? Creating a culture of sustainability. Corporate Governance, 15(1), 1–17. doi: 10.1108/CG-01-2013-0004

Karczewska, A. A. (2021). Relationship between pro-innovative organizational culture and development and profitability of companies. Scientific Quarterly "Organization and Management", 2(54), 43–57. doi: 10.29119/1899-6116.2020.54.3

Kumar, A. (2016). Redefined and Importance of Organizational Culture. Global Journal of Management and Business Research, 16(4), 14–18.

Lapiņa, I., Kairiša, I., & Aramina, D. (2015). Role of Organizational Culture in the Quality Management of University. Procedia-Social and Behavioral Sciences, 213, 770–774. doi: 10.1016/j.sbspro.2015.11.472

Madhani, P. M. (2018). Enhancing Customer Value Creation with Market Culture: Developing 7Cs Framework. IUP Journal of Management Research, 17(3).

Madhani, P. M. (2021a). Building a Customer-focused Culture with Rewards Alignment: A Cross-sectional Analysis. International Journal of Electronic Customer Relationship Management, 13(1), 45–80. doi: 10.1504/IJECRM.2021.115609

Madhani, P. M. (2021b). Building Organization Culture with Seven Cs Framework: Enhancing Business Performance. HRM Review, 17(2), 23–30.

Madueke, C. V., & Emerole, I. C. (2017). Organizational Culture and Employee Retention of Selected Commercial Banks in Anambra State. Saudi Journal of Business and Management Studies, 2(3), 244–252. doi: 10.21276/sjbms.2017.2.3.16

Načinović Braje, I. (2016). Organizacijska kultura. In L. Galetić (Ed.) Organizacija (pp. 371–398). Zagreb: Sinergija.

Oney-Yazıcı, E., Giritli, H., Topcu-Oraz, G., & Acar, E. (2007). Organizational culture: The case of Turkish construction industry. Engineering, Construction and Architectural Management, 14(6), 519–531. doi: 10.1108/09699980710828996

Oz, M., Kaya, F., & Ciftci, I. (2015). Evaluating the Organizational Culture Types of the 5-Star Hotel's in Istanbul in terms of the Cameron & Quinn Competing Values Model. Yaşar Üniversitesi E-Dergisi, 10 (40), 6684–6691.

Park, J., Lee, K.-H., & Kim, P. S. (2016). Participative Management and Perceived Organizational Performance: The Moderating Effects of Innovative Organizational Culture. Public Performance & Management Review, 39(2), 316–336. doi: 10.1080/15309576.2015.1108773

Pettigrew, A. M. (1979). On Studying Organizational Cultures. Administrative science quarterly, 24(4), 570–581. doi: 10.2307/2392363

Robbins, S. P., & Coulter M. (2018). Management (14th Edi.). New York: Pearson.

Robbins, S. P., & Judge, T. A. (2019). Organizational Behavior (18th Edi.). New York: Pearson.

Ryu, J.-m., Park, S., Park, Y.; Park, J.; Lee, M. (2021). Innovative Culture and Firm Performance of Medical Device Companies: Mediating Effects of Investment in Education and Training. International Journal of Environmental Research and Public Health, 18(8926). doi: 10.3390/ijerph18178926

Schein, E. H. (1992). Organizational Culture and Leadership. San Francisco: Jossey-Bass Wiley.

Schein, E. H. (1996). Culture: The Missing Concept in Organization Studies. Administrative science quarterly, 41(2), 229-240.

Schein, E. H. (1999). The Corporate Culture Survival Quide: Sense and Nonsense About Culture Changes. San Francisco: Jossey-Bass Wiley.

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. Annual review of psychology, 64, 361–388. doi: 10.1146/annurev-psych-113011-143809

Sen, Y. (2021). Essentials of Sustainability: A Roadmap for Businesses. In C. F. Macado & P. Davim (Eds.) Sustainable Management for Managers and Engineers (pp. 67–88). doi: 10.1002/9781119804345.ch4

Sikavica, P. (2011). Organizacija. Zagreb: Školska knjiga.

Šmider, M., & Petr Balog, K. (2012). Kakvu vrstu organizacijske kulture imaju naše knjižnice? Primjer Gradske i sveučilišne knjižnice u Osijeku. Knjižničarstvo: glasnik Društva knjižničara Slavonije, Baranje i Srijema, 15(1-2), 71–96.

Smircich, L. (1983). Concepts of Culture and Organizational Analysis. Administrative science quarterly, 28(3) 339–358. doi: 10.2307/2392246

Sofi, M. A., & Devanadhen, K. (2015). Impact of Organizational Culture on Organizational Performance: An Empirical Assessment of Banking Sector in Jammu and Kashmir India. International Journal of Innovative Research and Development, 4(9), 34–44.

Umemezia, E., & Osifo, S. J. (2021). Organisational Culture and Employee Engagement in the Banking Industry in Benin City. Ife Social Sciences Review, 29(1), 1–13.

Wei, Y., O'Neill, H., Lee, R. P., & Zhou, N. (2013). The impact of innovative culture on individual employees: The moderating role of market information sharing. Journal of Product Innovation Management, 30(5), 1027–1041. doi: 10.1111/j.1540-5885.2012.01000.x

Žugaj, M., Bojanić-Glavica, B., Brčić, R., & Šehanović, J. (2004). Organizacijska kultura. Varaždin: TIVA Tiskara.

E-Schools project - Framework for Digital Competencies of School Principals: Competencies in the Area Digital Infrastructure

Nikolina Žajdela Hrustek, Renata Mekovec, Valentina Kirinić
University of Zagreb
Faculty of Organization and Informatics,
Pavlinska 2, Varaždin, Croatia
{nikolina.zajdela, renata.mekovec, valentina.kirinic}@foi.unizg.hr

Abstract. *The efficient integration of digital technologies into all facets of school operation is the responsibility of principals of schools undergoing digital transformation in the 21st century. As part of the project "e-Schools: Development of the System of Digitally Mature Schools", which has completed its second phase of implementation in the Republic of Croatia, the Framework for Digital Competencies of School Principals was created.*
The necessity of educational institution (school) principals maintaining suitable digital infrastructure is emphasized in this study, along with some related challenges. The digital skills that the principal must have in order to properly manage the school's digital infrastructure were highlighted in particular. The key steps of the methodology for developing the framework were briefly described in the paper. Particular attention was paid to the presentation of the digital competencies developed within the field of digital infrastructure (a total of five competencies), as well as typical activities (a total of 24 typical activities).

Keywords. school, digital infrastructure, principals, education, digital competence

1 Introduction

The main goal of the research was to define the key digital competencies of principals related to the area of Digital Infrastructure by (1) exploring and reviewing the relevant literature and (2) working with experts in the field of digital competencies and the main stakeholders (principals).

In accordance with the set goal of the research, the authors defined the following research questions:

- What digital competencies should principals acquire in order to successfully manage the area of Digital Infrastructure?
- What typical activities should school principals perform in order to successfully manage the area of Digital Infrastructure?

The responsibility of a school principal in planning and implementing suitable digital infrastructure in schools is covered in the paper. Also, digital competencies of school principals in the area of Digital Infrastructure together with school principals' tasks and activities in the area of efficient use of digital infrastructure have been described. Some challenges related to digital competencies in the area of Digital Infrastructure have been also presented.

2 The Role of a School Principal in Planning and Implementing Adequate Digital Infrastructure in School

Digital infrastructure is prerequisite for the digital transformation of schools, which is required for students (future employees), instructors, and other school personnel, as well as the school principal/management. Digital competencies needed for new workplaces should be acquired throughout early education. Teachers should base their teaching and learning activities for students on digital technologies. All other school business functions should be assisted by digital technologies. This is not practicable in the absence of proper digital infrastructure.

The term digital infrastructure is particularly challenging to define, because much of it represents new and evolving technologies (Glaeser & Poterba, 2021).

According to Tondeur, et al. (2009, as cited in Vermeulen et al., 2017), "an ICT-infrastructure refers to more components than the technical structure (PC, internet connection and the availability of educative programs) alone; it comprises ICT policy, ICT vision,

and sufficient degree of competence by colleagues to use ICT".

School principals as leaders have an important role in fostering, supporting, and assuring each component of the digital/ICT infrastructure. Okeke (2019) discusses technology leadership as "a new concept in school leadership, and it involves school principals creating and sustaining support for effective use of ICT in teaching and learning."

Sheninger (2019) defines digital leadership and its seven pillars as the specific areas embedded in the culture of all schools that can be improved or enhanced through the purposeful use of technology as follows:

1. Student Engagement and Learning
2. Learning Spaces and Environment
3. Professional Learning and Growth
4. Communication
5. Public Relations
6. Branding
7. Opportunity

After conducting a thorough literature review, the following basic publications were chosen to be used in determining the set of required digital competencies:

- Proposal of Strategy for digital maturation of schools and school system in the Republic of Croatia (2018),
- Standards for Principals of Educational Institutions (proposal) (2017),
- ISTE standards for teachers (ISTE, 2008),
- Norwegian Professional digital competence framework for teachers (Kelentrić, Helland & Arstorp, 2017),
- Australian Professional Standard for Principals and the Leadership Profiles (Australian Institute for Teaching and School Leadership, 2014),
- European Framework for Digitally-Competent Educational Organisations: DigCompOrg (Kampylis, Punie & Devine, 2015),
- European Framework for the Digital Competence of Educators: DigCompEdu (Punie, 2017),
- DigComp 2.0: The Digital Competence Framework for Citizens (Vuorikari et al., 2016), and
- laws connected to primary and secondary school education.

The members of the working group employed methods of content analysis, group decision making, brainstorming, brainwriting and panel discussions to identify the elements that made up the mapping of the aforementioned documents as a starting point for the identification of key competencies. This phase's primary contribution may be seen in the list of fundamental areas, the elements that go with them, and the critical competencies for each element. Through five categories and 37 components, this first phase identified eighty-two capabilities.

No improvement or enhancement are possible without adequate digital/ICT (technical) infrastructure and principals are supposed to possess competencies in this area in order to actuate them.

3 Digital Competencies of School Principals in the Area of Digital Infrastructure

Five digital competencies are created for the area Digital Infrastructure. The main steps of the methodology used to construct the framework are briefly discussed in this chapter, along with the created competencies and a simple explanation of each.

The first step specified the areas of school planning and management that matched the Framework for the Digital Maturity of Schools. Five areas in total have been identified: (1) Planning, Management, and Leadership, (2) Digital Technologies in Learning and Teaching, (3) Development of Digital Competencies, (4) Digital Culture and (5) Digital Infrastructure.

After identifying the important principal tasks for each of these areas, a set of digital competencies was determined through an analysis of the relevant literature.

The following step involved a team of experts performing the content validation of competencies. More than twenty digital competency experts were involved in the content validation. In addition to content validation, the experts evaluated the significance of competence for a certain defined area.

In addition to experts in content validation and digital competency clarification, the principals themselves participated in the validation of the principals' digital competencies. Validation was accomplished through focus groups that followed a predetermined process and were facilitated by a competent moderator.

After updating the competencies based on qualitative validation of experts and principals who took part in the research, the final version of the competencies was obtained.

The final phase involved describing each competency and identifying typical actions associated with each competency. The research ultimately produced a total of 25 competencies and more than 170 typical activities in the five previously mentioned areas. A complete description of the methodology development of the Framework for Digital Competencies of School Principals is given in the paper entitled "Methodology for developing a framework for digital competencies of school principals in Croatia" (Žajdela Hrustek, Kirinić, Mekovec, 2022). In the rest of the chapter, the principal's digital competencies related to the area of

Digital Infrastructure are presented together with short descriptions of the competencies.

Competency 5.1 "The principal plans and coordinates upgrades to the school's digital infrastructure according to the needs identified and in collaboration with experts, staff, founders and the Ministry (*of Science and Education*)."

In their development plans, missions and visions, educational institutions (schools) may present different requirements regarding digital infrastructure. In the Republic of Croatia, the Ministry designs and implements a strategy for improving the digital infrastructure of the education system in coordination with the Croatian Academic and Research Network (CARNET). The needs of each individual school may be greater than what the Ministry has foreseen in the budget. In such circumstances, the principal should consider the possibilities of securing additional financial resources in partnership with the Ministry, the founders, the school board and the community of parents, guardians and foster parents.

Competency 5.2 "The principal supports and coordinates the development and upgrades of the school's network infrastructure based on identified needs and capabilities."

The Ministry of Science and Education of the Republic of Croatia, in cooperation with CARNET, is developing a network infrastructure upgrade strategy based on the requirements of educational institutions (schools). Some educational institutions could have greater needs if teachers and students are better motivated and engaged in the use of advanced digital technologies. As a result, a new, more advanced network architecture, higher speeds, with more access points, subnets and requirements for improved network access monitoring may be required.

Competency 5.3 "The principal manages the distribution of digital devices according to the recognized staff and student needs, planned schedule and established criteria."

The principal, together with professional associates, is responsible for determining the conditions and distribution of digital devices in accordance with the monitoring of activities and the requirements of teachers and students. The obsolescence of technology, which is a consequence of the rapid and continuous development of digital technologies, is another key criterion for starting the process of procurement and adequate distribution of digital devices.

Competency 5.4 "In collaboration with the school staff, the principal determines the purpose of using software to support business, learning and teaching processes."

The principal, together with professional associates and all other employees of the educational institution (school) creates software usage models, considering the potential of digital tools as well as unique aspects of teaching and business processes. The principal should be a motivator and initiator of the preparation and implementation of plans to move from simpler to more complex software usage models in order to improve learning and teaching processes as well as business processes.

Competency 5.5 "The principal ensures and coordinates technical support and the school's infrastructure maintenance system."

Nowadays, as educational institutions (schools) use more and more digital equipment, there is an increasing need for technical and program maintenance. However, the end users (teachers, other employees and students) do not have the necessary knowledge to perform such activities. The principal should cooperate with the Ministry and the founders in order to secure additional funds for the employment of experts or the restructuring of the personnel responsible for maintaining the infrastructure in the school or, alternatively, engage specialized economic entities.

4 School Principals' Tasks and Activities in the Area of Efficient Use of Digital Infrastructure

In the "Area of Efficient Use of Digital infrastructure" for five competencies a total of twenty-four typical activities were identified.

For the competence "The principal plans and coordinates upgrades to the school's digital infrastructure according to the needs identified and in collaboration with experts, staff, founders and the Ministry." three key tasks/activities are recognized for modernizing school digital infrastructure. Firstly, the principal's task is to familiarize themselves with the Ministry's adopted plan for a certain period. It is also very important that the principal, in collaboration with other stakeholders, makes a plan for developing the school's digital infrastructure in accordance with the Ministry's plan. The principal should also be responsible for communication with other stakeholders about how to realise the procurement of digital infrastructure that the school requires but is not addressed by the Ministry's plan.

Four typical tasks/activities related to the second competency of the Digital infrastructure area were identified, "The principal supports and coordinates the development and upgrade of the school's network infrastructure based on identified needs and capabilities." As a starting point, the principal must become acquainted with the Ministry's strategy for improving the schools' network infrastructure. The principal of the school should be ready for constant cooperation with other stakeholders in order to be able to adequately develop a plan for the development of the school's network infrastructure in accordance with the plan of the Ministry. Furthermore, the principal should actively interact with other stakeholders on how to

achieve the procurement of network infrastructure that the school requires, which is not included in the Ministry's plan. The principal should be prepared to arrange individualized access to specific online resources for certain teachers and students to foster increased involvement, understanding, and faster development.

"The principal manages the distribution of digital devices according to the recognised staff and student needs, planned schedule and established criteria." is the third competency in the area of digital infrastructure. Five key tasks/activities were recognized. Principals must manage the tracking of digital devices in classrooms and labs, devices retained by teachers, other staff, and students, and monitor their obsolescence with teachers and professional assistants. Principals must, also organize the observation of various preparation and teaching activities, as well as horizontal and vertical communication to determine the necessity for new or specific digital gadgets (advanced use of technology). It is also important that the principal constantly checks teacher and student evaluation results using digital devices. Principals must participate in defining the criteria for distributing digital devices. A further very important task of the principal is to continuously oversee the process of distributing digital devices and authorisations based on identified forms of use.

There are six tasks/activities defined regarding the fourth competence "In collaboration with the staff, the principal determines the purpose of using software to support business, learning and teaching processes." One of the principal's primary responsibilities is to stay current on specific software features that assist business, learning, and teaching activities. It is crucial that the principal familiarizes oneself with the standards for software use set forth by the relevant Ministry. The principal is responsible for, in collaboration with the teachers and other staff, identifying how and what types of software is used and also developing software use models as support for business, learning and teaching processes. An educational institution's principal, with a vision for a successful digital transition, determines and provides the prerequisites for making a transition to more complex software use models with the aim of increasing the quality of business, learning and teaching processes and continuously motivates the staff to apply more complex software use models.

For the last competency within the area of Digital infrastructure "The principal ensures and coordinates technical support and the school's infrastructure maintenance system." six tasks/activities are defined. The principal's task should be to ensure and to coordinate technical maintenance of digital devices. The principal should communicate with the founders and the Ministry in order to provide additional funds for financial compensation to employees with appropriate competencies for the time spent on maintenance work, as well as, if such employees do not exist, obtain permission to enter into technical support contracts with companies specialized in the maintenance of digital infrastructure. The principal is also responsible for communication with the Ministry to obtain permission to hire a person (professional assistant) with the necessary competencies for performing maintenance tasks in schools that use many digital devices. In collaboration with experts, the principal should make a protocol for fulfilling technical maintenance requirements. One of the principal's top responsibilities should be to stay current on technical maintenance requirements in order to provide extra training for digital device users. (Kirinić, et al., 2022).

An overview of the principal's digital competencies along with typical activities in the area of Digital Infrastructure is also shown in Table 1.

5 Challenges Related to Digital Competencies in the Area of Digital Infrastructure

Five focus groups were held, with the main purpose of examining the proposed framework's clarity and comprehensibility and determining how it could be improved. The focus groups comprised representatives from five distinct types of schools. The first focus group was with primary school representatives. The second focus group included high school representatives. The third focus group included members from art schools, and the fourth included representatives from vocational schools. The fifth focus group was held with representatives of the centre for upbringing and education, that is, institutions for the upbringing and education of students with greater difficulties.

The principals in the focus group examined the framework and revealed some of the challenges they confront in their everyday administrative activities.
Challenges in Area Digital infrastructure are as follows:

- Procurement of ICT equipment (as well as network infrastructure) is an important skill, but it is dependent on the founder and is not practiced uniformly throughout Croatia. The practice varies and is dependent on new funding sources. It is debatable how much it reflects the scope of the principal's work; an ICT expert should be included; the principal may coordinate it, but using an expert to assist.
- In order to keep up with trends, it is necessary to seek funding sources. The principal should be aware of equipment insurance calls/tenders and be able to apply for them.

Table 1. Competencies and typical activities of school principals in the Area of Digital Infrastructure

5.1. The principal plans and coordinates upgrade to the school's digital infrastructure according to the needs identified and in collaboration with experts, staff, founders and the Ministry.	**5.2. The principal supports and coordinates the development and upgrade of the school's network infrastructure based on identified needs and capabilities.**	**5.3. The principal manages the distribution of digital devices according to the recognised staff and student needs, planned schedule and established criteria.**	**5.4. In collaboration with the staff, the principal determines the purpose of using software to support business, learning, and teaching processes.**	**5.5. The principal ensures and coordinates technical support and the school's infrastructure maintenance system.**
The principal: - familiarises himself/herself with the Ministry's plan for upgrading the digital infrastructure of schools adopted for a certain period; - in collaboration with other stakeholders, makes a plan for developing the school's digital infrastructure in accordance with the Ministry's plan; - communicates with other stakeholders about how to realise the procurement of digital infrastructure that the school requires, but is not addressed by the Ministry's plan.	The principal: - familiarises himself/herself with the Ministry's plan for upgrading the schools' network infrastructure adopted for a certain period; - in collaboration with other stakeholders, makes a plan for developing the school's network infrastructure in accordance with the Ministry's plan; - communicates with other stakeholders about how to realise the procurement of network infrastructure that the school requires, but is not addressed by the Ministry's plan; - organises an individualised approach to individual network resources for particular teachers and students due to greater engagement, knowledge and faster development.	The principal: - coordinates recordkeeping of digital devices in the classrooms and labs, devices kept by the teachers, other staff and students, and also monitors their obsolescence; - coordinates the monitoring of various types of activities for preparing and holding classes, as well as vertical and horizontal communication for the purpose of identifying the needs for specific/new digital devices (advanced use of technology); - checks teacher and student evaluation results using digital devices; - participates in defining the criteria for distributing digital devices; - oversees the process of distributing digital devices and authorisations based on identified forms of use.	The principal: - keeps up with particular software features for supporting business, learning and teaching processes; - familiarises himself/herself with the competent Ministry's requirements regarding the standard form of software use; - in collaboration with the teachers and other staff, identifies how and what types of software is used; - in collaboration with the teachers and other staff, develops software use models as support for business, learning and teaching processes; - determines and provides the prerequisites for making a transition to more complex software use models with the aim of increasing the quality of business, learning and teaching processes; - motivates the staff to apply more complex software use models.	The principal: - ensures and coordinates technical maintenance of digital devices; - communicates with the founders and the Ministry to ensure additional funds to compensate the employees with adequate competencies for the time spent performing maintenance tasks; - communicates with the Ministry to obtain permission to hire a person (professional assistant) with the necessary competencies for performing maintenance tasks in schools that use many digital devices; - communicates with the founders and the Ministry to obtain permission for concluding a technical support contract with specialised business operators; - in collaboration with experts, makes a protocol for fulfilling technical maintenance requirements; - keeps up with technical maintenance requirements for the purpose of organising additional training courses for digital device users

- If the principle does not have sufficient financial resources, he will struggle to manage technical support. The founders have ultimate authority, whereas the principals can merely collaborate. There are no conventional answers; it differs between each school and per principal.
- Some principals have attempted to find financial resources to establish their own repository (central storage), but due to limited financial resources, this is not feasible for all schools. They should have an IT specialist on staff at least part-time.
- The principals lack the necessary expertise to manage an information security system. Because both teachers and students are involved in a large system, the principal faces challenges if he does not have the support of competent colleagues. Licenses exist, as do persons who renew and maintain them. Is there a legal procedure in place? Who gives technical assistance? A principal appears to be overburdened. Seminars are essential for the principal's preparation. If they acquire a procedure from someone, it won't be a problem.

6 Conclusion

As the leaders of educational institutions, principals have a significant role in the development, maintenance, and quality assurance of their digital and information technology (ICT) infrastructure.

The primary purpose of the research that is presented in this paper was to define the key digital competencies of principals related to the field of digital infrastructure. This was accomplished by searching for and studying the relevant literature, as well as enlisting the assistance of specialists in the field of digital competencies and the primary stakeholders (principals).

In this paper, the function of a school principal in the formulation and execution of an appropriate digital infrastructure in educational institutions is analysed and explored. In addition, descriptions have been provided of digital competencies held by school principals in the domain of digital infrastructure, as well as of the responsibilities and actions held by school leaders in the domain of making effective use of digital infrastructure.

The main steps of the methodology used to construct the framework are briefly discussed along with the created competencies and a simple explanation of each.

In addition to specialists in content validation and digital competency clarity, the principals themselves took part in the process of validating their own digital skills as part of the validation of the principals' digital competencies. Focus groups that followed a predetermined protocol and were conducted by an experienced moderator were used to carry out the validation process. The final version of the competencies was achieved after the competencies were updated based on qualitative validation by the experts and principals who participated in the research. This was done in order to obtain the final version of the competencies.

The key limitation of the research described in this paper is that it is restricted to establishing the input guidelines for developing the Framework. The framework was established with a focus on specific documents that should be considered, hence the full opus of needed skills and knowledge was not included in some components. Furthermore, only Croatian experts and principals participated in the validation, so the framework itself reflects only the situation in Croatia.

Future study should include updating the framework in light of ongoing changes in the development of digital technologies and digital competencies.

Acknowledgments

The research was conducted within the project "e-Schools: Development of the System of Digitally Mature Schools (Second Phase)" being financed from the European Regional Development Fund, European Social Fund and the state budget.

References

Australian Institute for Teaching and School Leadership (2014). *Australian Professional Standard for Principals and the Leadership Profiles.* Retrieved from https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standard-for-principals-and-the-leadership-profiles.pdf?sfvrsn=c07eff3c_24

Glaeser, E. L., & Poterba, J. M. (Eds.). (2021). *Economic analysis and infrastructure investment* (Vol. 28215). University of Chicago Press.

ISTE (2008). *ISTE standards for teachers.* Washington: International Society for Technology in Education

Kampylis P., Punie Y., & Devine J. (2015). *Promoting Effective Digital-Age Learning - A European Framework for Digitally-Competent Educational Organisations. No. JRC98209.* Joint Research Centre (Seville site), Retrieved from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC98209/jrc98209_r_digcomporg_final.pdf

Kelentrić, M., Helland K. & Arstorp, AT. (2017). *Professional digital competence framework for teachers*. The Norwegian Centre for ICT in education. Retrieved from

https://edudoc.ch/record/131449/files/pfdk_framework.pdf

Kirinić, V., Hutinski, Ž., Mekovec, R., Žajdela Hrustek, N., Jugo, G., Dejanović, L., Bilić Meštrić, K., Tingle, J., Maravić, J. (2022) *Framework for the Digital Competence of School Principals.* Retrieved from https://www.e-skole.hr/wp-content/uploads/2023/04/Okvir-za-digitalne-kompetencije-ravnateljica-i-ravnatelja-HRV.pdf

Okeke, N. L. (2019). School technology leadership: A new concept. *International Journal of innovative development and policy studies*, *7*(2), 50-56.

Proposal of Strategy for digital maturation of schools and school system in the Republic of Croatia. (2018). Retrieved from https://pilot.e-skole.hr/wp-content/uploads/2018/12/Strategija-digitalnog-sazrijevanja-cjeloviti-tekst.pdf

Punie Y. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*. (Editor: C. Redecker), EUR 28775 EN, Publications Office of the European Union, Luxembourg, Retrieved from https://publications.jrc.ec.europa.eu/repository/bitstream/JRC107466/pdf_digcomedu_a4_final.pdf

Sheninger, E. (2019). *Digital leadership: Changing paradigms for changing times.* Corwin Press.

Standards for Principals of Educational Institutions (proposal). (2017). Retrieved from https://www.azoo.hr/images/NATJECANJA2017/Prijedlog_standarda_za_ravnatelje_.pdf

Tondeur, J., Devos, G., Van Houtte, M., van Braak, J., & Valcke, M. (2009). Understanding structural and cultural school characteristics in relation to educational change: The case of ICT integration. *Educational Studies*, *35*(2), 223–235.

Vermeulen, M., Kreijns, K., Van Buuren, H., & Van Acker, F. (2017). The role of transformative leadership, ICT-infrastructure and learning climate in teachers' use of digital learning materials during their classes. *British Journal of educational technology*, *48*(6), 1427-1440.

Vuorikari, R., Punie, Y., Carretero Gomez S. & Van den Brande G. (2016). *DigComp 2.0: The Digital Competence Framework for Citizens, Update Phase 1: The Conceptual Reference Model. Luxembourg Publication Office of the European Union. EUR 27948 EN,* Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101254/jrc101254_digcomp%202.0%20the%20digital%20competence%20framework%20for%20citizens.%20update%20phase%201.pdf

Žajdela Hrustek, N., Kirinić, V., Mekovec, R. (2022). Methodology for developing a framework for digital competencies of school principals in Croatia. In Gómez Chova, Luis ; López Martínez, Agustín ; Lees, Joanna (Eds.), *Proceedings, 14th International Conference on Education and New Learning Technologies (EDULEARN22)* (pp. 5618-5625). Valencia: IATED, doi:10.21125/edulearn.2022.

Education and Learning Analytics

Till Winkler, Rony G. Flatscher
Cognitive Load in Programming Education: Easing the Burden on Beginners with REXX

Bojan Radišić, Ivan Dunđer, Sanja Seljan
Data Analysis of the Motivation and Factors for a Shorter Duration of Study

Melita Milić , Blaženka Divjak
Innovative Learning Design and Teachers' Competencies: Prerequisites for Meaningful Use of Digital Material in Schools

Marko Matus, Goran Hajdin, Igor Balaban
Joint Creative Classroom as a Response to Industry 4.0 Demands

Ivana Ružić, Igor Balaban
The Use of Social Robots as Teaching Assistants in Education: Literature Review

Blaženka Divjak, Josipa Bađari, Darko Grabar, Damir Horvat, Bart Rienties, Petra Vondra
Enhancing Learning Design through User Experience Research: Insights from a Survey in Four European Countries

Josipa Lipovac, Dorotea Potoč, Tomislav Jagušt
How Programming Helps Early-Age Learners Tackle Complex Concepts

Ivan Kekez, Diana Šimić
Learning Analytics Frameworks: A Review of Challenges and Practices in Higher Education

Cognitive Load in Programming Education: Easing the Burden on Beginners with REXX

Till Winkler, Rony G. Flatscher
Vienna University of Economics and Business
Institut für Wirtschaftsinformatik und Gesellschaft
Welthandelsplatz 1, 1020 Wien, Austria
{till.winkler, rony.flatscher}@wu.ac.at

Abstract. *To learn and teach programming is very difficult, often leads to poor results, and causes many students to drop out or turn away from the subject. Cognitive load theory can help to understand the challenges students face, improve programming education, and select an appropriate language for instruction. In this paper, we take a theoretical look at programming education and, in particular, language characteristics that reduce students' cognitive load and thus enable rapid learning and frustration-free productivity. We introduce the REXX language and some of its favorable characteristics that make it possible to teach novices programming within a single semester. In this limited time, students are empowered to program Microsoft products (Windows, Office), address the command line, grasp the basics of object-oriented programming, use Java classes, and create portable graphical user interfaces (GUIs) with JavaFX.*

Keywords. Programming education, Cognitive load theory, Human-oriented programming, REXX, ooRexx

1 Introduction

It has always been a topic of discussion as to which programming language should be taught. With Wikipedia listing 691 (“List of programming languages,” 2023) different programming languages, such a choice can be overwhelming. Moreover, when beginners ask an expert, they usually get answers based on individual preferences, which can be confusing. It can be equally confusing to rely on popularity ratings of programming languages. Until the mid-1980s, the most popular choices were Fortran, Pascal, or Ada; in the 1990s it was clearly C; in the late 2000s it was Java; and today it is, with a 28.4% certainty, Python (“Data is Beautiful,” 2019; PYPL, 2023). The fact that language popularity scores nowadays are often calculated based on the frequency of online searches, such as for tutorials, should not be ignored (PYPL, 2023). Since we believe that it is not the amount of support needed that matters, but the ease with which a language can be learned, we would like to present our experience with using REXX in programming classes. Specifically, we will discuss how language characteristic can place an unnecessary cognitive burden on students. In addition, we will give illustrative examples of REXX, a language we consider particularly suitable for teaching.

Learning a programming language can be challenging for beginners, as they need to grasp the syntax, semantic and language-specific concepts such as variables, data types, arithmetic, and others (Sands, 2019; Stachel et al., 2013). Moreover, students must quickly apply new knowledge to solve complex and often novel problems. It is well known that it is a combination of students' lack of experience, understanding new concepts, applying syntactic and semantic rules, and solving new complex problems that can be overwhelming (Sands, 2019). Programming courses are generally considered difficult, with high dropout rates and poor outcomes; some students cannot program loops even after several semesters (Robins et al., 2003). Many programming educators find that students achieve poor grades or, more importantly, become disillusioned with programming (Garner, 2002). This is in contrast to what we observe in REXX teaching. According to the course evaluation (WS22/23), 83.3% of students would definitely recommend the class to others and consider the demands to be reasonable (66.6%) or slightly taxing (33.3%).

The second author became acquainted with REXX on IBM mainframes in the 1980s and developed an experimental course using the PC version of REXX. To his surprise, it was possible to teach programming concepts much faster compared to VBScript, a language considered easy to learn at the time. REXX was developed at IBM (Cowlishaw, 1987) with the motivation of creating a "human-oriented" language—by keeping it small—that is easy to learn, code, remember, and maintain (Fosdick, 2005). At that time REXX was extremely successful; Amiga OS used it as a script language (“AmigaOS Manual: Arexx,” 2023)

and several companies developed interpreters. Today, REXX is still an integral part of IBM mainframes and has been also formalized as an ANSI/INCITS X3.274 standard (ANSI, 1996). In the 1990s, IBM developed an object-oriented successor to REXX called ooREXX, which is open source and has been available for all major systems since 2005 (ooRexx, 2023). Under Windows, ooREXX allows the direct use of COM/OLE, which enables direct interaction with many Microsoft products. For applying the acquired skills on other platforms, a Java bridge called BSF4ooREXX is available, which disguises Java as ooREXX and allows for the use of Java classes (BSF4ooRexx, 2023).

Over the past 35 years, what was once an experimental course has evolved into two programming courses[1] that teach students the necessary skills to solve complex programming tasks in a single semester (Flatscher & Müller, 2021). Within the first two months, during "Business Programming 1" (BP1), students learn basic programming concepts, fundamentals of object-oriented programming and everything necessary to use COM/OLE in Windows (BP-1, 2023; Flatscher & Müller, 2021). The following two months, during "Business Programming 2" (BP2), are dedicated to the Java bridge (BSF4ooRexx, 2023) and include the use of Java classes including the development of platform independent GUI applications with JavaFX (BP-2, 2023; Flatscher & Müller, 2021). Some students are even so motivated that they write seminar papers, bachelor's and master's theses that go far beyond what they originally learned (WU, 2023). We believe that this learning outcome and motivation is primarily related to language characteristics of REXX that reduce students' cognitive load and minimize frustration. Before looking more closely at specific language characteristics, we will introduce the perspective of cognitive load theory on learning, problem solving and programming education.

2 Cognitive Load Theory

Human expertise and problem-solving skills, are based on knowledge stored as so-called schemata in our long-term memory (Sweller & Van Merriënboer, 2005; Garner, 2002). A schema might be anything that can be treated as a single element; for instance, a word, a mathematical formula, or a particular programming concept (Garner, 2002). During learning, multiple new or previously disconnected pieces of information are bundled together into a single, more complex element or schema (Paas et al., 2003). For example, it is almost impossible to remember all the digits of a telephone number individually unless bundled into more complex elements, such as, country code (two digits), area code (four digits) and the remainder as three-digit blocks ("141" instead of "1 - 4 – 1"). In this way, a twelve element/digit number can be remembered easily. The same basic principle applies to any kind of learning, including physics, mathematics, spoken languages and programming. The general goal of teaching is to enable the construction of increasingly complex schemata and to facilitate their automation through practice (Paas et al., 2003). A practiced stick driver has automated the procedure of shifting gears to a point where he or she no longer needs to think about it, whereas a novice driver requires active processing of each step, which can be tiring and frustrating. Similarly, a skilled programmer can easily create a "selection block", while a beginner must actively think about the necessary structure, syntax, variables and boolean symbols.

More complex schemata can only be built if the brain is actively involved in the learning process, for which free working memory capacities are needed (Sweller & Van Merriënboer, 2005). In other words, students must actively think about new programming concepts. Unlike the nearly unbound long-term memory, however, our working memory can only deal with up to four elements or schemata at a time (Sweller & Van Merriënboer, 2005). People who are able to handle complicated programming tasks do not think more sophisticatedly or process more elements, but already have complex schemata that are treated as a single element. In comparison, an inexperienced programmer must process many different details (elements) in his limited working memory. When details or new information overwhelm the capacity of working memory, problem-solving performance and learning success decline (Sweller, 1988). During programming education in particular, as in any other problem-solving area, the cognitive load on novices must, therefore, be carefully managed (Garner, 2002; Paas et al., 2003). There are three different types of cognitive load that essentially fight for the limited resources of working memory. The intrinsic cognitive load is caused by the learning content itself: the programming language with its individual and interacting elements (Sands, 2019). Extraneous cognitive load is a burden on top of the content, that may be caused by information search or inappropriate teaching methods (Sweller & Van Merriënboer, 2005). Intrinsic and extraneous cognitive load can add up to such an extent that there is no capacity left for germane cognitive load. Germane cognitive load is necessary for learning through thinking about new information and concepts (Paas et al., 2003). The chosen programming language and teaching methods must facilitate the construction of schemata without overwhelming limited cognitive capacities (Garner, 2002).

[1] In the spirit of open education, the course material is freely available (see BP-1, 2023; BP-2, 2023).

2.1 Intrinsic Cognitive Load

It is especially the degree of interactivity between novel elements that can produce high intrinsic cognitive load (Garner, 2002). Simply learning vocabulary, for example, produces a relatively low burden because each word—an individual element—can be learned separately (Garner, 2002). Learning grammar, on the other hand, produces more intrinsic cognitive load because the words in a sentence are connected and their interactivity must be taken into account (Garner, 2002). In this sense, learning a programming language is an extremely high cognitive burden, since abstract concepts and the syntax, so to speak, and the grammar of a language has to be learned (Sands, 2019; Stachel et al., 2013). It is often assumed that the intrinsic cognitive load caused by the learning content cannot be reduced (Sands, 2019; Garner, 2002). This is not entirely true for programming, as we can choose a language with fewer abstract concepts and a simpler syntax, which is an advantage that other fields do not have.

Many educators—here referring to C and VisualBasic.NET—see "...the excessive amount of class time spent on teaching the language syntax…" (Al-Imamy et al., 2006, p. 280) as a major issue. While a C-style syntax has influenced languages such as Java, PHP, Go, or Swift, it is challenging for beginners (Denny et al., 2011; Stefik & Siebert, 2013). Since learning syntax is a common challenge, tools and languages have been developed to bypass syntax altogether, but even so, general-purpose programming languages are still predominantly used in classrooms (Stefik & Siebert, 2013).

Some syntactic choices made by language designers are perceived as easier to understand because they are more similar to knowledge or schemata from other domains (Stefik & Siebert, 2013). Unfortunately, most languages require consideration of unnecessary elements and interactions that students take time to learn. In Java, for example, a keyword specifying the data type is required before the name of a variable is stated in the declaration (Sands, 2019). While an experienced Java programmer doesn't have to think about it—having automated schemata—this can be a burden for beginners. In general, strictly typed variable declarations pose a major challenge for beginners, with dynamically typed languages being perceived as more intuitive (Stefik & Siebert, 2013). From a cognitive load perspective, omitting such declarations reduces the number of elements and interactions a novice must consider in working memory, thereby freeing up capacity.

While experienced programmers are already familiar with abstract characteristics of programming languages, beginners tend to find aspects that are not literal or are rooted neither in English or mathematics difficult to understand (Stefik & Siebert, 2013). For instance, novices are able to use statements like repeat ten times more accurately than traditional C-style looping syntax (Stefik & Siebert, 2013). The word repeat, or loop is simply more common in English and can be understood literally, as opposed to for. Moreover, the use of a single equal sign is perceived by beginners as easier to grasp than that of a double equal sign (Stefik & Siebert, 2013). The meaning of a single equal sign is a schema developed in mathematics education, while a double equal sign is rather uncommon. An intuitive language should be designed so that prior, non-programming knowledge can be applied as expected (McIver & Conway, 1996). When choosing a language to teach, we need to put ourselves in the beginner's shoes and recognize how many new aspects are necessary to understand, and what existing schemata from other fields can be useful. This is recognized by many teachers who choose a language primarily for pedagogical reasons rather than popularity or industry relevance (Mason et al., 2012). In terms of students' future and cognitive load theory, it makes sense to focus on a general-purpose language because it is much easier to transfer schemata from one general-purpose language to another. It is simply not possible to predict which language will be popular or desired when students start working in the industry. Therefore, we must enable students to master one language without frustrating them, as they can easily switch to another language later if needed.

Learning success is of course also strongly influenced by the teaching methods, which should aim to keep the extraneous cognitive load as low as possible in order to free up cognitive processing resources for learning (Sweller & Van Merriënboer, 2005).

2.2 Extraneous Cognitive Load

Extraneous cognitive load is an additional burden that is not required for learning (germane load) and is not directly related to the content (intrinsic load) (Sands, 2019). Such a cognitive burden can, for instance, be caused by redundant, unnecessary or too frequently expressed information (redundancy effect) (Sweller & Van Merriënboer, 2005). Additionally, extraneous cognitive load may be imposed by lengthy web searches for information needed to complete a task, or by spreading relevant information across multiple lessons, textbooks or reference manuals (locations or times) (Sands, 2019). Various methods such as pair programming or presenting worked or nutshell examples are known to reduce cognitive burden. During pair-programming learners can split minimally demanding tasks—typing, navigation, or file management—and highly demanding tasks—syntax development or solution search—among themselves, thereby reducing the cognitive burden (Sands, 2019). Presenting practical—worked or nutshell—examples where students are shown a solution step-by-step from start to finish, helps to break down a complex and novel problem into meaningful steps and provide scaffolding for other problems (Sands, 2019; Stachel et al., 2013).

While extraneous cognitive load can mostly be reduced by suitable teaching methods and materials,

part of the load is also caused by the programming language. For example, when a language forces to use an unfamiliar operating system or a complex IDE, basic tasks such as typing, file management, or navigation become an unnecessary additional cognitive burden.

2.3 Germane Cognitive Load

To learn, novices must actively invest free cognitive resources; that is, an appropriate germane cognitive load should be elicited (Sweller & Van Merriënboer, 2005). Freeing up working memory capacity by reducing intrinsic and extraneous cognitive load is only effective if students are motivated to actively invest cognitive effort in schemata construction (Paas et al., 2003). For this, it is important that students are not frustrated with the teacher, grading, or programming in general. In many classes, a common problem is that students use a passive elaboration strategy; they do not use free capacities to self-elaborate new concepts (Sweller & Van Merriënboer, 2005). Common teaching methods to promote active elaboration in class are to have students annotate worked examples or complete missing code from a well-structured program (Garner, 2002).

It is well-known that practicing programming concepts in variable situations has a positive effect on schemata building and educational transfer (Paas et al., 2003). In order to have enough time to practice, the chosen programming language should have characteristics that are easy to understand and therefore do not take up unnecessary time in class.

3 Language Characteristics

REXX coding can be achieved with a simple editor (e.g., Notepad, gedit) or more complex IDEs (e.g., IntelliJ). While gedit is equipped with REXX syntax highlighting by default, Intellij requires a readily available plugin (Seik, 2023). This allows students to select a tool with which they are most familiar, thereby reducing extraneous cognitive load. While for most languages a simple editor is sufficient, for Python it is advisable to use an IDE, as it is necessary to create intended blocks or include and manage packages for basic functions.

REXX was developed with the goal of creating a "human-oriented" language that is small, as well as being easy to learn, code, remember, and maintain (Fosdick, 2005). In the limited time available for teaching, large languages such as C++ or Java can only be taught by focusing on a subset of the entire language, and intentionally ignoring important aspects (McIver & Conway, 1996). This can be confusing because textbooks or online tutorials rarely adhere to the same subset, and beginners may encounter features that were intentionally not taught. In comparison, REXX is a small but powerful language that can be taught in a short time. All necessary knowledge is bundled in a single reference manual (ooRexx, 2023). This eliminates the tedious search for information and thereby reduces extraneous cognitive load (Sands, 2019). In addition, the reference manual itself provides brief and meaningful explanations, syntax diagrams and nutshell examples. Figure 1 shows a syntax diagram for the Strip method; such diagrams are used for all methods and functions in the reference manual (ooRexx, 2023). With its multimodal presentation (description and visualization) of key knowledge and its nutshell examples, the manual does a good job of reducing unnecessary cognitive load (cf. Sands, 2019; Stachel et al., 2013).

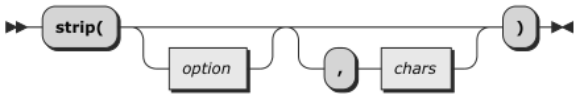


Figure 1. Syntax diagram (ooRexx, 2023, p. 206)

Many believe that making and correcting mistakes is the best way to learn. However, the inadequacy of error messages is a problem that dates back to COBOL, but is still a problematic in C++ or Java (Becker et al., 2016). Errors and related messages of a compiler or interpreter should be understandable without knowing technical jargon (McIver & Conway, 1996). Unfortunately, error messages are often "… terse, confusing, too numerous, misleading, and sometimes seemingly wrong...", this way "...they become a source of frustration and discouragement" (Becker et al., 2016, p. 21). We consider the error messages of the ooRexx interpreter to be clear and, above all, precisely pointing to the source of an error. Most importantly, REXX's free-form syntax, its case-insensitive nature, and its use of dynamic data types avoid many common errors from the outset. Avoiding such errors helps reduce frustration or disillusionment with programming, which can motivate students to invest the cognitive load required to construct schemata and automate these through practicing (Garner, 2002; Paas et al., 2003).

3.1 Free-form Syntax

REXX has a free-form syntax where the positioning of the code is irrelevant. By default, the interpreter merges multiple blanks into a single one before execution. If this behavior is not desired, quotation marks (" or ') can be placed directly next to each other or two vertical bars can be used directly as concentration operators (||). Strings in REXX can be merged by listing them one after another in a single expression and delimiting them with blanks. A string encapsulated by quotation marks is not changed. Figure 2 provides examples of this.

In teaching, the free form of REXX allows for the creation of readable, consistent, and intentionally eye-catching syntax that helps to convey new concepts to

novices who typically have difficulty grasping the signals of novel concepts (McIver & Conway, 1996).

```
1  say  "Hello World!"              /* output: Hello World!     */
2  say " This"  'is'  "REXX!"       /* output:  This is REXX!   */
3  say "Good""bye" || '!'           /* output: Goodbye!         */
```

Figure 2. Free-from syntax and string merge

Such flexibility is also important for learning success, since fewer syntactic rules, details, or elements are relevant, and this therefore imposes a lower intrinsic cognitive load. In contrast to this flexibility, spaces or indentations have a semantic meaning in Python (e.g., for conditional statements). Although it was probably a good intention to eliminate grouping constructs, which reduces the number of elements (e.g., parentheses, do, end, …) and enforces structure, students do not seem to be able to master the concept of consistent indentation (McIver & Conway, 1996). This Python characteristic is contrary to non-programming knowledge that novices typically have, and can be considered to be a case of excessive cleverness (McIver & Conway, 1996). A text written in a natural language is understandable even with random indentations, and this is how beginners implicitly expect a programming language to behave. REXX's free-form syntax ensures that this reasonable expectation is met. Such violations of non-programming expectations, as committed by Python, are probably the "worst pedagogical sin" a programming language can commit (McIver & Conway, 1996, p. 4).

3.2 Case-insensitivity

Unlike in most programming languages, the case of symbols used in REXX is irrelevant. It does not "bother" the interpreter whether a beginner writes *do*, *Do*, *dO* or *DO* by mistake or on purpose. The REXX interpreter will uppercase all characters outside of quoted strings before executing them. This applies equally to all aspects of the language, including variable names, statements, functions, methods, method options, and so on. Figure 3 provides an example of this. While the strip method (see Figure 1) removes leading and trailing characters by default this behavior can be changed by an option, and a character can also be specified to replace blanks. For example, if you write "*Leading"*, "*leading"*, "*LeaDing"*, "*l"* or "*L"*, which all give the same result, only the leading blanks will be removed. This example shows that in addition to being case-insensitive, an option can also be spelled out, which makes its effect literally understandable. Such literal comprehensibility, further reduces the amount of learning (intrinsic cognitive load) for novices. A Python beginner, on the other hand, must first learn the meaning of *strip()*, *rstrip()*, or *lsrtip()* and build up the schema that an "l" here is an abbreviation for "*leading"*. While it is obviously clever to use abbreviations, forcing such behavior is another case of excessive cleverness.

```
1  a = "  This"   'is'   "REXX!"   /* a merged string          */
2  Say A                           /* output:  This is REXX!   */
3  SAY a~Strip("LEADing")          /* output: This is REXX!    */
4  say A~strip("l")                /* output: This is REXX!    */
```

Figure 3. Case-insensitivity

Considering variables, a Python novice must be careful when naming or referring to these, because a single case difference makes them distinct; ***O****ranges* and ***o****ranges* in this case are in fact two different things (variables). Such a distinction between cases is an additional element or rule that novices must learn, which unnecessarily increases the intrinsic cognitive load and may lead to frustrating syntax errors. Case dependence also violates the expectation of natural language schemata that an Orange remains an orange regardless of its case.

If someone new to a natural language makes a grammatical error—analogous to a syntax error—he or she can still accomplish the intended task of communication if the other person has a basic level of generosity and flexibility. However, a typical compiler or interpreter is by no means generous or flexible, but will mercilessly reject any slight deviation. This can be frustrating for students because they cannot achieve their goal of creating an executable program. The free-form and case-insensitive nature of REXX makes the interpreter more generous and flexible, and allows students to write a form of pseudo-code without frustration.

3.3 Data Type and Arithmetic

The REXX language has a single data type, a string value, which is immutable. Arithmetic is possible if the string contains numbers. The REXX interpreter defines the datatype implicitly with assignment or in the context of instructions. Compared to strictly typed languages, this eases the intrinsic cognitive load on students (Stefik & Siebert, 2013). When assigning variables, REXX students must consider fewer elements (e.g., no declaration of integer, float, ...) and their interaction with the rest of the program. It is not necessary to think about the required calculation precision in advance, as is the case in mathematics classes.

REXX's arithmetic, defined in ANSI/INCITS X3.274, formed the basis for the definition of ISO/IEC/IEEE standards that have been used to implement decimal arithmetic in languages such as Java, Python, and others (ANSI, 1996; Cowlishaw, 2022). By default, nine significant digits are used for the calculation, but this precision can be adjusted if desired. In REXX, variable names can start with a letter, an underscore, an exclamation mark or a question mark, followed by the same set of symbols and additional numbers and dots. All variables that

contain a dot become compound variables, which can be used to represent associative arrays. In this way, associative arrays can be declared without much effort and can be used like a typical variable. An example of this can be seen in Figure 4.

```
1  var = 6 * 7              /* assign and evaluate 6 * 7                */
2  say var                  /* output: 42                               */
3  stem.1 = 4               /* assign 4 to compound variable            */
4  say var - stem.1 / 0.7   /* output: 36.2857143                       */
5  numeric digits 20        /* now use 22 digit precision               */
6  say var - stem.1 / 0.7 /* output: 36.285714285714285714              */
```

Figure 4. Basic arithmetic and stem variables

3.4 Instructions

The ANSI/INCITS REXX standard defines assignment, keyword and command as three distinct instruction types (ANSI, 1996). An assignment instruction in REXX consists of a variable name, a single equal sign (=) as assignment operator, and an expression that contains the string that is assigned. The assignment *"var = 6 * 7"* would evaluate the expression (a multiplication) and assign the result 42 to the variable var (line 1 in Figure 4).

A keyword statement begins with a keyword; for example, *address*, *say*, *if*, *call*, *do*, *loop*, *parse* and others. Note that these keywords reflect their meaning in literal English. In this way, students can further draw on the schemata they have acquired in English classes. For illustration, Figure 5 shows a REXX program and Figure 6 shows a Python program with the same functionality.

```
1    /* an assignment instruction:                                      */
2  a = "Hello World!"          /* assigns "Hello World!" to a          */
3    /* an assignment instruction:                                      */
4  say a                       /* output: Hello World!                 */
5    /* an command instruction:                                         */
6  "dir a.txt"                 /* command: list the file a.txt         */
7   /* variable RC contains the command's return code                 */
8  if rc = 0 then say "found!" /* 0 means success                      */
9       else say "some problem occurred, rc="rc  /* shows rc          */
```

Figure 5. Instructions in REXX

A quoted string, including a variable or an expression evaluated as a string, is recognized by the REXX interpreter as a command instruction (line 6 in Figure 5). By default, the command is executed as if it were typed in a command line. The return code is made available immediately via the *rc* variable (line 8 in Figure 5). This feature made REXX popular on mainframes as it facilitates addressing the operating system, editors and utilities. If experience with the command line is available, solutions can be found with the existing system functions even without great programming knowledge.

Figure 5 contains the *if* keyword instruction with a dependent then and an else keyword instruction (line 8f in Figure 5). Depending on the programmer's preference, these instructions can be on separate lines. The indentation here is a preference decision and does not change the semantics of an instruction. In comparison, indentions in Python (see Figure 6) have semantic meaning and are mandatory, which limits flexibility and dictates programmer preferences. To understand or even write the Python program in Figure 6, many more details must be considered. For example, a module called *subprocess* must be imported (line 6 in Figure 6), its *run()* method called to submit the command to the system (line 8 in Figure 6), and the strictly int-typed return code fetched (line 9 in Figure 6). It should also be noted that two equal signs (==) represent an equality and one sign (=) represents an assignment operator (line 10 in Figure 6). Also, the built-in function *str()* must be known if concentration is desired (line 11 in Figure 6). Only if the students then also manage to put the colons (:) and the indentation correctly do they achieve a working program.

```
1  # an assignment instruction
2  a="Hello Word!"  # assigns "Hello World!" to a
3  # no keyword instruction, using built-in function()
4  print(a)
5  # no command instruction, using module subprocess instead
6  import subprocess
7  # execute command
8  completedProcess=subprocess.run("dir a.txt", shell=True)
9  rc=completedProcess.returncode  # fetch return code, an int
10 if rc==0:
11     print("found!")                  # indentation mandatory
12 else:
13     print("some problem occurred, rc="+str(rc)) # to string
```

Figure 6. Instructions in Python

The amount of time required to explain all the necessary Python concepts in class before students can productively write such a program is enormous. This is not only problematic given the limited time in class, but also puts a strain on student cognitive capacity and motivation. From the perspective of cognitive load theory, a much greater intrinsic cognitive load is generated, straining the limited resources of working memory for the necessary germane cognitive load.

3.5 Built-in and External Functions

REXX defines about 80 built-in functions, the number of which has been kept stable over the last 40 years. Even though the number of built-in functions may seem limited, they are powerful and more than enough to be productive. For example, functions that require additional packages in Python, such as root calculations (e.g., *sqrt()*), are already integrated. REXX can be extended with external functional libraries using the *::requires* directive. Such libraries are easy to write and are usually organized around domain-specific functions and are only included on a per-program basis.

As with the strip method (see Figures 1 and Figure 3), most built-in functions have a default behavior that can be changed by options. The *date()* function, for instance, returns the date as a "6 Mar 2023" string, while *date('s')* returns the string suitable for sorting as "20230306". In teaching, we consider it helpful that built-in functions and methods work without specifying options, so that you can use them from the beginning without worrying about details.

Mastering Java syntax is often seen as a major obstacle even for good students (Denny et al., 2011). The BSF4ooREXX Java bridge (BSF4ooRexx, 2023) enables students to use these functions without having to deal with the demanding Java syntax. Figure 7 shows a simple example, where BSF4ooREXX is included as if it were an external function (line 6 in Figure 7) and the *javax.swing.JFrame* class is invoked and the message show sent to it (line 1 in Figure 7).

```
1  frame=.bsf~new("javax.swing.JFrame", "Hello, my beloved
     world - from ooRexx!")
2  frame~setSize(410,20)   /* set width and height          */
3  frame~visible=.true     /* make JFrame visible           */
4  call SysSleep 10        /* sleep for ten seconds         */
5
6  ::requires "BSF.CLS"    /* get access to Java bridge     */
```

Figure 7. Invoke javax.swing.JFrame class

The result is a user interface frame titled "Hello, my beloved world - from ooRexx!". The output can be seen in Figure 8, which shows how easy it can be to create GUI programs for any modern operating system. The ease with which external functions can be written and included, and the simplicity of how the operating system can be addressed and COM/OLE objects or Java classes can be used, makes REXX more than "just" a language for beginners.

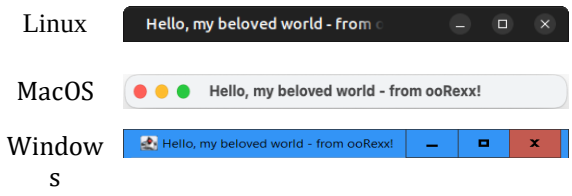


Figure 8. Result of code in Figure 7

3.6 Object-oriented

As an object-oriented language, useful base classes, data encapsulation, polymorphism, class hierarchy, method inheritance and concurrency are provided in ooRexx (ooRexx, 2023). ooRexx, the object-oriented paradigm of REXX, uses the tilde (~) as an explicit message operator. The programmer communicates with objects by sending them messages that name a method with potential options (or arguments). The receiving object looks for this method, invokes it on behalf of the programmer, and returns all the results that this method and its options may lead to. This explanation suffices to have students understand the concepts of insulation and inheritance. Without introduction, the object-oriented paradigm was already used in Figure 3, where the String object received the message *~strip("leading")*, which returned the string without leading spaces. Even with object-oriented programming, the ooRexx concepts manage to help beginners get started without unnecessary teaching time and cognitive load.

4 Conclusion

We consider cognitive load theory as a useful perspective to improve programming education and to choose an appropriate language. We see REXX's language characteristics to be the most important success factor in enabling students to learn productive programming quickly—within a few months—by minimizing unnecessary cognitive burden. These characteristics prevent troublesome errors and reduce the frustration associated with teaching and learning programming. Our experience has shown that students who have learned REXX subsequently learn other languages considered relevant by the industry, such as Visual Basic, Python, and especially Java, much more quickly and efficiently (Flatscher, 2023). We hope that this article will encourage future research on cognitive load in programming education and a consideration of REXX as an introductory language.

References

AmigaOS Manual: Arexx. (2023, May 27). In *AmigaOS Wikipedia.* https://wiki.amigaos.net/wiki/AmigaOS_Manual:_Arexx.

ANSI. (1996). *ANSI X3.274-1996—Programming Language REXX.* Retrieved from https://www.rexxla.org/rexxlang/standards/.

Becker, B. A., Glanville, G., Iwashima, R., McDonnell, C., Goslin, K., & Mooney, C. (2016). Effective compiler error message enhancement for novice programming students. *Computer Science Education*, *26*(2-3), 148-175.

BP-1. (2023, February, 16). *Business Programming 1.* Retrieved from https://wi.wu.ac.at/rgf/wu/lehre/autowin/material/foils/.

BP-2. (2023, February, 16). *Business Programming 2.* Retrieved from https://wi.wu.ac.at/rgf/wu/lehre/autojava/material/foils/.

BSF4ooRexx. (2023, May, 26). *Makes all of Java directly available to ooRexx and vice versa.*

Retrieved from https://sourceforge.net/projects/bsf4oorexx/.

Cowlishaw, M. (2022). *General Decimal Arithmetic*. Retrieved from https://speleotrove.com/decimal/.

Cowlishaw, M. (1987). The design of the REXX language. *ACM SIGPLAN Notices*, *22*(2), 26-35.

Data is Beautiful. (2019, October 7). *Most Popular Programming Languages 1965 – 2019* [Video file]. Youtube. https://www.youtube.com/watch?v=Og847HVwRSI.

Denny, P., Luxton-Reilly, A., Tempero, E., & Hendrickx, J. (2011, June). Understanding the syntax barrier for novices. In *Proceedings of the 16th annual joint conference on Innovation and technology in computer science education* (pp. 208-212).

Flatscher, R.G., (2023). Proposing ooRexx and BSF4ooRexx for Teaching Programming and Fundamental Programming Concepts. *In Proceedings of ISECON 2023*. Forthcoming. Plano.

Flatscher, R. G., & Müller, G. (2021). " Business Programming"–Critical Factors from Zero to Portable GUI Programming in Four Hours.

Fosdick, H. (2005). *Rexx programmer's reference*. John Wiley & Sons.

Garner, S. (2002). *Reducing the cognitive load on novice programmers* (pp. 578-583). Association for the Advancement of Computing in Education (AACE).

List of programming languages. (2023, May 27). In *Wikipedia*. https://en.wikipedia.org/wiki/List_of_programming_languages.

Mason, R., Cooper, G., & de Raadt, M. (2012, January). Trends in Introductory Programming Courses in Australian Universities–Languages, Environments and Pedagogy. In *Proceedings of the Fourteenth Australasian Computing Education Conference* (Vol. 123, pp. 33-42).

McIver, L., & Conway, D. (1996, January). Seven deadly sins of introductory programming language design. In *Proceedings 1996 International Conference Software Engineering: Education and Practice* (pp. 309-316). IEEE.

ooRexx. (2023, April 19). ooRexx (Open Object Rexx) Files. Retrieved from https://sourceforge.net/projects/oorexx/files/oorexx/.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, *38*(1), 63-71.

PYPL. (2023). *PYPL PopularitY of Programming Language*. Retrieved from https://pypl.github.io/PYPL.html.

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer science education*, *13*(2), 137-172.

Sands, P. (2019). Addressing cognitive load in the computer science classroom. *Acm Inroads*, *10*(1), 44-51.

Seik, A., (2023, February, 12). *ooRexx Plugin for IntelliJ IDEA and ooRexxDoc*. Retrieved from https://sourceforge.net/projects/bsf4oorexx/files/Sandbox/aseik/ooRexxIDEA/GA/2.1.0/.

Stachel, J., Marghitu, D., Brahim, T. B., Sims, R., Reynolds, L., & Czelusniak, V. (2013). Managing cognitive load in introductory programming courses: A cognitive aware scaffolding tool. *Journal of Integrated Design and Process Science*, *17*(1), 37-54.

Stefik, A., & Siebert, S. (2013). An empirical investigation into programming language syntax. *ACM Transactions on Computing Education (TOCE)*, *13*(4), 1-40.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12(2)*, 257-285.

Sweller, J., & Van Merriënboer, J. J. G. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *53*(3), 147-177.

WU (2023, February, 16). *Selected Seminar, Diploma, Bachelor and Master Theses.* Retrieved from https://wi.wu.ac.at/rgf/diplomarbeiten/.

Data Analysis of the Motivation and Factors for a Shorter Duration of Study

Bojan Radišić
Josip Juraj Strossmayer University of Osijek
Faculty of Tourism and Rural Development in Požega,
Vukovarska 17, 34000 Požega, Croatia
bradisic@ftrr.hr

Ivan Dunđer, Sanja Seljan
University of Zagreb
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
{idundjer, sseljan}@ffzg.unizg.hr

Abstract. *The duration of study is of interest to all stakeholders in education. This paper analyses the influence of motivation and other factors on the duration of study of students at professional studies in the field of economics at the Faculty of Tourism and Rural Development in Požega. In order to find out what former students considered important for finishing their studies, a survey was initially undertaken. In the second part of this research, real and objective data from the student information system called ISVU was analysed to find the correlations between the duration of study and selected data set variables that were extracted from ISVU.*

Keywords. Data analysis, data exploration, prediction, duration of study, motivation of students, correlation analysis, multiple linear regression

1 Introduction

Analysis of the motivational factors that influence duration of study is a point of interest of all stakeholders in the educational process: universities, students and employers. The duration of study affects both students and parents due to associated costs of studying, reflects the motivation of students and the quality of study, influences the employability and living standard, and directly impacts the satisfaction of all entities included in the educational process.

According to research conducted on a sample of 1,840 participants, some of the major weaknesses of the Croatian higher education system are the low percentage of study completion and significantly longer duration of study (Rimac, 2021). As stated in the mentioned research, there are many problems that affect the success of students or that influence the duration of studies. For instance, there are issues with adaptation to student life, difficulties with transitioning to a new and challenging level of education, and there are troubles that are related to personal circumstances.

Motivation is one of the very important factors that affects the success of students in terms of completing courses and studies on time. An important determinant of academic success is the motivation that energizes and directs behavior towards academic achievements (Wigfield et al., 2016).

Despite motivation being a subjective experience of students, many researchers in the field of education use EDM (Educational Data Mining) to explore the aspects of motivation and various factors, such as demographics, socioeconomics, and success in the study. This is done by applying statistical methods and machine learning algorithms for predicting the success or failure of students during their studies.

It is important to investigate and explain the different factors that affect the duration of study and the motivation of students. Factors can be perceived as subjective if they reflect the opinions of students collected through surveys. However, some other factors can also be perceived as objective, such as average grades, information on high school graduation success etc.

This paper is structured in the following way. After the Introduction section, a Literature review section presents relevant recent work, which is followed by a larger section that describes the conducted research methodology along with research results. Limitations and ideas for future work are shown in the fourth section, after which conclusions are drawn.

2 Literature review

A lot of research related to the success of students during their studies has been conducted with regard to different aspects, such as psychological (González et al., 2020), social and predictive elements (Xie et al., 2023), oftentimes based on the use of different statistical methods and machine learning algorithms.

Student motivation that is related to academic success and student well-being was recently explored by analysing a set of factors, such as self-efficacy, ability to self-regulate, engagement in learning, perceived learning value, reasons for studying, achievement motives, goals and adaptability (Kotera et al., 2022).

A research paper highlighted the importance of historical data in predicting student performance, as well as the relevance of background information such as numerical skills, family and motivation (Cavazos & Garza, 2018).

Another study analysed the influence of several factors on overall student performance. Factors that influenced student achievement were gender, application results and homework results (Liufu et al., 2020). All factors had a strong impact on students' final grades. The authors also concluded that promoting technological innovation at the university and strengthening the network of technology application opportunities with employers might help students to improve their learning outcomes.

A multiple linear regression model was used on data obtained from surveys conducted among engineering students in order to predict student performance (Kumar et al., 2020). The results showed that there were four independent variables that were significant for modelling academic performance of students: gender, relationship with their parents, secondary school success and first semester grades.

Research on a sample of 395 students concluded that the success of students is greatly influenced by specific factors, such as past failures with a negative correlation, implying that students are more likely to fail when they have a past failure class record (Beckham et al., 2023). Also, a negative correlation is observed in the case of age, which means that older students fail exams more often than their younger colleagues. However, positive correlations were also noticed, e.g. the mother's education and the student's motivation to continue their higher education.

On the other hand, there are also research papers that deal with the prediction of student dropouts, and thus provide an opportunity for educational institutions to identify students who are at risk of dropping out, and to take necessary countermeasures. For instance, according to an article that is based on a set of data collected from students, there are several factors that might eventually lead to the failure of a student in a course (Hegde & Prageeth, 2018), e.g. four previous course failures, more than three absences from classes, health problems, and the inability to adjust to student lifestyle.

A study on a sample of 13,696 students who studied at a Latin American university from the first semester of 2008 to the second semester of 2020 analysed various influences on student dropouts (Gutierrez-Pachas et al., 2023). The correlation analysis shows that the number of completed semesters has a strong negative correlation with dropouts in all cases. Furthermore, dropouts have a strong negative correlation with the final grade point average and the proportion of finished courses in relation to the total number of enrolled courses. Also, dropouts and the proportion of the number of absences in relation to the total number of course attendances have a moderate positive correlation.

3 Research methodology and results

This section presents results of two separate studies. Namely, the research is based on:

i) subjective data based on a survey conducted among former students who already finished their studies;
ii) objective data that was extracted from the so-called ISVU system (Information System of Higher Education Institutions).

Both studies ought to investigate the various factors affecting study duration.

The first study was conducted as an open-ended survey to examine the subjective factors that influence the duration of studies, according to the answers of students who have completed their studies. It was conducted among former students of the Faculty of Tourism and Rural Development in Požega, Josip Juraj Strossmayer University of Osijek. The survey was sent to 60 email addresses, among which 37 former students responded with answers, out of which 29 of them completed a three-year professional study in the field of economics. A total of 29 former students of economics were analysed in order to reveal the subjective factors that influence the duration of the study.

In the second study, student data from a student information system called ISVU – *Informacijski sustav visokih učilišta* (Information System of Higher Education Institutions) was extracted and examined in order to detect objective factors that affect the duration of study, to establish correlations with the duration of study, and to identify the predictors that mostly influence the duration of study by using multiple linear regression.

3.1 Study one: subjective data analysis

This study analyses the subjective motives that impact the desire to complete studies, and as such represent the personal perception of individual students. Motives highlighted by students who completed their studies within four years are examined in detail. The regular duration of study for professional studies is three years, however, in this research the authors additionally consider four years as regular.

The study was conducted on 29 former students of professional studies of economics. The average duration of study was 5.54 years according to internal Faculty reports. Fig. 1 shows the duration of study for students who participated in the survey.

76% of the surveyed students successfully completed their studies within 4 years, whereas 10% of them studied between five to six years. 7% of the students studied between four to five years, or more than six years.

According to the regulations of the University of Osijek, "a student has the status of a full-time student during the prescribed duration of studies, and at most for a period that is one third longer than the prescribed

duration of studies, i.e. until the end of the academic year in which that period expires".

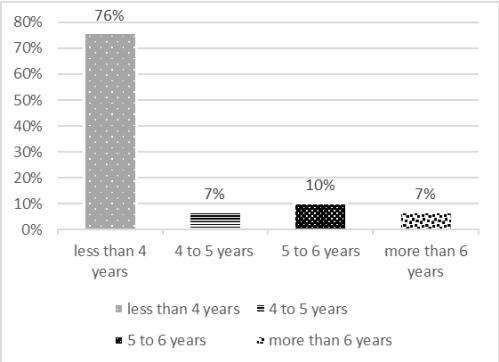


Figure 1. Duration of study

This means that full-time students lose their status after four years of studying, and that if they still want to continue their studies, they must change their status to part-time students.

The following figures in this section refer to students who completed their studies within four years.

Fig. 2 shows the age of the students at study enrolment. It presents only students whose duration of studies was shorter than four years. Most of the surveyed students (50%) were older than 30 years. Also, 41% were students under the age of 20. There were no students between 20 and 25 years old. It should also be noted that the study program is mostly enrolled by students after finishing high school in order to continue their education or by employees who have been employed for more than ten years.

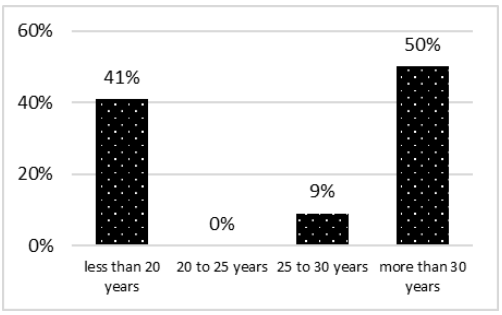


Figure 2. Age of students at study enrolment

Fig. 3 shows the ratio of full-time and part-time students who participated in the survey. Part-time students were a total of 55%, while full-time students made up 45% of all participants.

The largest number of students (59%) who completed their studies within four years achieved an average grade between 3.5 and 4.2, as shown in Fig. 4. An average of less than 3.5 was achieved by 27% of the surveyed students, whereas 14% of them achieved an average grade higher than 4.2.

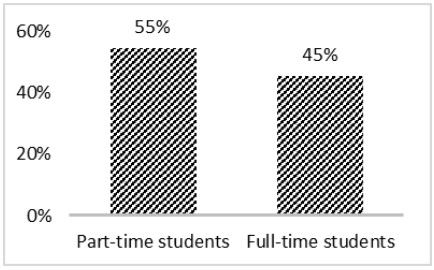


Figure 3. Type of study (enrolment type)

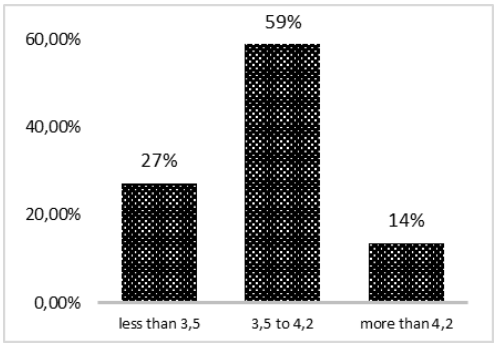


Figure 4. Average grade

Fig. 5 shows the distance from home to the place of study. Most students (64%) live within 30 km from the place of study, while 14% of all students live at a distance of more than 100 km from the place of study.

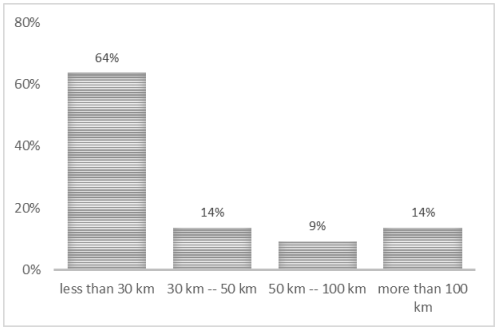


Figure 5. Distance from home

Fig. 6 shows the opinions of surveyed students who completed their studies within four years on what they considered to be important factors that influenced the total duration of their studies.

A total of 91% of the students agree or strongly agree that personal satisfaction/advancement is the most important factor that motivated them to complete their studies.

They also agree or strongly agree that continuing education (82% in total) and desire for employment (73%) are important factors for completing studies. This is to some extent expected, since after finishing high school, students have two choices: continuing their education or looking for a job.

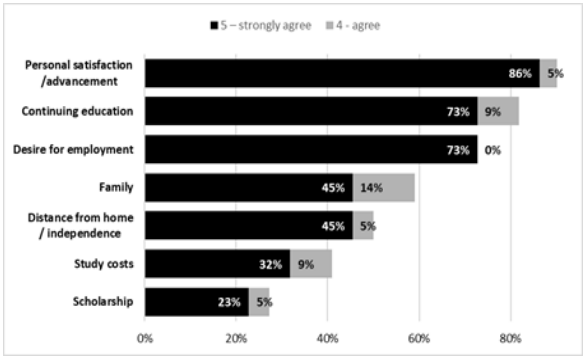


Figure 6. Motives of students for completing their studies

Most of the participants agree or strongly agree that family (59%) and distance from home (50%) are important factors that influence the successful completion of studies. Also, 50% of students who lived independently agree or strongly agree that distance from home is an important factor. The least number of respondents agree that the success of completing studies is influenced by study costs (41%) and scholarships (28%).

According to the observed data, students who finished their studies on time are mostly part-time students over 30 years old with an average grade between 3.5 and 4.2, and that study in a place lesss than 30 km from home. Motives that affect students to complete their professional studies of economics on time are: personal satisfaction/advancement (91%), need for continuing education (82%) and desire for employment (73%), family (59%) and distance from home (50%). All these are important factors that influence the successful completion of studies.

3.2 Study two: objective data analysis

According to the data extracted from the student information system called ISVU – *Informacijski sustav visokih učilišta* (Information System of Higher Education Institutions), 1,059 students enrolled the professional study in the field of economics at the Faculty of Tourism and Rural Development in Požega in the period from 2013 to 2018. The data was processed using two approaches: descriptive statistics and multiple linear regression.

The studies were successfully completed by 460 (43%) enrolled students. A total of 164 (36%) students completed their studies on time, i.e. within four years. The average duration of study was 5.39 years, whereas the median was 4.76 years.

3.2.1 Analysis of data from ISVU and discussion

Fig. 7 shows a comparison of the duration of study of full-time and part-time students.

It is noticeable that **full-time students** completed their studies on average in 4.56 years (median 4.16 years), which is almost two years earlier than part-time students – their average is 6.64 years (46% longer), with a median of 6.50 (56% longer).

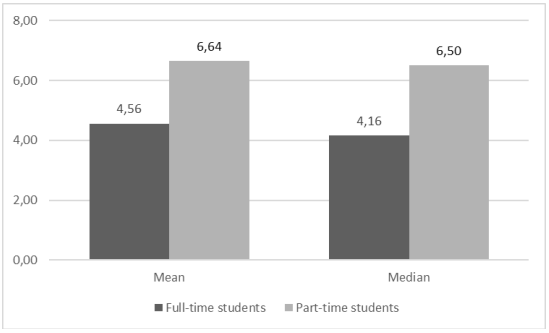


Figure 7. Average duration of study

It is reasonable to infer that full-time students wish to finish their studies earlier because they either plan to get a job and start working, or they plan to continue their education.

These two – continuing education and desire for employment – were also highlighted by the students in the survey as two of the three most important motivational factors for completing their studies.

Additionally, it might be speculated that part-time students are more driven by a desire for personal satisfaction/advancement to complete their studies. Part-time students, by definition, should already be employed, and therefore should be motivated to finish their studies for personal satisfaction or possibly for advancement at work. However, this should be analysed more carefully in future research for more precise results and conclusions.

The following figures in this section refer to students who completed their studies within four years.

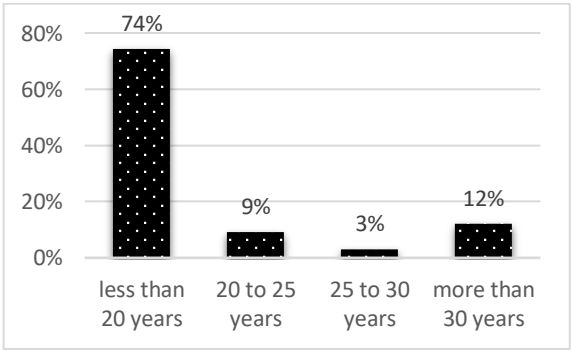


Figure 8. Age of students at study enrolment

Fig. 8 shows the age of the students at study enrolment. Most of the students (74%) were younger than 20 years. According to ISVU data, it should also be noted that this particular study program is mostly enrolled by students after finishing high school in order to continue their education.

Fig. 9 shows the ratio of full-time and part-time students. Part-time students were a total of 18%, while

full-time students made up 72%, according to ISVU data.

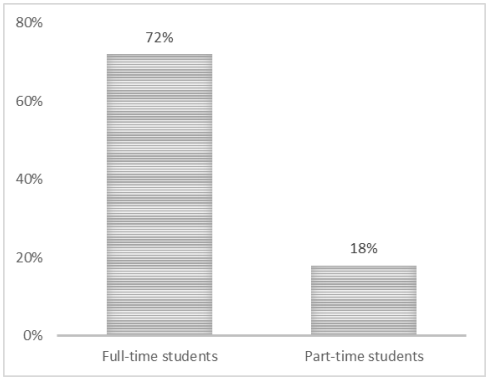


Figure 9. Type of study (enrolment type)

The largest number of students (63%) who completed their studies on time achieved an average grade less than 3.5. An average grade between 3.5 and 4.2 is achieved by 29% of the students, as shown in Fig. 10. In addition, only 8% of them achieved an average grade higher than 4.2.

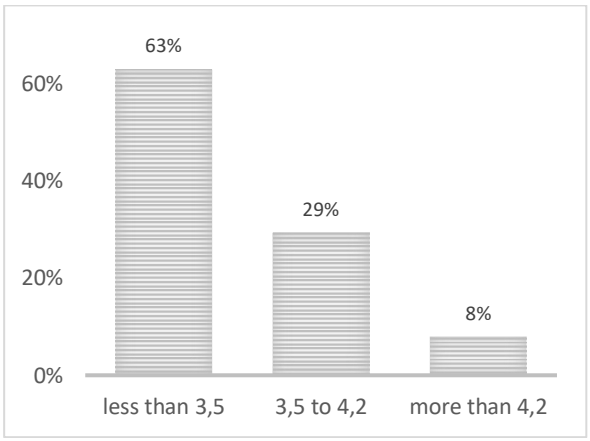


Figure 10. Average grade

Among the students who finished their studies on time, almost 60% of them were students from other counties, whereas 40% were from the Požega-Slavonia County, as shown in Fig. 11.

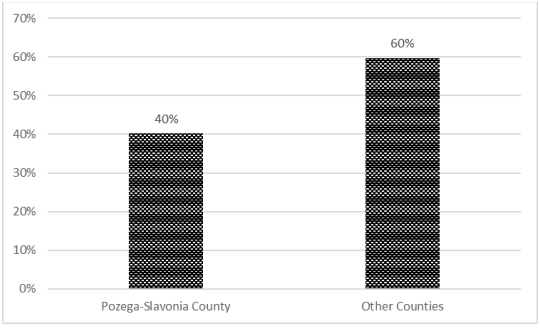


Figure 11. County residence

In comparison, according to the initial survey, most of the students (64%) live within 30 km from the place of study.

According to the data extraxted from ISVU, mostly full-time students younger than 20 years old with an average grade less than 3.5 from outside the Požega-Slavonia County are more motivated to complete their studies within four years.

3.2.2 Multiple linear regression and discussion

Fifteen variables were selected from the ISVU database of the Faculty of Tourism and Rural Development in Požega: duration of study, age at the time of enrolment, grade point average (average grade achieved by students during their studies), type of study (part-time or full-time), type of finished high school (such as economy and trade, public transport, electrical engineering, gymnasium, agriculture, mechanical engineering, catering and tourism, musical art, construction, geodesy and building materials, mathematics, physics and computer sciences, woodworking, leather working, maritime, river and port transport, food, textiles), Economics_in_out (finishing a high school of economics or not), counties of Croatia, PSZ_in_out (residence in Požega-Slavonia County or not), gender (male or female), high school score, state exam score, total score (high school score plus state exam score), English language level B score, Croatian language level B score and mathematics level B score. In this case, only data from students who completed their studies were considered (460 of them, i.e. 43% out of all students from ISVU).

Correlations between the duration of study (the dependent variable in this research) and all other mentioned variables (also known as predictors or predictor variables) were calculated with the help of SPSS.

Table 1 shows only variables that have a significant correlation with the duration of study for all graduated students. There is a medium Pearson correlation (Cohen, 1998) at the 0.01 (1%) significance level between the duration of study and the type of study (-0.44), total score (high school score plus state exam score) (-0.44), high school score (-0.44), state exam score (-0.43), English language level B score (-0.40), Croatian language level B score (-0.41) and mathematics level B score (-0.43).

A low Pearson correlation at the 0.01 significance level between the duration of study and gender (-0.22), grade point average (-0.18) and finishing high school of economics or not (-0.15). Also, there is a low Pearson correlation at the 0.05 significance level between the duration of study and the type of finished high school (0.12). The negative values for the variables that reflect high school grades and graduation success, as well as the grade point average during study, show that students with better grades complete their studies earlier.

Table 1. Correlation between the duration of study and other predictors

	Duration of study		
	Pearson Correlation	Sig. (2-tailed)	N
Part-time or full-time students	-,436**	,000	455
Gender	-,219**	,000	455
Grade point average	-,182**	,000	455
Type of finished high school	,120*	,010	455
Finished or not High School of Economics	-,149**	,001	455
High school score	-,435**	,000	455
Croatian language level B score	-,405**	,000	452
English language level B score	-,400**	,000	412
Mathematics level B score	-,425**	,000	452
State Exam score	-,426**	,000	455
Sum of high school score and state exam score	-,439**	,000	455

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The multiple linear regression model shows the linear relationship between the independent variables and the dependent variable. It considers the effect of multiple independent variables (i.e. predictor variables) on a dependent variable, and determines a quantitative relationship between them (Vinodhini & Chandrasekaran, 2018). The multiple linear regression equation is shown in (1):

$$Y = k_0 + k_1 x_1 + k_2 x_2 + \cdots + k_n x_n \quad (1)$$

where
Y=dependent variable,
x_i=predictors,
k_i=slope coefficient for every predictor,
k_0=intercept (the value of Y when all predictors are zero).

Due to excessive multicollinearity, the following predictors were removed from the model: high school score, state exam score, English language level B score, Croatian language level B score and mathematics level B score. Therefore, only the total score (high school score plus state exam score) is considered.

Table 2 shows a comparison of two multiple linear regression models. The R-squared value of both models is ca. 0.32, which indicates that the model explains 32% of the variation in the duration of study.

The adjusted R-squared value of ca. 0.31 is insignificantly lower than the R-squared value. Higher R-squared values can be obtained by simply increasing the number of predictors in the model, so an adjusted R-squared is used to prevent this. Using the adjusted R-squared value can avoid overfitting the value, as it takes into account the number of predictors and penalizes models with a large number of predictors.

Table 2. Comparison of models

Model Summary[c]			
Model		1	2
R		,563[a]	,568[b]
R Square		,317	,323
Adjusted R Square		,309	,309
Std. Error of the Estimate		1,8458079	1,8455635
Change Statistics	R Square Change	,317	,006
	F Change	41,620	1,030
	df1	5	4
	df2	449	445
	Sig. F Change	,000	,391

a. Predictors: (Constant), Part-time or full-time students, Type of finished high school, Gender, Age, Sum of high school score and state exam score

b. Predictors: (Constant), Part-time or full-time students, Type of finished high school, Gender, Age, Sum of high school score and state exam score, Croatian Counties, Grade point average, Living or not in Pozega Slavonia County, Finished or not High School of Economics

c. Dependent Variable: Duration of study

The first model consists of five predictors: total score (high school score plus state exam score), type of finished high school, gender, age at the time of enrolment, and type of study (part-time or full-time).

The R-squared change score shows the improvement in R-square when a new predictor is added. In this very case, for the second model it is only ca. 0.01 which is almost a negligible improvement over the first model. This allows the remaining predictors to be ignored: counties of Croatia, grade point average, PSZ_in_out (residence in Požega-Slavonia County or not), Economics_in_out (finishing high school of economics or not).

The most effective multiple linear regression model is established between the dependent variable "duration of study" and the predictive variables: total score (high school score plus state exam score), type of finished high school, gender, age at the time of enrolment, type of study (part-time or full-time), as presented in the first model in Table 2. Overall, this means that the duration of study is influenced by age, gender, total score (high school score plus state exam score), type of finished high school, age at the time of enrolment and type of study (part-time or full-time).

Table 3. Correlation between the duration of study and other predictors

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	13,019	,798		16,310	,000
	Part-time or full-time students	-1,884	,318	-,410	-5,921	,000
	Age	-,121	,018	-,328	-6,730	,000
	Gender	-,854	,219	-,153	-3,898	,000
	Type of finished high school	,049	,024	,081	2,065	,039
	Sum of high school score and state exam score	-,002	,001	-,254	-3,984	,000

a. Dependent Variable: Duration of study

Table 3 shows the contribution of the five optimal predictors to the duration of study. The beta score in the column of standardized coefficients shows the influence of the predictor on the dependent variable when the variance explained by all other predictors in the model is subtracted.

The highest absolute value is scored by the type of study (part-time or full-time) which has the greatest influence of -0.41. This means that if the predictor that reflects the type of study (part-time or full-time) increases by one standard deviation, and other predictors remain constant, then the duration of study will decrease by 0.41 of standard deviations.

The column "Sig." shows that all five predictors make a statistically significant unique contribution (<0.05) to the final value of the duration of study. Unstandardized coefficients in column B are used to describe the model. The first number represents a constant, whereas other values represent the slope coefficient for each predictor. Using equation (1), the multiple linear regression model in this case is:

$$Y = 13.019 - 1.884 \cdot \text{type of study} - 0.121 \cdot \text{age at the time of enrolment} - 0.854 \cdot \text{gender} + 0.049 \cdot \text{type of finished high school} - 0.002 \cdot \text{total score}.$$

It can be noticed that the type of study, age and total score (high school score plus state exam score) have the greatest influence on the duration of study. The model implies that older full-time female students with better scores in high school and final exams will complete their studies in a shorter period of time.

Table 4 shows only variables that have a significant correlation with the duration of study for students who completed their studies within four years.

Table 4. Correlation between the duration of study and other predictors (within four years)

Correlations

		Part-time or full-time students	Age at the time of enrolment	Gender	Grade point average	Finished or not High School of Economics	High school score	Sum of high school score and state exam score
Duration of study	Pearson Correlation	,276**	-,270**	-,171*	-,179*	-,209**	,174*	,162*
	Sig. (2-tailed)	,000	,000	,028	,022	,007	,026	,038
	N	164	164	164	164	164	164	164

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

There is a low Pearson correlation at the 0.01 significance level between the duration of study and type of study (0.28), type of finished high school (-0.21) and age at the time of enrolment (-0.27). There is also a low Pearson correlation at the 0.05 significance level between the duration of study and gender (-0.17), grade point average (-0.18), high school score (0.17), and total score (high school score plus state exam score) (0.16). It can be concluded that the age at the time of enrolment and the type of study can be highlighted as the variables that have the greatest influence on the duration of studies (within four years).

Table 5 shows only variables that have a significant correlation with the duration of study for students who completed studies in more than four years (prolonged study duration).

Table 5. Correlation between the duration of study and other predictors (more than four years)

Correlations

		Part-time or full-time students	Gender	High school score	Croatian language level B score	English language level B score	Mathematics level B score	State Exam score	Sum of high school score and state exam score
Duration of study	Pearson Correlation	-,420**	-,178*	-,427**	-,382**	-,403**	-,401**	-,399**	-,421**
	Sig. (2-tailed)	,000	,002	,000	,000	,000	,000	,000	,000
	N	296	296	296	294	275	294	296	296

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

There is a low Pearson correlation at the 0.01 significance level between the duration of study and type of study (-0.42), total score (high school score plus state exam score) (-0.42), high school score (-0.43), state exam score (-0.40), English language level B score (-0.40), Croatian language level B score (-0.38) and mathematics level B score (-0.40).

A low Pearson correlation at the 0.01 significance level between the duration of study and gender (-0.18) is also observed. Therefore, it is evident that the high school score, the total score (high school score plus state exam score) and the type of study can be highlighted as the variables that have the greatest influence on the duration of studies in the context of studying for more than four years.

By comparing the previous two tables, some differences and similarities can be observed. The type of study (part-time or full-time) has a positive correlation in the case of students who completed their studies within four years. On the other hand, students who study longer than four years have a negative correlation for the type of study.

It means that full-time students who study longer than four years finish their studies earlier, although not within four years. In both cases, female students finish their studies earlier than their male colleagues, regardless of the duration of study.

4 Limitations and future research

The main limitation of this research is the small number of students who participated in the survey, and especially the small number of students who studied for longer than four years.

Therefore, it was not possible to make a comparison with students who finished their studies earlier. For future research the authors plan to consider the inclusion of students from similar institutions and from other geographical areas so that comparisons on multiple levels can be made.

5 Conclusion

The aim of this research was to analyse data related to the influence of various motivational factors and to predict optimal factors that impact the duration of study in the case of a professional 3-year study program.

The research was conducted on students of professional studies in the field of economics, at the Faculty of Tourism and Rural Development in Požega. It was based on two studies: i) on subjective data (survey) among 29 former students who already finished the three-year professional study in economics, ii) on objective data on 460 students who finished the professional study of economics in the period from 2013-2018 – extracted from ISVU (Information System of Higher Education Institutions).

Results of the first study show that students who finish their studies on time, agree or strongly agree that continuing education (82%), desire for employment (73%) and personal satisfaction/advancement (91%) are very important motivational factors for completing their studies. Also, family (59%) and distance from home (50%) are highlighted as well.

Results of the second study show that full-time students younger than 20 years, with an average grade less than 3.5 from outside of Požega-Slavonia County, are more motivated to complete their studies on time, according to ISVU data.

There is a significant difference between students residing in the Požega-Slavonia County and those outside it, which directly affects the costs of studies and living. This is a clear important motivational factor for students from outside the county to complete studies on time.

When analyzing correlations between the duration of study and fifteen variables, the results show a significant correlation at the 1% significance level for factors that influence the duration of study for the following variables: type of study (part-time or full-time), total score (high school score plus state exam score), English language level B, Croatian language level B and mathematics level B.

The multiple linear regression model discovers five optimal factors for predicting the duration of study. Type of study (full-time or part-time), age and total score (high school score plus state exam score) have the greatest impact, followed by the type of finished high school and gender.

The multiple regression model also predicted that older female full-time students with better grades from high school and final exams will finish their study in a shorter period of time.

References

Beckham, N. R., Akeh, L. J., Pratama Mitaart, G. N., & Moniaga, J. V. (2023). Determining factors that affect student performance using various machine learning methods. *Procedia Computer Science*, 216, 597–603. doi:10.1016/j.procs.2022.12.174

Cavazos, R., & Garza, S. E. (2018). Learning Models for Student Performance Prediction. *Lecture Notes in Computer Science*, 171–182. doi:10.1007/978-3-030-02840-4_14

Cohen, J. W. (1988). *Statistical power analysis for the behavioural science* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

González Olivares, Á. L., Navarro, Ó., Sánchez-Verdejo, F. J., & Muelas, Á. (2020). Psychological Well-Being and Intrinsic Motivation: Relationship in Students Who Begin University Studies at the School of Education in Ciudad Real. *Frontiers in Psychology*, 11. doi:10.3389/fpsyg.2020.02054

Gutierrez-Pachas, D. A., Garcia-Zanabria, G., Cuadros-Vargas, E., Camara-Chavez, G., & Gomez-Nieto, E. (2023). Supporting Decision-Making Process on Higher Education Dropout by Analyzing Academic, Socioeconomic, and Equity Factors through Machine Learning and Survival Analysis Methods in the Latin American Context. *Education Sciences*, 13 (2), 154, doi:10.3390/educsci13020154

Hegde, V., & Prageeth, P. P. (2018). Higher education student dropout prediction and analysis through educational data mining. *Proceedings of the 2nd International Conference on Inventive Systems and Control (ICISC)*. doi:10.1109/icisc.2018.8398887

Kotera, Y., Aledeh, M., Barnes, K., Rushforth, A., Adam, H., & Riswani, R. (2022). Academic Motivation of Indonesian University Students: Relationship with Self-Compassion and Resilience. *Healthcare*, 10(10), 2092. doi:10.3390/healthcare10102092

Kumar, A., Eldhose, K. K., Sridharan, R., & Panicker, V. V. (2020). Students' Academic Performance Prediction using Regression: A Case Study. *Proceedings of the International Conference on System, Computation, Automation and Networking (ICSCAN)*. doi:10.1109/icscan49426.2020.9262346

Liufu, Y., Meng, X., & Wang, J. (2020). Analysis and Suggestions on Influencing Factors of Student Achievement under Multiple Linear Regression. *Proceedings of the 12th International Conference on Education Technology and Computers*. doi:10.1145/3436756.3437053

Rimac, I. (2021). Istraživačko izvješće projekta EUROSTUDENT VII, Ministarstvo znanosti i obrazovanja Republike Hrvatske. Retrieved from: https://www.eurostudent.hr/userfiles/files/Istrazivanje-Eurostudent-VII-(5A)_FINAL.pdf

Vinodhini, A., S., & Chandrasekaran, R. M. (2018). Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers. *Journal of Computer Science*, 14(5), 654–662. doi:10.3844/jcssp.2018.654.662Wigfield, A., Tonks, S., & Klauda, S. L. (2016). Expectancy-value theory. In K. R. Wentzel & D. B. Miele (Eds.), Handbook of Motivation in School, 2nd Edition (pp. 55–74). New York, NY: Routledge.

Xie, M., King, R. B. & Luo, Y. (2023). Social motivation and deep approaches to learning: a nationwide study among Chinese college students. Higher Education, 85, 669–687. doi:10.1007/s10734-022-00860-6

Innovative learning design and teachers' competencies: prerequisites for meaningful use of digital material in schools

Melita Milić, Blaženka Divjak
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin
{mzahtila, blazenka.divjak}@foi.hr

Abstract. *This study provides an in-depth analysis of the benefits and risks associated with the adoption of digital materials in schools. By conducting a systematic literature review and utilizing an AI tool, the study also explores the correlation between teachers' digital and pedagogical competencies and the use of these materials. The findings underscore the necessity of continuous teacher training to effectively implement innovative teaching and learning approaches that incorporate well-designed instruction and meaningful integration of digital resources. The utilization of digital materials fosters active student engagement and positively influences learning outcomes. However, it is important to consider potential challenges such as technical issues, discipline management, and waning student enthusiasm over time.*

Keywords. Digital textbook, e-textbook, open educational material, teachers' competences, SLR, learning

1 Introduction

The rapid advancement of technology has required education to adapt and support student learning and progress, both within and outside the traditional school premises. To facilitate this, teachers have increasingly embraced the use of textbooks supplemented with digital materials, such as animations, interactive simulations, videos, audio recordings, educational apps, presentations, and other forms of digital content intended to support learning and teaching, aiming for efficient delivery of relevant content and the achievement of intended learning outcomes.

A digital textbook is a specific type of digital resource that replicates the content and structure of a traditional textbook (sequence of topics, similar to their print counterparts), while digital materials encompass a broader range of digital resources used for educational purposes.

Digital materials are often used to enhance instruction, provide additional learning resources, and facilitate interactive and engaging learning experiences. Unlike digital textbooks, digital materials are not limited to replicating the structure and content of traditional textbooks but can encompass a wide range of multimedia and interactive resources.

Digital textbooks offer advantages over their traditional counterparts by incorporating multimedia, animations, and virtual reality, enabling faster and more engaging content delivery (Kim & Yu, 2019; Wang & Xing, 2019).

While the existing literature highlights the benefits associated with the use of digital materials in education, it is important to acknowledge the obstacles that impede their widespread adoption and usage in schools (Vorotnykova, 2019). These obstacles encompass various factors such as the availability of IT infrastructure, the provision of adequate support services, and even health concerns like the impact of e-books on users' vision (Masango et al., 2020).

In this paper, we aim to investigate the relationship between teachers' digital competences and the utilization of digital textbooks and other digital materials in schools. To achieve this, we employ a classical systematic literature review (SLR) enriched by the application of an artificial intelligence (AI) tool. The subsequent sections of this paper provide detailed descriptions of the research methodology, research questions, analysis of the research findings, discussion of the results, and a conclusion. Additionally, we propose avenues for future research in this field.

2 Research methodology and research questions

This paper addressed three research questions.

RQ1: What are the characteristics associated with the use of digital materials in schools?

The characteristics under investigation include the frequency of usage (occasional-continuous), the location of usage (school-based or independent study at home), the medium through which they are accessed (computer, tablet, mobile phone), their alignment with innovative learning and teaching methods/strategies (e.g., flipped classroom, project-based learning, problem-solving, inquiry-based learning), the publisher of the digital materials (commercial or public entity), and the payment method for accessing the digital materials (independent purchase or combined with a physical textbook).

RQ2: How does the use of digital materials for teaching and learning (TL) in schools relate to teachers' digital competences?

RQ3: Is there a relationship between the use of digital materials for TL and the maturity level of the school or educational system?

To address the research questions mentioned earlier, a systematic literature review (SLR) methodology was employed. Two databases, Web of Science (WoS) and Scopus, were selected to conduct comprehensive literature searches. Additionally, an AI-based tool called Research Rabbit (RR) was utilized to enhance the literature research process. RR is a text-based research recommender publicly available since 2021. Its purpose is to support researchers in their literature exploration without the need to switch between search and database modes. Using RR, researchers can input initial literature items, and the tool guides them through networks of related sources. The interface of RR facilitates the mapping of literature in a specific area by presenting lists and graphs of interconnected papers. These lists and graphs are dynamically adjusted in real time based on the researcher's selection (Krzton, 2023). The RR tool was used in this study to explore broader areas where digital textbooks have been discussed, while also evaluating the effectiveness of AI tools in literature research.

The databases were accessed on October 6, 2022, and April 6, 2023, with a search conducted across all fields using three main topics: digital maturity ("digital readiness" OR "digital maturity"), school (school* OR "pre-tertiary education"), and digital textbooks ("digital textbook*" OR "electronic textbook*" OR "open digital material*" OR "open educational resources" OR "open educational content" OR "open educational material*").

Unfortunately, no articles were found that matched the specified query in any of the databases. Even the query combining the topics of digital maturity and digital textbooks did not yield any results. However, an article by (Milić & Divjak, 2022) covered the query with the keywords "digital maturity" and school*.

Given these findings, the search was continued without the inclusion of the digital maturity keyword. The final query focused on the topic of school and digital textbooks, using the following query: (school* OR "pre-tertiary education") AND ("digital textbook*" OR "electronic textbook*" OR "open digital material*" OR "open educational resources" OR "open educational content" OR "open educational material*"). While RR did not identify any papers matching this query, numerous relevant articles were found in the WoS and Scopus databases. To refine the search, the results were narrowed down to articles written in English within the last ten years, following the approach outlined by (Divjak et al., 2022). The various stages of this search process are depicted in Figure 1.

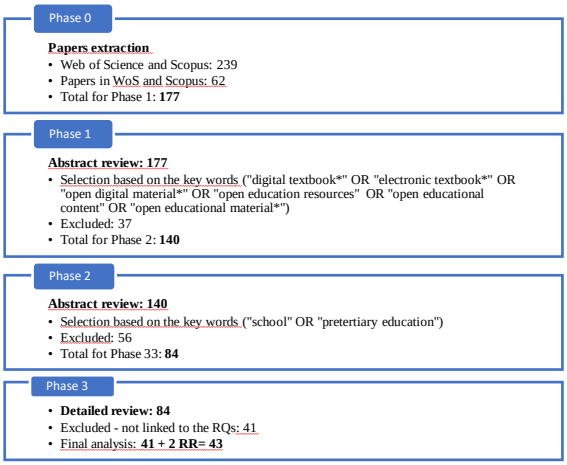


Figure 1. Methodology of the SLR research

A comprehensive search of the WoS and Scopus databases, considering the specified keywords and limitations, resulted in the identification of 239 scientific articles. Among these, 62 articles were found in both databases, leaving a total of 177 unique scientific articles for Phase 1 of the analysis.

During Phase 1, the abstracts of the articles were examined to identify those directly related to the topic of digital textbooks. Out of the initial set, 37 articles were unrelated to the specified keywords, narrowing down the selection to 140 articles for Phase 2. In Phase 2, the abstracts of the remaining articles were further scrutinized to determine their relevance to the keywords "school" or "pre-tertiary education." The majority of the 56 articles that met this criterion focused on the use of digital materials in higher education, which led to their exclusion from further analysis. As a result, 84 articles remained for thorough examination in Phase 3.

In Phase 3, a detailed review and analysis were conducted on the 84 selected articles. Upon closer inspection, 41 articles were excluded from further analysis as they did not provide answers to any of the research questions, placed excessive emphasis on the technology itself, or lacked appropriate research design. Consequently, a total of 43 papers were identified as suitable for in-depth discussion and analysis. Table 1 provides a classification of these papers based on the country where the studies were conducted. Notably, the highest number of papers originated from the USA (8), followed by China and

Indonesia (4). Generally, a significant portion of the studies (15) were conducted in Asian countries.

Table 1. List papers for discussion (additional 2 papers from RR are marked with the asterisk)

USA	(Bouck et al.,2016), (Bruhn & Hasselbring, 2013), (Chen, 2017), (Fouh et al., 2014) *, (Gerard et al., 2022), (Roskos & Neuman, 2014), (Richardson & Roebuck Sakho, 2022), (Lee et al., n.d.)
China	(Shen et al., 2015), (Sun & Jiang, 2015), (Wang & Xing, 2019), (Wijaya et al., 2022)
Indonesia	(Farisi, 2014), (Hermita et al., 2023), (Pratama & Firmansyah, 2021), (Sinaga et al., 2022)
Hong Kong	(Chiu, 2017), (Lo et al., 2021), (Lo et al., 2022) *, (Wong et al., 2016)
Korea	(Joo et al., 2017), (Lim et al., 2022), (So et al., 2015)
Germany	(Reinhold et al., 2020), (Reinhold et al., 2021), (Rezat 2021)
Uganda	(Busulwa & Bbuye, 2018), (Kabugo, n.d.)
Morocco	(Moundy et al., 2021), (Moundy et al., 2022)
Sweden	(Grönlund et al., 2018), (Håkansson Lindqvist, 2019)
South Africa	(Masango et al., 2020)
Australia	(Crook & Sharma, n.d.)
Kazakhstan	(Prutko et al., 2020)
Serbia	(Radović et al., 2020)
Portugal	(Lucas, 2020)
Ukraine	(Vorotnykova, 2019)
Taiwan	(Shadiev et al., 2015)
Slovakia	(Zacok et al., 2019)
Tunisia	(Tlili et al., 2022)
Spain	(Grimalt-Álvaro et al., 2019)
Lithuania	(Dudaitė & Prakapas, 2019)

Furthermore, to gain additional insights into the selected papers, a network analysis was conducted using the Gephi tool (Figure 2). All 43 papers, along with their citations, were imported into Gephi for visualization. The network analysis revealed the most influential papers based on the number of citations and links to other papers. Notably, three papers emerged as particularly influential: (Joo et al., 2017), (Shadiev et al., 2015), and (Chiu, 2017). These papers were published in reputable scientific journals and employed survey or experimental approaches in their research. (Joo et al., 2017) focused on Korean middle school students enrolled in an English class that utilized a digital textbook. The study investigated the structural relationships among students' expectations, perceived enjoyment, perceived usefulness, and satisfaction when using digital textbooks. Their findings shed light on the factors influencing students' experiences with digital textbooks. In the study by (Shadiev et al., 2015), an experimental approach was employed to examine the effects of mobile systems on learning achievement and cognitive load among students in Taiwan. The experimental group, which used digital textbooks and mobile systems, exhibited better performance on tests compared to the control group. Furthermore, the use of technology resulted in reduced cognitive load for the students. A survey conducted by (Chiu, 2017) involved 306 secondary school teachers in Hong Kong. The study explored the acceptance of digital textbooks and identified anxiety as a significant barrier and positive attitude as a key catalyst for their acceptance among teachers.

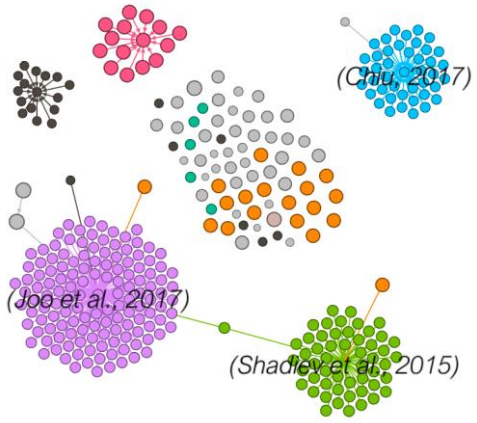


Figure 2. Analyzed papers along with their citations, displayed in the Gephi tool

Out of the 41 papers uploaded to the RR tool, a substantial number of similar papers (15,448) were identified. Further filtering excluded 48 studies published before 2012 and 7 studies published after October 6, 2022. Additionally, 142 studies authored by researchers whose articles were uploaded to RR, as well as 253 suggested articles by other authors, were found. Consequently, a total of 450 articles underwent a two-step selection process. Through searching the titles, 43 articles were initially identified, and by reading the abstracts, 12 papers remained eligible for a detailed analysis. It is worth noting that while RR suggested numerous articles, there was a significant overlap between the authors of the uploaded papers and the suggested authors. Moreover, several articles were excluded due to their focus on entirely different research domains, such as psychology or medicine.

Following a meticulous analysis, only two additional papers, (Fouh et al., 2014) and (Lo et al., 2022), were deemed relevant and were included alongside the articles identified through the classical systematic literature review. Consequently, the final literature analysis comprised a total of 43 articles.

3 Results

RQ 1. Characteristics of using digital materials in school

A comprehensive analysis of the selected articles revealed that 88% of them addressed aspects related to RQ 1, and their findings were categorized based on the characteristics specified in the research question. While the price of digital materials is recognized as an important factor, teachers also prioritize other considerations. For instance, studies conducted by (Wong et al., 2016), (Busulwa & Bbuye, 2018), and (Lo et al., 2021) emphasized that teachers primarily value the quality and availability of the materials, followed by the necessary hardware and software, with price being of lesser concern. On the other hand, (Vorotnykova, 2019) emphasized the importance of digital materials being freely accessible to students.

Interestingly, over 80% of the articles did not discuss the price of digital materials or the publisher. However, the available data indicated that the initiative for digital materials primarily originated from national or regional authorities. Among the analyzed articles, seven highlighted that digital materials were provided through private initiatives or private publishers, while nine articles described the significant influence of entities such as ministries of education, public open educational resource (OER) initiatives, UNICEF, and K-12 foundations in the creation and availability of digital materials. Only five papers specifically addressed the use of open educational content, while the remaining articles predominantly focused on digital textbooks.

It is worth noting that two-thirds of the analyzed articles were published after the onset of the Covid-19 pandemic. As anticipated, teachers increasingly relied on OER during the pandemic compared to before (Lo et al., 2022), which aligns with the findings of (Divjak et al., 2022).

Furthermore, findings from (Pratama & Firmansyah, 2021) and (Chiu, 2017) highlighted that students who are enthusiastic about adopting the latest mobile technology tend to prefer digital content. Additionally, (Shadiev et al., 2015) emphasized the positive impact of multimedia tools in enabling students to effectively utilize authentic contextual resources for learning (e.g. language learning).

In the majority of the analyzed articles, it was observed that students enjoy working with interactive teaching aids in the classroom and find such activities interesting and engaging. Moreover, (Wang & Xing, 2019) pointed out that students perceive digital materials as easy to use. Teachers also noted various benefits associated with the use of digital materials, such as promoting active student participation, offering practical and effective learning experiences, fostering increased student motivation to learn, and leading to improved learning outcomes compared to traditional teaching methods. These findings were supported by studies conducted by (Moundy et al., 2022), (Radović et al., 2020), (Chiu, 2017), (Lim et al., 2022), and (Dudaitė & Prakapas, 2019).

Furthermore, it was found that increased student engagement when using digital materials can result in better academic performance, as highlighted by the research of (Moundy et al., 2021) and (So et al., 2015).

On the other hand, (Fouh et al., 2014) pointed out that the limited adoption of mobile devices in the classroom is attributed to device-specific characteristics and usability issues, such as small screens. Teachers also acknowledged several challenges and negative aspects associated with the use of digital materials. For instance, difficulties in maintaining student discipline, observed a decline in students' interest and enthusiasm for using interactive tools as the school year progresses, and encountered occasional technical issues (Dudaitė & Prakapas, 2019). Additionally, older studies pointed out that inadequate technology infrastructure, characterized by slow response times and frequent technical difficulties, may discourage students from fully embracing digital textbooks (Shen et al., 2015). Furthermore, a three-year study conducted by (Lim et al., 2022) emphasized the importance of continuously updating the user interface and user experience of digital textbooks to meet students' preferences and demands. This ensures that students perceive digital textbooks as attractive and engaging learning tools. Figure 3 presents a SWOT analysis of the use of digital textbooks in schools, based on the insights gathered from the analyzed literature.

Regarding the medium of digital materials used in schools, it is challenging to make a definitive assessment since almost half of the analyzed articles did not provide specific information on this aspect. Furthermore, it is important to note that drawing conclusions about the current situation in a particular country based solely on research articles may not be comprehensive, and it is necessary to examine available international and national reports. However, it is evident that the use of digital devices has increased in the aftermath of the Covid-19 pandemic (OECD, 2022). This observation is supported by the analyzed articles, as only one article mentioning the use of tablets was published prior to the pandemic, while the rest were published after its onset.

Interestingly, most teachers expressed concerns about utilizing mobile phones for learning and teaching purposes (Busulwa & Bbuye, 2018). A similar finding was reported in a study conducted by (So et al., 2015), which revealed that teachers tend to be more conservative compared to students and parents when it comes to implementing mobile devices in classrooms. However, it is important to note that these studies were conducted before the pandemic, and the current situation regarding the use of mobile devices may have evolved.

While the frequency of digital materials usage is not explicitly defined in over half of the analyzed articles, some insights can still be gleaned.

Approximately 37% of the articles discuss the continuous use of digital materials in schools, highlighting their consistent integration into the learning process. On the other hand, occasional use of digital materials is mentioned in 11% of the articles, typically for activities related to the learning of new content.

The majority of the analyzed articles (62%) were published after the onset of the Covid-19 pandemic (Figure 3), indicating a significant shift in the use of digital devices in schools in line with (OECD, 2022). It was found that 45% of the articles explicitly mentioned the use of devices in schools, while no information was provided in 45% of the articles. This data is particularly noteworthy because studies such as (Lo et al., 2021) and (Moundy et al., 2022) reported increased utilization of digital devices by students for remote learning, especially during the Covid-19 period. These studies also highlighted the use of flipped classroom approaches for teaching and learning, which were only mentioned in a limited number of articles (compared to 45% of articles that explicitly stated the absence of flipped classroom usage).

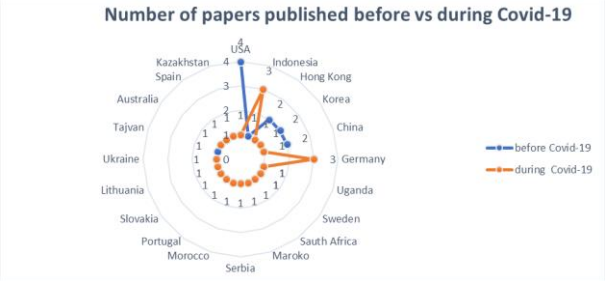


Figure 3. Number of papers published before and during pandemic caused by Covid-19

Furthermore, research by (Joo et al., 2017) revealed that students with high expectations regarding the usefulness of digital materials tend to be more satisfied users. However, this does not necessarily guarantee their continued usage. These findings align with the research conducted by (Reinhold et al., 2020), which emphasized the connection between students' active engagement in classroom activities and their academic success, regardless of the type of textbook utilized.
While many teachers considered the use of digital materials as a crucial educational practice (Roskos & Neuman, 2014), it was noted by (Crook & Sharma, n.d.) that teachers still tended to employ teacher-centered approaches in their instruction, irrespective of whether digital technologies were incorporated or not.

RQ 2. Use of digital materials for TL in schools related to the digital competences of teachers

Regarding the second research question on the use of digital materials for teaching and learning (TL) in schools in relation to teachers' digital competences, insights were obtained from 11 (23%) of the papers. One study by (Masango et al., 2020) investigated the competences of teachers in integrating digital textbooks in 356 South African schools, with complete questionnaire responses received from 51 schools. The research indicated that 47.1% of teachers believed they possessed the necessary competences for using digital textbooks. However, 35.3% of teachers expressed neutrality, suggesting doubts about the obtained results. Furthermore, (Wijaya et al., 2022) found in their research that some teachers used digital textbooks because they believed such materials helped improve the quality of learning.

Moreover, (Grimalt-Álvaro et al., 2019) and (Masango et al., 2020) stated that teachers with basic digital knowledge did not require additional training in the use of digital textbooks. Among the articles providing insights on the second research question, half of them emphasized the need for continuous professional development for teachers. Noteworthy papers highlighting this aspect include (Grönlund et al., 2018), (Håkansson Lindqvist, 2019), (Hermita et al., 2023), (Gerard et al., 2022), and (Crook & Sharma, n.d.). Additionally, (Gerard et al., 2022) emphasized that experienced teachers who have used digital textbooks can provide valuable professional development support to their colleagues. Teachers expressed the need for quality support to effectively implement digital content in teaching and learning, as mentioned in studies by (Masango, 2022), (Crook & Sharma, n.d.), (Gerard et al., 2022), and (Håkansson Lindqvist, 2019).

Digital textbooks were primarily used when introducing new material, as indicated by (Grimalt-Álvaro et al., 2019). Furthermore, (Wijaya et al., 2022) highlighted that mathematics teachers used digital textbooks to enhance the quality of learning in their classrooms. Another important aspect emphasized by (Richardson & Roebuck Sakho, 2022) was the significance of developing new curricula that incorporate advanced technologies such as virtual and augmented reality, algorithmic learning, and artificial intelligence. They emphasized the advantage of integrating these technologies into the curriculum and learning design to improve interaction between students and teachers and better prepare students for their future professional work.

A particularly interesting finding from the research conducted by (Håkansson Lindqvist, 2019) was teachers' perception of the substantial time spent on reviewing and selecting digital textbooks to adapt them to individual students' needs to support their learning.

RQ 3. Is there a relationship between the use of digital materials for TL and the maturity level of the school or educational system?

Our study did not identify any scientific papers that comprehensively address all three research areas of digital maturity, school education, and the utilization of digital textbooks.

4 Discussion

The analyzed articles predominantly discuss the implementation of digital materials, including digital textbooks, in teaching and learning. In order to present a comprehensive perspective, we have included a SWOT analysis table (Table 2) that summarizes the strengths, weaknesses, opportunities, and threats associated with their utilization, based on the findings of the analyzed articles.

Table 2. SWOT analysis

Strengths	**Weaknesses**
- digital material (DM) is practical and effective - students use DM without training - use of DM increase student activity within the class - DM is particularly efficient in introducing new content - improved learning outcomes through systemic use of DM	- student discipline issues arise with DM usage - waning student interest in interactive tools over time - students employ mobile gadgets in unintended ways during class - teachers lack sufficient familiarity and efficient use of DM
Opportunities	**Threats**
- enhance out of the class activity (learning at home) - availability of DM - availability of authentic context for learning - education authorities support purchase of digital textbooks - preparing collaborative digital textbooks - positive attitude towards technology among students	- poor technology - lack of access to suitable technology - technology-related anxiety - continuous updates to user interface and user experience - DM use not always student-centered - variation in digital material quality -limited availability of teacher training for effective DM use

As previously mentioned, digital textbooks are predominantly used for introducing new content (Grimalt-Álvaro et al., 2019). However, concerns persist regarding teachers' familiarity with digital textbooks (Wong et al., 2016), (Busulwa & Bbuye, 2018), (Lo et al., 2021), and their ability to integrate them with innovative teaching and learning strategies that support student-centered approaches. Teachers expressed the need for substantial time to review and select quality digital materials (Håkansson Lindqvist, 2019), showing less concern about the price.

While additional training in the use of digital materials may not be necessary for teachers with basic digital knowledge, possessing basic IT skills and access to digital materials does not guarantee their effective use (Grimalt-Álvaro et al., 2019) and (Masango et al., 2020), particularly in terms of implementing pedagogically effective approaches. Notably, despite the utilization of digital materials, teaching still tends to be teacher-centered, without the incorporation of innovative methods (Crook & Sharma, n.d.), which is consistent with the findings of the analyzed studies. Specifically, 45% of the analyzed articles indicated the absence of innovative teaching and learning methods, while only 5% confirmed the use of the flipped classroom approach.

The need for continuous professional development for teachers was emphasized by (Grönlund et al., 2018), (Håkansson Lindqvist, 2019), (Hermita et al., 2023), (Gerard et al., 2022), and (Crook & Sharma, n.d.). This need was also recognized by local authorities and teachers themselves (OECD, 2022). Moreover, teachers who have effectively utilized digital textbooks can assist in training their peers and identifying the best forms of professional development support (Gerard et al., 2022).

Furthermore, (Wong et al., 2016) and (Grönlund et al., 2018) highlighted the insufficient familiarity of teachers with digital textbooks. (Grönlund et al., 2018) expressed doubts about whether teachers have fully incorporated the new collaborative design of digital textbooks into their teaching practices or if they still perceive them as static books. This suggests that making full use of digital tools necessitates new learning designs that take advantage of rich collaborative digital materials. The need for modern textbooks that teachers and students can creatively utilize was recognized in previous research (Hajdin & Divjak, 2016). To conclude, quality digital materials may enhance student activity and have a positive impact on achieving learning outcomes, but only when paired with sound learning design. Despite students' enthusiasm for adopting the latest mobile technology and digital content (Pratama & Firmansyah, 2021) and (Chiu, 2017), most teachers have concerns about using mobile phones in classrooms (Busulwa & Bbuye, 2018) due to the difficulties in restricting their use solely to learning purposes.

Certain studies highlighted the effectiveness of continuous use of digital materials in schools tied to school and educational system resilience and readiness. This readiness was a prerequisite for successfully transitioning to predominantly digital learning and teaching, as demanded by unprecedented challenges such as those triggered by the COVID-19 pandemic (Svetec & Divjak, 2021).

It is particularly interesting to note that no studies connecting the use of digital textbooks with the digital maturity of schools, or the education system were found, indicating a potential avenue for future research.

In conclusion, there is a need for systematic investigation into the implementation of digital textbooks in schools and its connection to teachers' digital competences and, indirectly, the digital maturity

of the school. This will ensure greater incorporation of innovative forms of teaching and learning, ultimately preparing students for the jobs of the future.

One limitation of this study is that only papers written in the last ten years, in English, and indexed in the WoS or Scopus databases or accessible through the Research Rabbit tool were analyzed.

Finally, recognizing Research Rabbit's limitations is important. Despite its potential, there are constraints. Since the SLR was conducted in a traditional manner, the utility of the RR tool was limited. Additionally, as with many machine learning applications, researchers have limited control over the selection and ranking of outputs, making the research process less reproducible and difficult to explain. The RR tool did not provide information on how its algorithms work or the dataset on which the search relies (Krzton, 2023).

5 Conclusion

This paper offers a literature review via SLR and an AI tool to examine the implementation of digital materials, particularly digital textbooks, in schools. The review identifies positive/negative effects, opportunities and challenges. Out of the 43 papers analyzed, the majority focused on the use of digital materials in schools, but only a quarter of them specifically examined the relationship between digital materials and teachers' digital skills. Surprisingly, none of the papers investigated the link between the use of digital materials and the digital maturity of schools. The use of digital materials was found to encourage active student participation and positively impact learning outcomes, particularly when integrated into innovative learning designs. Negative effects include technical issues, discipline maintenance, and a decrease in student interest and enthusiasm over time. Teachers' lack of familiarity with digital textbooks and efficient utilization was identified as a challenge. Having basic digital skills was considered a necessary but insufficient condition for successful use of digital materials. For meaningful use, continuous teacher training is vital, integrating digital materials into innovative teaching and learning strategies. Future research should explore the connection between the use of digital textbooks, the digital readiness of the educational system, and the digital maturity of schools.

References

Bouck, E. C., Weng, P.-L., & Satsangi, R. (2016). Digital versus Traditional: Secondary Students with Visual Impairments' Perceptions of a Digital Algebra Textbook. *Journal of Visual Impairment & Blindness*, *110*(1), 41–52. https://doi.org/10.1177/0145482X1611000105

Bruhn, A. L., & Hasselbring, T. S. (2013). Increasing Student Access to Content Area Textbooks. *Intervention in* School *and Clinic*, *49*(1), 30–38. https://doi.org/10.1177/1053451213480030

Busulwa, H. S., & Bbuye, J. (2018). Attitudes and coping practices of using mobile phones for teaching and learning in a Uganda secondary school. Open Learning: *The Journal of Open, Distance and e-Learning*, 33(1), 34–45. https://doi.org/10.1080/02680513.2017.1414588

Chen, X. (2017). A comparative study of visual representations in conventional, digitized and interactive high school science textbooks. *Journal of Visual Literacy*, *36*(2), 104–122. https://doi.org/10.1080/1051144X.2017.1386388

Chiu, T. K. F. (2017). Introducing electronic textbooks as daily-use technology in schools: A top-down adoption process. *British Journal of Educational Technology*, 48(2), 524–537. https://doi.org/10.1111/bjet.12432

Crook, S. J., & Sharma, M. D. (n.d.). Bloom-ing Heck! The Activities of Australian Science Teachers and Students Two Years into a 1:1 Laptop Program Across 14 High Schools. 16.

Divjak, B., Rienties, B., Iniesto, F., Vondra, P., & Žižak, M. (2022). Flipped classrooms in higher education during the COVID-19 pandemic: Findings and future research recommendations. International *Journal of Educational Technology in Higher Education*, 19(1), 9. https://doi.org/10.1186/s41239-021-00316-4

Dudaitė, J., & Prakapas, R. (2019). Influence of use of Activinspire interactive whiteboards in classroom on students' learning. *Digital Education Review*, 299–308. https://doi.org/10.1344/der.2019.35.299-308

Farisi, M. I. (2014). Bhinneka Tunggal Ika [Unity in Diversity]: From Dynastic Policy to Classroom Practice. *JSSE - Journal of Social Science Education*, 12014 Insights into Citizenship Classrooms: The Art of Documentation & Description. https://doi.org/10.4119/JSSE-687

Fouh, E., Breakiron, D.A., Hamouda, S., Farghally, M.F., Shaffer,C.A., (2014) Exploring students learning behavior with an interactive etextbook in computer science courses. *Computers in Human Behavior*, 41. 478-485. https://doi.org/10.1016/j.chb.2014.09.061.

Gerard, L., Wiley, K., Debarger, A. H., Bichler, S., Bradford, A., & Linn, M. C. (2022). Self-directed Science Learning During COVID-19 and Beyond. *Journal of Science Education and Technology*, 31(2), 258–271. https://doi.org/10.1007/s10956-021-09953-w

Grimalt-Álvaro, C., Ametller, J., & Pintó, R. (2019). Factors Shaping the Uptake of ICT in Science

Classrooms. A Study of a Large-Scale Introduction of Interactive Whiteboards and Computers. *International Journal of Innovation in Science and Mathematics Education*, 27(1). https://doi.org/10.30722/IJISME.27.01.002

Grönlund, Å., Wiklund, M., & Böö, R. (2018). No name, no game: Challenges to use of collaborative digital textbooks. Education and Information Technologies, 23(3), 1359–1375. https://doi.org/10.1007/s10639-017-9669-z

Hajdin, G. & Divjak, B.(2016) Use of informatics textbooks in school classroom. *Journal of information and organizational sciences*, 40 (1), 21-44.

Håkansson Lindqvist, M. (2019). Talking about digital textbooks. The teacher perspective. The International *Journal of Information and Learning Technology*, 36(3), 254–265. https://doi.org/10.1108/IJILT-11-2018-0132

Hermita, N., Wijaya, T. T., Yusron, E., Abidin, Y., Alim, J. A., & Putra, Z. H. (2023). Extending unified theory of acceptance and use of technology to understand the acceptance of digital textbook for elementary School in Indonesia. *Frontiers in Education*, 8, 958800. https://doi.org/10.3389/feduc.2023.958800

Joo, Y. J., Park, S., & Shin, E. K. (2017). Students' expectation, satisfaction, and continuance intention to use digital textbooks. *Computers in Human Behavior*, 69, 83–90. https://doi.org/10.1016/j.chb.2016.12.025

Kabugo, D. (n.d.). *Utilizing Open Education Resources to Enhance Students' Learning Outcomes during the* COVID-*19 Schools Lockdown: A Case of Using Kolibri by Selected Government Schools in* Uganda. 12.

Kim, P., & Yu, J.-S. (2019). A Study on Online Delivery of Digital Textbooks in Korea. *Universal Journal of Educational Research*, 7(5A), 92–102. https://doi.org/10.13189/ujer.2019.071511

Krzton, A. (2023, March 21). Welcome to the Machine: Ir/Responsible Use of Machine Learning in Research Recommendation Tools. *ACRL 2023 Proceedings*. ACRL 2023 Conference Proceedings, American Library Association.

Lee, S., Lee, J.-H., & Jeong, Y. (n.d.). *The Effects of Digital Textbooks on Students' Academic Performance*, Academic *Interest, and Learning Skills*.

Lim, K., Go, J., Kim, J., Son, J., Jang, Y., & Joo, M.-H. (2022). Sustainable Effect of the Usefulness of and Preference for Digital Textbooks on Perceived Achievements in Elementary Education Environments. *Sustainability*, 14(11), 6636. https://doi.org/10.3390/su14116636

Lo, C. K., Cheung, K. L., Chan, H. R., & Chau, C. L. E. (2021). Developing flipped learning resources to support secondary school mathematics teaching during the COVID-19 pandemic. *Interactive Learning Environments*, 1–19. https://doi.org/10.1080/10494820.2021.1981397

Lo C.K., Tlili A, Huang X. (2022). The Use of Open Educational Resources during the COVID-19 Pandemic: A Qualitative Study of Primary School Mathematics Teachers in Hong Kong. *Education Sciences*. 12(11):744. https://doi.org/10.3390/educsci12110744

Lucas, M. (2020). External barriers affecting the successful implementation of mobile educational interventions. Computers *in Human Behavior*, *107*, 105509. https://doi.org/10.1016/j.chb.2018.05.001

Masango, M. M. (2022). A Paperless Classroom: Importance of Training and Support in the Implementation of Electronic Textbooks in Gauteng Public Schools. 20(3), 15.

Masango, M. M., Van Ryneveld, L., & Graham, M. A. (2020). Barriers to the Implementation of Electronic Textbooks in Rural and Township Schools in South Africa. *Africa Education Review*, 17(6), 86–118. https://doi.org/10.1080/18146627.2022.2064310

Milić, M. & Divjak, B. (2022). Digital Maturity of Schools - Explanation, Literature Review and Analysis, *Central European Conference on Information and Intelligent Systems*

Moundy, K., Chafiq, N., & Talbi, M. (2021). Comparative Analysis of Student Engagement in Digital Textbook Use during Quarantine. *Education Sciences*, 11(7), 352. https://doi.org/10.3390/educsci11070352

Moundy, K., Chafiq, N., & Talbi, M. (2022). Digital Textbook and Flipped Classroom: Experimentation of the Self-Learning Method Based on the Development of Soft Skills and Disciplinary Knowledge. *International Journal of Emerging Technologies in Learning (IJE*T), 17(07), 240–259. https://doi.org/10.3991/ijet.v17i07.28933

OECD. (2022). *Education at a Glance 2022: OECD Indicators*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/education-at-a-glance-2022_3197152b-en;jsessionid=oktyOPo3720lBNrtdqOoNnU13gQXJkfdnbeR18Le.ip-10-240-5-190

Pratama, A. R., & Firmansyah, F. M. (2021). How can governments nudge students to become ebook readers? Evidence from Indonesia. *Digital Library Perspectives*, 37(3), 275–288. https://doi.org/10.1108/DLP-07-2020-0066

Prutko, A. S., Kudussov, A. S., Mussenova, E. K., & Kambarova, Zh. T. (2020). Development of the electronic physics textbook for students of secondary school. *Bulletin of the Karaganda University. 'Physics' Series*, *100*(4), 87–94. https://doi.org/10.31489/2020Ph4/87-94

Radović, S., Radojičić, M., Veljković, K., & Marić, M. (2020). Examining the effects of Geogebra applets on mathematics learning using interactive mathematics textbook. *Interactive Learning Environments*, 28(1), 32–49. https://doi.org/10.1080/10494820.2018.1512001

Reinhold, F., Hoch, S., Schiepe-Tiska, A., Strohmaier, A. R., & Reiss, K. (2021). Motivational and Emotional Orientation, Engagement, and Achievement in Mathematics. A Case Study With One Sixth-Grade Classroom Working With an Electronic Textbook on Fractions. *Frontiers in Education*, 6, 588472. https://doi.org/10.3389/feduc.2021.588472

Reinhold, F., Strohmaier, A., Hoch, S., Reiss, K., Böheim, R., & Seidel, T. (2020). Process data from electronic textbooks indicate students' classroom engagement. *Learning and Individual Differences*, 83–84, 101934. https://doi.org/10.1016/j.lindif.2020.101934

Rezat, S. (2021). How automated feedback from a digital mathematics textbook affects primary students' conceptual development: Two case studies. *ZDM – Mathematics Education*, *53*(6), 1433–1445. https://doi.org/10.1007/s11858-021-01263-0

Richardson, S., & Roebuck Sakho, J. (2022). Creating equitable access: Using OER for socially just educational leaders. *Journal for Multicultural Education*, 16(5), 443–455. https://doi.org/10.1108/JME-12-2021-0224

Roskos, K., & Neuman, S. B. (2014). Best Practices in Reading: A 21st Century Skill Update. *The Reading Teacher*, 67(7), 507–511. https://doi.org/10.1002/trtr.1248

Shadiev, R., Hwang, W.-Y., Huang, Y.-M., & Liu, T.-Y. (2015). The Impact of Supported and Annotated Mobile Learning on Achievement and Cognitive Load. The Impact of Supported and Annotated Mobile Learning on Achievement and Cognitive Load. *Educational Technology & Society*, 18(4), 53–69.

Shen, H., Luo, L., & Sun, Z. (2015). What Affect Lower Grade Learner's Perceived Usefulness and Perceived Ease of Use of Mobile Digital Textbook Learning System? An Empirical Factor Analyses Investigation in China. *International Journal of Multimedia and Ubiquitous Engineering*, 10(1), 33–46. https://doi.org/10.14257/ijmue.2015.10.1.4

Sinaga, P., Setiawan, W., & liana, M. (2022). The impact of electronic interactive teaching materials (EITMs) in e-learning on junior high school students' critical thinking skills. *Thinking Skills and Creativity*, *46*, 101066. https://doi.org/10.1016/j.tsc.2022.101066

So, H. J., Choi, H., & Yoon, H. G. (2015). Understanding users' perceived needs and concerns toward mobile application integration in primary science education in Korea. International *Journal of Mobile Learning and Organisation*, 9(4), 315. https://doi.org/10.1504/IJMLO.2015.074515

Sun, Z., & Jiang, Y. (2015). How the young generation uses digital textbooks via mobile learning terminals: Measurement of elementary school students in China: How the young generation uses digital textbooks. British *Journal of Educational Technology*, *46*(5), 961–964. https://doi.org/10.1111/bjet.12299

Svetec, B., & Divjak, B. (2021). Emergency Responses to the COVID-19 Crisis in Education: A Shift from Chaos to Complexity. *EDEN Conference Proceedings*, 1, 513–523. https://doi.org/10.38069/edenconf-2021-ac0051

Tlili, A., Zhao, J., Yang, K., Wang, Y., Bozkurt, A., Huang, R., Bonk, C. J., & Ashraf, M. A. (2022). Going beyond books to using e-books in education: A systematic literature review of empirical studies. Interactive *Learning Environments*, 1–25. https://doi.org/10.1080/10494820.2022.2141786

Vorotnykova, I. (2019). Organizational, Psychological and Pedagogical Conditions for the Use of E-Books and e-Textbooks at School. Turkish *Online Journal of Distance Education*, 89–102. https://doi.org/10.17718/tojde.598227

Wang, X., & Xing, W. (2019). Understanding Elementary Students' Use of Digital Textbooks on Mobile Devices: A Structural Equation Modeling Approach. *Journal of Educational Computing Research*, 57(3), 755–776. https://doi.org/10.1177/0735633118758643

Wijaya, T. T., Zhou, Y., Houghton, T., Weinhandl, R., Lavicza, Z., & Yusop, F. D. (2022). Factors Affecting the Use of Digital Mathematics Textbooks in Indonesia. *Mathematics*, 10(11), 1808. https://doi.org/10.3390/math10111808

Wong, B. T. M., Li, K. C., Yuen, K. S., & Wu, J. W. S. (2016). Adopting and adapting open textbooks: School teachers' readiness and expectations. *International Journal of Services and Standards*, 11(2), 160. https://doi.org/10.1504/IJSS.2016.077954

Zacok, L., Bernat, M., & Bernatova, R. (2019). Verification of New Electronic Technical Textbooks In The Current School. *Ad Alta-Journal Of Interdisciplinary Research*, *9*(2), 371–375.

Joint Creative Classroom as a Response to Industry 4.0 Demands

Marko Matus, Goran Hajdin, Igor Balaban
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{marko.matus, goran.hajdin, igor.balaban}@foi.unizg.hr

Abstract. *This paper presents an analysis of the results collected from the eight Joint Creative Classrooms (JCC), a new approach to learning and teaching in computer science and related courses, which was piloted within a European project among five different universities. Factors that affect student satisfaction and engagement, as well as gains in their skills were studied. In total, 96 responses related to student reactions before and after the JCC sessions, as well as 110 responses related to learning outcomes, were analysed. The results indicate that the students would recommend their JCC to others, suggesting that the overall satisfaction with JCC is quite high. Research results indicate various benefits to learning within JCC. According to the results, student engagement and peer collaboration are positively related to student satisfaction.*

Keywords. Industry 4.0, Joint Creative Classroom, Higher Education

1 Introduction

Industry 4.0 brings changes in how people work and create value in contemporary society. Increasing automation, robotization, and digitalisation of business processes are transforming the demands of the labour market (Guadalupe et al., 2023) requiring new knowledge and skills from the workforce, thus placing new tasks in front of the educational system. To answer the challenges of higher specialisation of the labour market, higher education must provide students with the knowledge and skills necessary to adapt to the rapid changes in the industry and face the challenges of digital transformation (Baygin et al., 2007). New jobs are being created due to the need to implement the concept of Industry 4.0, which means that employers will demand new technologies, particularly digital skills of their employees (Grenčíková et al., 2021). Higher education institutions are essential drivers of these changes and must embrace new teaching techniques based on technologies such as the Internet of Things (IoT), Cloud computing, Virtual and Augmented reality, simulations, Artificial intelligence (AI), Data analytics, and others. In addition, there is significant pressure on governments to recognise the requirements of such education.

One of the main advantages of Industry 4.0-influenced technology in higher education is improved accessibility to education (Mian et al., 2020). Many talents are lost due to the unavailability of quality motivation or comprehensive materials. Additionally, greater efficiency is achieved using interactive content and simulations (Chen et al., 2020). Better collaboration and deeper learning approaches can foster greater adoption of educational materials, the formation of learning communities, the development of new curriculum programmes by faculties, and the collaboration between teachers and students (Adams Becker et al., 2017).

The teaching process that is influenced by Industry 4.0-influenced technology has multiple advantages over traditional teaching methods. These benefits will have a positive impact on students and the institutions they represent through their work (Baygin et al., 2016; Sudibjo et al., 2019; Goh & Abduh-Wehab, 2020).

To be successful in teaching, higher education must focus on student satisfaction as one of the factors contributing to its quality. Since satisfaction is a complex construct, higher education institutions should focus on relevant factors that will contribute to the defined goal (Alves & Raposo, 2009; Hang & Tam, 2018). Thus, this paper focusses on the satisfaction and skills change of students after participating in Joint Creative Classrooms (JCC) that were designed as part of the project "Accelerating the transition towards Edu 4.0 in HEIs" (Teach4Edu4) conducted from 2020 to 2023 (Rienties et al., 2023). JCCs are computer science courses developed in collaboration with two or more partner universities using the learning design approach aligned with Education 4.0 with a total workload of 2-3 ECTS. They can be seen as a unique student-centred environment that enables students from different partner institutions to join the same classroom and, in addition to domain knowledge, to gain mobility from home experience and transferable skills (e.g., Communication, Reliability, Teamwork, Organisation, etc.). At the same time, at least two teachers from

different universities are also working collaboratively to design such courses, create the content, and run them in different settings. Some JCCs were run completely online, while others were run in blended mode.

The purpose of this paper is to explore the factors that influence student satisfaction and how individual skills have changed after JCC sessions. For that approach, we used the Kirkpatrick's model, a globally recognised method of evaluating the results of online learning programmes (Gandomkar, 2018). The Model allows programme designers and researchers to assess formal and informal training methods and to rate them against four levels of criteria: reaction, learning, behaviour, and results. In this paper we will focus on learners' reaction, and their learning gains which are the first two levels in the Kirkpatrick's Model.

2 Research goals

Since Kirkpatrick's model describes four fundamental levels of evaluation without suggesting any relationships within the constructs, the research objectives of this paper are as follows:

- RO1: To identify factors that affect students' satisfaction with the Joint Creative Classrooms;
- RO2: To determine how individual students' skills have changed after completing the Joint Creative Classrooms;
- RO3: To investigate how student engagement in learning and peer collaboration influences their satisfaction.

3 Methods

In this section we will describe the participants (sample), data collection procedure and measurement instrument creation.

3.1 Participants and data collection procedure

The Teach4Edu4 project started its activities in November 2020. A total of 21 teachers and 162 students from 5 different universities in Europe participated in the Teach4Edu4 project. In total, 8 JCCs were held on different topics within the computer science field. In this paper, we focus on the student's satisfaction through the analysis of 92 responses from JCC participants and 110 responses related to learning. The difference in the number of students arises because not all students decided to participate in the surveys at the end of the JCC, nor did all of them complete the JCC. The sample profile is presented in Table 1.

3.2 Instrument

The first questionnaire was designed to measure student reactions (the first level of the Kirkpatrick model) and consisted of three constructs: satisfaction (7 items), engagement (3 items) and peer work (2 items). For each item, students had to indicate their agreement or disagreement on a Likert scale ranging from 1 - "strongly disagree" to 4 - "strongly agree". The final set of items is presented in Table 4 after validation of the measurement instrument. There are numerous examples that describe satisfaction, but we focused on a set of measures proposed by Ashby et al. (2011) and Li et al. (2016) that are most suitable for our research. Literary sources by Dixon (2015) and Handelsman et al. (2005) were used for the construct of engagement.

Table 1. Sample profile

	Gender	Reaction aspects group (n=92)	Learning aspects group (n=110)
Gender	Male	65 (71%)	78 (70,9%)
	Female	24 (26%)	29 (26,4%)
	Prefer not to say	3 (3%)	3 (3%)
Average age	Male	23,8	23,9
	Female	24,6	24,9
	Prefer not to say	25,3	25,6
Standard deviation	Male	2,65	2,99
	Female	2,66	2,98
	Prefer not to say	1,38	1,32

The second questionnaire was related to the learning aspect of JCCs (the second level of the Kirkpatrick's model) and consisted of 10 items related to skills and 2 items related to attitudes. The instrument was administered to students at the beginning and at the end of each JCC to measure the gains in specific skills and expectations from the JCC. Examples of questions before JCC started and after it ended are presented in Table 2. The main difference between the two columns presented in Table 2 is that the first column shows the outcomes that students want to learn before the JCC started while the second column refers to fulfilment of the outcomes.

Table 2. Items related to learning

Before JCC	After JCC
I would like to learn where and when I choose.	I could learn where and when I chose.
It would be great if the course was personalized to my learning.	Learning was personalised to my needs.
I would like to be able to study how I like to study (rather than following a predetermined structure by the teacher).	I could study the way I like to study.
I would like to work on projects with other students.	I worked on projects with other students.
I would like to gain hands-on, authentic experiences and real-world skills.	I gained hands-on, authentic experiences, and real-world skills.
I would like to learn how to interpret and reason with data.	I learnt how to interpret and reason with data.
I would like the assessments to be innovative.	I participated in new ways of assessment.
I would like to contribute to the design and implementation of the JCC.	My voice and input were used to (re)design, implement, and/or improve the JCC.
I would like to be more independent.	I have become more independent in my study.
I would like to study an approach to learning and teaching that emphasises the development of skills and competencies necessary in a modern workplace using up-to-date technology.	I would like to study an approach to learning and teaching that emphasises the development of skills and competencies necessary in a modern workplace using up-to-date technology.
I would like to develop my soft skills (such as teamwork and creativity).	I was able to strengthen my soft skills (such as teamwork and creativity).
I would like to develop my skills to work effectively in a modern workplace.	I was able to strengthen my skills to work effectively in a modern workplace.

4 Results

Before we could interpret the results with confidence, the validity and reliability of the instrument needed to be confirmed for the first level of the Kirkpatrick model. Table 3 shows the results of the instrument validity and reliability with the sample of 92 students. As seen in Table 3, the instrument showed good validity and reliability. Confirmatory factor analysis was also conducted and items below the 0.6 threshold were eliminated to achieve the satisfactory validity of the instrument scales. To assess the reliability of the responses within the measuring instrument, Crombach alpha (CA) was calculated with a values above 0.7, which is considered acceptable according to Pallant J. (2007). Average Variance Extracted (AVE) is used to asses the convergent validity of a construct. Considering that all the latent variables have an AVE value of more than 0.5, therefore, convergent validity has been achieved (Pallant J., 2007). Construct Reliability (CR) which measure of internal consistency in scale items is very high because all the values are above 0.85 (Pallant J., 2007). The final list of items for further analysis is presented in Table 4.

Table 3. Construct reliability and validity

	CA	AVE	CR
Engagement	0,75	0,655	0,850
Peer Collaboration	0,86	0,877	0,935
Satisfaction	0,88	0,589	0,908

Table 4. Instrument items in three (sub) constructs for the first level of Kirkpatrick's model

Satisfaction
SAT1 Overall, I would recommend this JCC to other colleagues. (Li et al., 2016)
SAT2 Overall, this JCC met my expectations. (Li et al., 2016)
SAT3 Overall, I enjoyed studying this JCC. (Li et al., 2016)
SAT4 Overall, I was satisfied with the JCC materials provided. (Li et al., 2016)
SAT5 Overall, I was satisfied with the method of delivering different JCC materials and learning activities. (Li et al., 2016)
SAT6 Overall, I was satisfied with the assessment during this JCC. (Li et al., 2016)
SAT7 I was satisfied with the support provided by my facilitator. (Li et al., 2016)
ENGAGEMENT
ENG 1 I carefully read the JCC materials. (Dixon, 2015)
ENG 2 I applied the JCC materials to my life. (Dixon, 2015; Handelsman et al., 2005)
ENG 3 I was very motivated to learn the JCC content. (Dixon, 2015; Handelsman et al., 2005)
PEER WORK
PW1 I enjoyed working with students from the partner JCC institution.

PW2 I have learnt a lot working with students from the partner JCC institution.

To address the first research objective (RO1), the results were interpreted based on the mode function within each category because not all categories had an identical number of items (questions). Mode presents a value which appears most often in the data set. Survey responses were based on a 1-4 Likert scale and are presented in Figure 1. Overall, the highest satisfaction was expressed with the highest possible rating of four in the Overall JCC (SAT1) category, where students indicated that they would recommend this JCC to others, suggesting that the overall satisfaction is quite high. In all other satisfaction categories, the most common rating is three which points to a positive agreement with the statements presented in Table 4 (SAT 2-7). The observed satisfaction categories are Materials (SAT 4-5), Assessment (SAT 6), and Tutor/guidance (SAT 7).

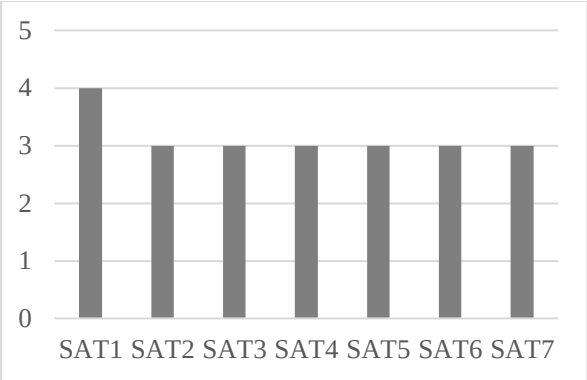


Figure 1. Results of the first research objective (Mode values)

Based on the results of the T test shown in Table 5 which shows whether there was a statistically significant difference between two groups of answers (Pallant J., 2007) or the responses before and after the implementation of the JCC, the second research objective (RO2) significantly demonstrated that the students were able to learn where and when they wanted and in the way they desired, to a greater extent than expected as shown in Table 2. The reasons that can be inferred are the flexibility of the instruction itself, as some teachers recorded lectures and provided interactive video materials, enabling students to adjust their own schedules. Additionally, the students truly worked on projects with other students through collaborative learning environments, which allowed them to improve skills such as communication, teamwork, organisation, problem solving, stress management, and more. However, students believe that the approach to learning and teaching did not emphasise the development of skills and competencies needed in the modern work environment using contemporary technology. The students also learnt to a lesser extent than expected how to interpret and think about data. The students also feel that the JCCs did not lead them to develop learning strategies and they believe that they were able to strengthen their skills for effective work in a modern work environment to a significantly lower extent than expected. A possible reason for such a result could be high student expectations, but this needs to be further researched.

Table 5. T-test values

Question	Df value	T Stat value	P value
1.	267	-3,781	0,001
3.	267	-5,649	0,001
4.	267	-4,121	0,001
6.	198	1,707	0,05
9.	197	3,123	0,001
10.	267	2,482	0,05
12.	193	3,954	0,001

To address the third objective of this research (RO3), the Spearman correlation was calculated. Correlation analysis is used to describe the strength and direction of the liner relationship between two variables, while Spearman rank order correlation is designed for use with ordinal level or ranked data and isuseful when your data does not meet the criteria for Pearson's correlation (Pallant J., 2007). A total of 92 student responses were analysed and the literary source used in interpreting the results was Pallant J. (2007). The answers were analysed to the questions shown in Table 4 and the results are shown in Table 6.

Table 6. Correlations between main constructs

	Engagement	Peer collaboration	Satisfaction
Engagement	1,000		
Peer collaboration	0,074**	1,000	
Satisfaction	0,73*	0,51*	1,000

*p<0,001, **p>0,5

The first correlation conducted compared the responses between student engagement and satisfaction, resulting in a value of 0,73. This value indicates a relatively high positive correlation between these variables. Since in most cases, more engaged students are more active and dedicated to learning, they usually achieve better results, which fulfils them and makes them more satisfied. The second correlation conducted compared the responses regarding peer collaboration and satisfaction, resulting in a value of 0,51. This value suggests a highly positive correlation between these variables, closer to the lower limit. On the basis of the responses, it can be concluded that the students provided support to each other, which can

positively impact their satisfaction. However, it should be noted that if there is poor team organisation, the responses of all students are not relevant because students who are more burdened will certainly feel less satisfaction if they must work more than other students. The third correlation between engagement and peer collaboration results in a value of 0,074, suggesting a minimal correlation between these two variables. Such a value can be observed because not every student who collaborates in peer work have identical engagement motivation, and not all the students are equally burdened.

5 Conclusions

Industry 4.0 has a significant impact on higher education, requiring adaptation of curricula, integration of new technologies, and collaboration with industry. This research analysed JCCs as a new concept of learning and teaching in line with the requirements of Industry 4.0. The research results showed that the students strongly agree that they would recommend this JCC to other colleagues, which means that overall JCC satisfaction is quite high. In other satisfaction categories like materials, assessment or tutor/guidance, satisfaction is evaluated with the most common rating of 3 out of 4. Furthermore, the implementation of JCC has positively influenced the development of certain student skills, such as flexible learning and teamwork, but did not emphasise the development of skills needed in the modern environment. The possible reason for such a result could be high student expectations or lack of attention to the mentioned skills, but this needs to be further researched. According to Williams (2003), it is essential for coaches and administrators to underscore the significance of teamwork and collaborative skills in the realm of policymaking and staff/faculty development initiatives, placing particular emphasis on the role of teamwork in the development of courses and the implementation of technology. Student engagement and peer collaboration have a positive correlation with student satisfaction, which is consistent with previous research studies conducted by the authors Grey and DiLoreto (2016). JCCs are an innovative approach to higher education that promotes the development of key skills and knowledge for the contemporary job market. Future research can focus on differences between traditional teaching methods and adapted JCC, key skills that influence student outcomes, perception of modern approaches in higher education, and possible improvements in the application of modern technologies in JCC among gender groups. It is important to embrace new learning trends while critically considering their implementation, and to develop skills and leverage satisfaction in their application as motivation for further learning and improvement.

Acknowledgement

This research was conducted as part of the project "Accelerating the transition towards Edu 4.0 in HEIs" (Teach4Edu4), approved by the Erasmus+ program - KA2 - Cooperation for innovation and the exchange of good practices, which has started its activities in November 2020.

References

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., Ananthanarayanan, V., ... & Wolfson, N. (2017). NMC horizon report: 2017 library edition.

Alves, H., & Raposo, M. (2009). The measurement of the construct satisfaction in higher education. *The service industries journal*, *29*(2), 203-218.

Ashby, A., Richardson, J. T., & Woodley, A. (2011). National student feedback surveys in distance education: an investigation at the UK Open University. *Open Learning: The Journal of Open, Distance and e-Learning*, *26*(1), 5-25.

Baygin, M., Yetis, H., Karakose, M., & Akin, E. (2016, September). An effect analysis of industry 4.0 to higher education. In *2016 15th international conference on information technology based higher education and training (ITHET)* (pp. 1-4). IEEE.

Briggs, W. L., Handelsman, M. M., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *Journal of Educational Research*, *98*(3), 184-191.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264-75278.

Dixson, M. (2015). Measuring Student Engagement in the Online Coursecourse: The Online Student Engagement Scale (OSE). *Online Learning*, *19*(4).

Gandomkar, R. (2018). Comparing Kirkpatrick's original and new model with CIPP evaluation model. *Journal of Advances in Medical Education & Professionalism*, *6*(2), 94-95.

Goh, P. S. C., & Abdul-Wahab, N. (2020). Paradigms to drive higher education 4.0. *International Journal of Learning, Teaching and Educational Research*, *19*(1), 159-171.

Gray, J. A., & DiLoreto, M. (2016). The effects of student engagement, student satisfaction, and perceived learning in online learning environments. *International Journal of Educational Leadership Preparation*, *11*(1), n1.

Grenčíková, A., Kordoš, M., & Navickas, V. (2021). The impact of industry 4.0 on education contents. *Business: Theory and Practice*, *22*(1), 29-38.

Guadalupe Mori, V. H., Ogosi Auqui, J. A., Rosales Huamani, J. A., & Arenas Ñiquin, J. L. (2023). Industry 4.0: Implementation of Technologies in Medical Manufacturing Companies. In *Intelligent Sustainable Systems: Selected Papers of WorldS4 2022, Volume 1* (pp. 619-626). Singapore: Springer Nature Singapore.

Hang, N. P. T., & Tam, P. T. (2018). Impacting the industry 4.0 on the training quality and student's satisfaction at Lac Hong University. *Journal of Management Information and Decision Sciences*, *21*(1), 1-18.

Li, N., Marsh, V., & Rienties, B. (2016). Modelling and managing learner satisfaction: Use of learner feedback to enhance blended and online learning experience. *Decision Sciences Journal of Innovative Education*, *14*(2), 216-242.

Marzano, G., & Martinovs, A. (2020, May). TEACHING INDUSTRY 4.0. In *SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference* (Vol. 2, pp. 69-76).

Mian, S. H., Salah, B., Ameen, W., Moiduddin, K., & Alkhalefah, H. (2020). Adapting universities for sustainability education in industry 4.0: Channel of challenges and opportunities. *Sustainability*, *12*(15), 6100.

Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. McGraw-hill education (UK).

Rienties, B., Ferguson, R., Gonda, D., Hajdin, G., Herodotou, C., Iniesto, F., ... & Isidori, M. V. (2023). Education 4.0 in higher education and computer science: A systematic review. *Computer Applications in Engineering Education*.

Sudibjo, N., Idawati, L., & Harsanti, H. R. (2019, December). Characteristics of Learning in the Era of Industry 4.0 and Society 5.0. In *International Conference on Education Technology (ICoET 2019)* (pp. 276-278). Atlantis Press.

Williams, P. E. (2003). Roles and competencies for distance education programs in higher education institutions. *The American Journal of Distance Education*, *17*(1), 45-57

The use of social robots as teaching assistants in education: literature review

Ivana Ružić
1st Primary school Čakovec
K. Tomislava 43, 40000 Čakovec
ivana.ruzic@skole.hr

Igor Balaban
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42 000 Varaždin
igor.balaban@foi.unizg.hr

Abstract. *This literature review examines the impact of social robots as teaching assistants on student engagement in learning activities. The study identified and analysed 16 studies that investigate the use of social robots mostly in preschool or primary education. The results reveal that social robots, such as NAO, Pepper, and Skusie, have been used in various educational contexts, including language learning, STEM education, special education, and collaborative learning. The findings highlight the positive effects of social robots on student motivation and engagement, leading to improved learning outcomes. The review emphasizes the importance of taking ethical considerations into account, data privacy, and the role of human educators when integrating social robots into educational settings.*

Keywords. social robot, education, motivation, engagement

1 Introduction

Robotics is one of the expressions of technology and takes place in various contexts of life (Kalaitzidou & Pachidis, 2023). The use of advanced technologies that include robotics and artificial intelligence in education motivates students to learn at different educational levels. These technologies encourage active learning in an innovative, effective, motivating, and individualized way. Active learning equals active engagement of students in discussions, problem solving, case studies, role plays, and other methods with a higher degree of learners' responsibility. Many studies show that the integration of advanced technologies has a positive impact on learning about various topics.

A social robot in education refers to the use of robotic systems designed to interact with students, improve engagement, and support their learning experiences (Belpaeme et al. (2018); Kubilinskiene et al. (2017)). These robots are equipped with artificial intelligence (AI) and various sensors to perceive and respond to their environment and engage with students in a social and educational manner.

The impact and influence of social robots in teaching/learning process has been researched in different educational contexts, accentuating their great potential in supporting learners and teachers (Donnermann et al. (2020)). We describe some cases found in the literature according to the areas in which social robots assisted.

Personalized Learning: Social robots can adapt their interactions based on individual student's needs. They can assess student's abilities, tailor content and feedback, as well as provide personalized teaching and support (Chen et al. (2020)).

Language Learning: Robots can help with language acquisition by engaging students in conversation, pronunciation practice, and vocabulary building. They can provide real-time feedback, correct pronunciation errors, and simulate immersive language scenarios, creating interactive and engaging learning experiences in which children participate with high motivation (Sisman et al. (2018)).

STEM Education: Social robots can facilitate hands-on learning experiences in science, technology, engineering, and mathematics (STEM) subjects. They can guide students through experiments, help them understand complex concepts, and foster problem-solving skills (Konijn & Hoorn (2020)).

Special Education: Robots can provide valuable support to students with special needs. They can assist in communication, development of social skills, and emotional regulation (Lopez-Caudana et al. (2021)). Robots can create a non-threatening environment that promotes engagement and learning for children with autism spectrum disorder, for example.

Collaborative Learning: Robots can act as collaborative partners for group activities and projects. They can encourage teamwork, facilitate discussions, and provide prompts or challenges to foster critical thinking and creativity. The robot as a collaborative learning companion could also contribute to the child's social development (Ekström & Pareto, 2022).

Emotional Support: According to Escobar-Planas et al. (2022), social robots can offer emotional support

to students by acting as empathetic companions. They can engage in conversations, listen attentively, and provide encouragement. This can be particularly beneficial for students experiencing stress, anxiety, or loneliness.

Virtual Field Trips: Robots equipped with cameras and remote operation capabilities can allow students to explore remote locations or otherwise inaccessible environments. This enables virtual field trips, enhancing students' understanding of geography, history, or scientific phenomena (Oh et al. (2018)).

Coding and Robotics Education: Robots can serve as tools for teaching coding and robotics concepts. Students can program robots to perform specific tasks, solve problems and learn computational thinking skills along the way (Konijn & Hoorn (2020)).

The integration of social robots in education aims to enhance student engagement, provide personalized support, and foster the development of various skills. However, it is essential to take ethical considerations into account, followed by data privacy, and the role of human educators in the learning process when implementing social robots in educational settings.

Therefore, this study explores empirical findings on the use of social robots as assistants in education with a special interest in the impact of social robots on student engagement in learning activities across the educational sector.

2 General objective and research questions

The most important factor for active participation in the learning process and achieving success is learner motivation (Leitão et al. (2021)). Motivation can be described as a stimulus to action that defines intrinsic and varied extrinsic layers of motivation. Intrinsic motivation is determined by autonomy, competences, and socialisation, while extrinsic motivation is determined by external regulation: introjection, identification, and integration.

Engagement can affect learners' intrinsic or extrinsic motivation, and gamification as a strategy with the purpose of generating commitment can lead to changes in behaviour (Rojas-López et al. (2019)). It is defined as the passion to participate in and perform the assigned learning activities (Rojas-López et al. (2019)). It is encouraged through the use of active teaching methods. Learner engagement can be categorized into three categories: behavioural, cognitive, and emotional (Rojas-López et al. (2019)). The positive attitude of the learner towards participating in the teaching process is called behavioural participation. Cognitive engagement implies the disposition to think and comprehend one topic or concept, which involves self-regulation. The reactions and feelings of the students during the learning process are known as emotional engagement. Many studies show that the use of advanced technologies as a teaching strategy increases the engagement of the learners.

Therefore, the general objective of this study is to investigate the connection between the use of social robots as a teaching assistant and the teaching-learning processes. The following research questions are at the center of interest:

RQ1: What are the major fields of research in the context of social robots in education?

RQ2: How does the use of social robots as teaching assistants affect student engagement and motivation in learning activities?

3 Material and method

The literature review process was divided into four stages: identification, screening, eligibility, and inclusion, according to Boland et al. (2017) and the complete process is illustrated on Figure 1.

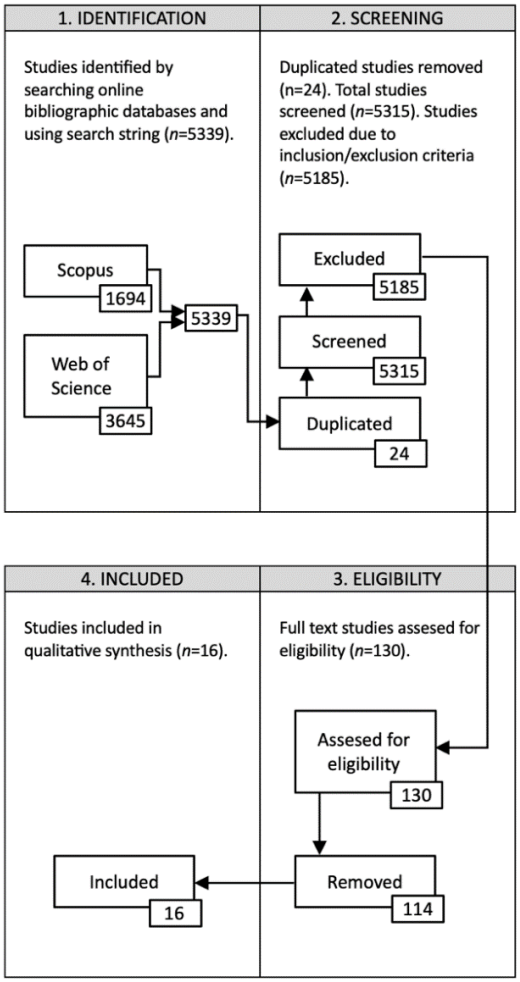


Figure 1: PRISMA flow diagram of the literature review process

In the first identification stage, a structured search strategy was created for use on the scientific databases Scopus and Web of Science. A search string: *social* and *robot* and *education*, was used for database search.

For the second stage, screening, 3,645 results from Web of Science and 1,694 results from Scopus were identified. The following additional selection criteria were used:

1. Published in English,
2. Published within the time frame 2019-2023,
3. Document type was article,
4. Full text was available,
5. The subject areas were Computer and Social Science (Scopus) and Educational Research in Education (Web of Science).

As a result, 45 studies were obtained from Scopus and 109 from Web of Science. 24 duplicates were detected and excluded, and 130 studies were reviewed by titles and abstracts.

Other means of availability, such as the author's contacts, were also used for the acquisition of studies. In this way, 5 studies were acquired, while 6 studies could not be acquired.

In the next stage, the final eligibility criterion was applied – studies should have included empirical findings on the use of social robots as assistants in education. From the focus are excluded all those studies that research robots in education that are not social robots, and studies that explore social robots, but these robots are not in the function of teaching assistants. 114 studies were removed because they did not meet the required criteria. A total of 16 studies were included in the fourth stage of the literature review.

4 Results

The results of the research are presented in Table 1 and they show the type of robot used, the number of participants and their age, the educational field (languages, STEM, other) and the duration of the educational activity.

Table 1: Presentation of the research results

Authors	Robot	Participants		Subject	Duration
		No	Level		
Sisman et al. (2018)	NAO	232	secondary	language	4 months
Escobar-Planas et al. (2022)	Haru	84	primary	problem-solving	1 session
Tolksdorf et al. (2021)	NAO	21	preschool	language	4 sessions
Connolly et al. (2022)	NAO	13	adults	health	-
Demir-Lira et al. (2020)	NAO	38	preschool	language	30 mins
Arar et al. (2021)	Emys	54	primary	language	8 weeks
Guggemos et al. (2020)	Pepper	462	high school	academic writing	45 mins
Yueh et al. (2020)	Julia	36	primary	literacy	60-80 min
van den Berghe et al. (2020)	NAO	104	preschool	language	7 sessions
Chen et al. (2020)	Tega	59	primary	language	48 sessions
Konijn & Hoorn (2020)	NAO	86	primary	mathematic	3 x 5 min
Velentza et al. (2021)	NAO	138	university	engineering	1 session
Chalmers et al. (2022)	NAO	29	teachers	-	8 weeks -9 months
Kim & Tscholl (2021)	Skusie	24	preschool	language	49 sessions
Kim et al. (2021)	Skusie	11	preschool	STEM	30 sessions
Ekström & Pareto (2022)	Pepper	92	teachers	mathematics	2 years

The results show that NAO is one of the most popular social robots in education, used in as many as 8 out of 16 studies. The robots Pepper and Skusie were used in two studies each, and the robots Haru, Emys, Julia and Tega were used in one study each.

The number of participants who actively participated in the research ranges from 11 to as many as 462. Studies are most often conducted among children of preschool and primary school age, with five studies conducted for each age category. Only one study was conducted among secondary school students, one among high school and one among university students. Two studies were conducted among teachers and one among adult participants (patients).

The use of social robots is most common in teaching foreign languages, as many as seven studies have been conducted in this area. The next field of education in terms of the popularity of use of social robots is the STEM field (mathematics, engineering). Four studies were conducted in the STEM field. One study per field was conducted in the areas of literacy, academic writing, health education and problem solving, while the field was not specified for one study.

Considering the duration of teaching/learning and the use of social robots in the educational process, both short-term and long-term studies were carried out. Eight studies conducted educational activities for a period of two months to two years. Seven studies conducted educational activities as a one-time activity, while for one study the duration of the educational activity was not specified.

Considering RQ1, major fields of research were identified by taking into account the main findings and focus of the studies analysed: 1) Changes in learning; 2) Gains in knowledge; 3) Student-robot interaction; 4) Student engagement and enjoyment. In the next sections we briefly describe each of them along with the key findings.

4.1 Changes in learning

Konijn & Hoorn (2020) investigated the impact of the social robot on learning and the contribution of the robot's more social behaviour to the learning effect. They found that social robots are capable of creating a positive learning effect, but robot's more social behaviour did not add to learning. They were not teasing out a relationship, they tested if increasing the number of social cues was helpful in remembering.

Ekström & Pareto (2022) explored teachers' perceptions of a learning activity based on learning-by-teaching where the robot is designed to act as a didactic tool and a social actor. They found that robot-based learning activities can contribute to learning, develop skills, increase children's metacognitive awareness, and especially can increase children's interest, motivation, and participation. Robot as a didactic tool achieves relatively short-term goals with increased knowledge in a defined area. From the other perspective, the robot as a social actor has a rather long-term goal to support the general development of children.

Kim & Tscholl (2021) researched the advantages of an embodied social robot to engage children in play and learning from the perspective of embodied cognition. They found three embodied phenomena: the embodiment of early mathematics and scientific knowledge and reasoning, the appropriation of physical space, and the embodied collaboration. Embodiment occurs not only in thinking but also in social and emotional experiences. We can conclude that social robots have the potential to enable embodied learning experiences.

4.2 Gains in knowledge

Demir-Lira et al. (2020) showed in their research that children successfully learnt a foreign language from a social robot as an assistant, as well as they learnt from a human teacher. The observers expressed great excitement and enjoyment due to the social robot acting as an assistant, which is associated with the novelty and the anthropomorphic tendencies of the robot.

The use of social robots is most common in educational institutions, but examples show successful use in other areas of adult education as well. Connolly et al. (2022) researched using social robots in digital health for patient education. They observed significant improvements in patient knowledge, meaning the social robot provides health education effectively. Users (patients) reported high levels of acceptance and engagement with this method of intervention and reported low or slight levels of associated diabetes distress.

4.3 Student – robot interaction

It is usual for social robots to assist in the learning process in a variety of different ways, but there is a research on social reference in the context of a child–robot interaction. According to Tolksdorf et al. (2021) interaction with a social robot led to more social referencing in children compared to an interaction with a human partner and instances of social referencing of children did not decrease over the course of a long-term interaction.

Kim et al. (2021) investigated the use of a social robot to moderate interactions among culturally and linguistically diverse children, with a focus on design challenges and solutions to facilitate positive peer interactions. They proved that children could benefit from their design principles of using social robots for flexible children's exploration, friend-like communication, tasks relying on familiar experiences while stimulating imagination, and use of children's native languages.

The degree to which children anthropomorphise a robot as an assistant and whether this anthropomorphism relates to their learning of a foreign language were investigated by van den Berghe et al. (2020). Their results show the importance of taking children's anthropomorphism into consideration when designing robot - assisted learning sessions. van den Berghe et al. found that children generally anthropomorphised the robot and a weak but significant correlation was established between children's increased anthropomorphism and their word knowledge. Children who came to perceive the robot more as a human knew more words after the tutoring sessions.

Social robots need to be designed in a way that they are child-centred and collaborative because robot behaviours and collaboration paradigms affect children's perception about the robot. Escobar-Planas et al. (2022) prove in their research that the cognitive reliability of the robot shapes the helping relationship between the children and the robot, while the robot's expressivity impacts the perception of the robot's support ability and friendship. Children's perceptions about the robot remain individual, even if they interact with it in pairs, although a good collective task-performance seems to empower children's perception of the robot in terms of friendship and reliability.

Guggemos et al. (2020) investigate the acceptance of social robots by higher education students in the field of social sciences. The perceived characteristics of the social robot: trustworthiness, adaptiveness, social presence and appearance, indirectly predict the intention to use a social robot for learning purposes. The most important characteristic for predicting students' behavioural intention to use a social robot is its adaptiveness. An anxiety about making mistakes in handling the robot and privacy issues are not significant predictors.

The social robot as a reading companion to facilitate children's reading participation, as

alternatives to group storytelling activities in libraries, was investigated by Yueh et al. (2020). Their results showed that the children perceived the robot companion as more favourable and desirable to read with than a human co-reader. According to the results of the comparison, cognitively, it was found that human and robot companions facilitated the comprehension of the children's language in different ways and that the children performed similarly well with both types of companions. Affectively, the robot co-reader induced more social interaction during the reading sessions.

4.4 Student's engagement and enjoyment

Arar et al. (2021) researched the use of social robots in foreign language teaching and found that the use of social robots enhances the effectiveness of the educational process and significantly improves learning outcomes. Furthermore, given the facilities offered by the social robot through its support for foreign language learning to children, the authors recommend the use of social robots for improving the quality of learning outcomes and comfort in the schooling conditions.

According to Sisman et al. (2018) engagement is the most influential construct among all sources of attitude towards the integration of robots into the lesson. They have determined a high rate of correlation between engagement and intention as well as between enjoyment and intention. Furthermore, Sisman et al. (2018) highlighted that enjoyment is of particular significance to transform a teaching practise into a personalised experience of 'education', which, in turn, might influence the level of students' engagement in the lesson.

In the teaching process, a social robot can assume the role of an assistant and the role of a peer. The impact of peer learning with a social robot on children's learning and emotional engagement was investigated by Chen et al. (2020). They found that children who interact with the robot in the role of an assistant learnt more target vocabulary words than children who interact with the robot. Children interacting with the peer robot were more expressive in their facial affect display and more engaged than children interacting with the assistant robot, they learnt the most target words and showed the highest enjoyment.

Velentza et al. (2021) found that the social robot in the role of assistant increases the student's enjoyment. In their research, students had statistically significant higher scores in the enjoyment questionnaire in comparison with the human - teacher condition. It can be concluded that the presence of the robot played a key role in student motivation. The scores during the final exam confirmed that the presence of the robot motivated the students, as shown by the statistically significant higher scores in comparison to students who did not have any course with the robot as assistant.

Teachers' perceptions of the benefits and challenges, pedagogical practises that helped with student engagement, and where the social robots fit in the curriculum were investigated by Chalmers et al. (2022). Their findings show that the robot could be used to enhance the curriculum, from introducing robotics, coding and computational thinking to using the robot to engage students with foreign language and mathematics. Most teachers adopted a constructivist social approach to teaching with the social robot and reported that students were highly engaged with their learning when programming and working with the robot.

5 Discussion

Throughout literature review it was revealed that social robots are used mostly in preschool or primary education. The results also revealed that social robots, such as NAO, Pepper, and Skusie, have been used in various educational contexts, including language learning, STEM education, special education, and collaborative learning.

It was also noted that using social robots in education raises several ethical considerations and data privacy concerns. As technology advances and social robots become more prevalent in educational settings, it is essential to address these issues to ensure the responsible and safe use of these devices. Risks include lack of transparency, data privacy issues, dependency on robots, reduced human interaction, and potential job displacement (Boch et al. (2020)).

It is necessary to define how data will be collected and processed from social robots, to what extent this data should be stored or uploaded to the cloud, how to inform and get the users' enlightened consent to do so and how to prevent unauthorised external actors from accessing personal detailed information (Boch et al. (2020). According to the same authors, due to new technological features, allowing robots to collect more data in their environment than before, the creation of particular regulations need to be considered. All of the above issues raise the question of regulations and public policy formulations for social robots, which will be an important subject of study and research in the near future (Subramanian (2017)).

Collaboration among educators, policymakers, developers, and privacy experts is crucial to establish clear guidelines and policies that priorities students' well-being and rights (Boch et al. (2020)).

In respect to the first research question (RQ1), the research highlighted major fields of research where social robots can be used effectively in education, including personalized learning, language learning, STEM education, special education, collaborative learning, emotional support, virtual field trips, and coding and robotics education. These applications demonstrate the versatility of social robots in satisfying the diverse educational needs and contexts and further

support the claim that robots can be implemented in various contexts (Kalaitzidou & Pachidis, 2023) supporting the learning experiences of students (Belpaeme et al. (2018); Kubilinskiene et al. (2017)).

Related to the second research question (RQ2), the results show that the integration of social robots in education has the potential to enhance students' engagement, provide personalized support, and foster the development of various skills. Social robots in education contribute to the child's social development (Ekström & Pareto, 2022)" improve engagement, and support their learning experiences (Belpaeme et al. (2018); Kubilinskiene et al. (2017); (Chen et al. (2020); (Chalmers et al.(2022); (Sisman et al. (2018)). According to Escobar-Planas et al. (2022), social robots can offer emotional support to students. The robot's presence motivated the students and increased the student's enjoyment (Velentza et al. (2021)), enhanced the effectiveness and improved learning outcomes (Arar et al. (2021)) This is inline with other studies that have shown that the use of advanced technology in education has a positive effect on student motivation (Francis (2017); (Yang et al. (2021). On the other hand, research shows that students who are intrinsically motivated and those who are motivated by attaining high grades tend to engage more with learning using advanced technology (Dunn & Kennedy (2019); (Stockwell & Reinders (2019)). At the same time, the role of teachers and their attitudes can also impact technology implementation (Stockwell & Reinders (2019)).

6 Conclusion

This literature review has been focused on exploring the impact of social robots as teaching assistants on student engagement and motivation in learning activities. Following a structured methodology, a total of 16 relevant studies were included in the analysis. These studies explored various aspects of social robot implementation, such as the type of robot used, the number and age of participants, the educational fields covered, and the duration of educational activities.

The findings indicated that social robots, particularly NAO, are commonly used in educational settings, with a focus on teaching foreign languages and STEM subjects. The studies showed positive outcomes in terms of learning gains, increased engagement, and improved knowledge acquisition when social robots were employed as teaching assistants. Children's interaction with social robots was found to be beneficial, leading to enhanced social referencing, positive peer interactions, and increased word knowledge.

Furthermore, the study emphasized the importance of ethical considerations, data privacy, and the role of human educators in the learning process when implementing social robots in educational settings. It is crucial to retain a balance between the use of technology and human interaction to ensure an optimal learning experience for students.

In future research, it is recommended to explore additional aspects of social robot implementation, address ethical concerns, and investigate long-term effects on student engagement and motivation, but also to expand the research on the use of social robots in education affecting knowledge construction, self-regulation, and development of communication and cooperation skills, development of critical thinking and real problem-solving skills. By harnessing the power of advanced technologies like social robots, teachers can create dynamic and personalized learning environments that inspire and empower students.

References

Arar, C., Belazoui, A., Telli, A., (2021). Adoption of social robots as pedagogical aids for efficient learning of second language vocabulary to children. Journal of e Learning and Knowledge Society 17 (3), 119. – 126., doi: 10.20368/19718829/1135551

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. Science Robotics, 3, 1-9.

Boch, A., Lucaj, L., & Corrigan, C. (2020). A robotic new hope: Opportunities, challenges, and ethical considerations of social robots. *Technical University of Munich*, 1-12.

Boland, A., Cherry, G., & Dickson, R. (2017). Doing a Systematic Review: A Student's Guide. Sage.

Chalmers, C., Keane, T, Boden, M, & Williams, M. (2022). Humanoid robots go to school. Education and Information Technologies (2022) 27:7563–7581, doi: 10.1007/s10639-022-10913-z

Chen, H., Park, H. W., & Breazeal, C., (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on child learning and emotional engagement. Computers & Education (2020), doi: 10.1016/j.compedu.2020.103836.

Connolly, C., Walsh, J. C., Worlikar, H., Ryan, L., Murray, A., O'Connor, S., Kelly, J., Coleman, S., Vadhira, V. V., Newell, E., & O'Keeffe, D. T., (2022) Exploring new frontiers of education using humanoid robots – a case study of patient centred innovation in digital health education, Irish Educational Studies, 41:1, 107-115, doi: 10.1080/03323315.2021.2022514

Demir-Lira, Ö. E., Kanero, J., Oranç, C., Koskulu, S., Franko, I., Göksun, T. & Küntay, A. C., (2020). L2 Vocabulary Teaching by Social Robots: The Role of Gestures and On-Screen Cues as

Scaffolds. Front. Educ. 5:599636., doi: 10.3389/feduc.2020.599636

Donnermann, M., Schaper, P., & Lugrin, B. (2020). Integrating a social robot in higher education–a field study. In Proceeding of the 29th IEEE International Conference on Robot and Human Interactive Communication (pp. 573–579). IEEE. doi: 10.1109/RO-MAN47096.2020.9223602

Dunn T.J. & Kennedy M., Technology Enhanced Learning in higher education; motivations, engagement and academic achievement, Computers & Education (2019), doi: 10.1016/j.compedu.2019.04.004.

Ekström, S. & Pareto L. (2022). The dual role of humanoid robots in education: As didactic tools and social actors. Education and Information Technologies (2022) 27:12609–12644. doi: 10.1007/s10639-022-11132-2

Escobar-Planas, M., Charisi, V., and Gómez, E. (2022.) "That Robot Played with Us!" Children's Perceptions of a Robot after a Child-Robot Group Interaction. Proc. ACM Hum.-Comput. Interact. 6, CSCW2, Article 393 (November 2022), 23 pages. doi: 10.1145/3555118

Francis, James, "The Effects Of Technology On Student Motivation And Engagement In Classroom-Based Learning" (2017). All Theses And Dissertations. 121.

Guggemos, J., Seufert, S., & Sonderegger, S., (2020), Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT, British Journal of Educational Technology, 51(5), 1864-1883., doi: 10.1111/bjet.13006

Kalaitzidou, M., Pachidis, T.P. (2023). Recent Robots in STEAM Education. Educ. Sci. 13, 272. doi: 10.3390/educsci13030272

Kim, Y., Marx, S., Pham, H. V., & Nguyen, T. (2021). Designing for robot-mediated interaction among culturally and linguistically diverse children. Educational Technology Research and Development, 69, 3233-3254., doi: 10.1007/s11423-021-10051-2

Kim, Y., & Tscholl, M. (2021). Young children's embodied interactions with a social robot. Educational Technology Research and Development, 69, 2059-2081., doi: 10.1007/s11423-021-09978-3

Konijn, E.A. & Hoorn, J.F., (2020). Robot tutor and pupils' educational ability: Teaching the times tables, Computers & Education (2020), doi: 10.1016/j.compedu.2020.103970.

Kubilinskiene, S., Zilinskiene, I., Dagiene, V., & Sinkevièius, V. (2017). Applying robotics in school education: A systematic review. Baltic Journal of Modern Computing, 5, 50. doi: 10.22364/bjmc.2017.5.1.04

Leitão, R., Maguire, M., Turner, S., Guimarães, L. (2022): A systematic evaluation of game elements effects on students' motivation, Education and Information Technologies (2022) 27:1081–1103, doi: 10.1007/s10639-021-10651-8

Lopez-Caudana, E., Lopez-Orozco, C. F., Bárbara, C. M., Reyes, G. E. B., Ponce, P., & Chong-Quero, J. E. (2021). Physical therapy using robotics: A project-based learning experience for undergraduate students. Australasian Journal of Educational Technology, 37(5), 32-42.

Oh, Y., Parasuraman, R., McGraw, T., & Min, B. C. (2018, March). 360 VR based robot teleoperation interface for virtual tour. In Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI).

Rojas-López, A., Rincón-Flores, E. G., Mena, J., García-Peñalvo, F. G., Ramírez-Montoya, M. S. (2019): Engagement in the course of programming in higher education through the use of gamification, Universal Access in the Information Society, doi: 10.1007/s10209-019-00680-z

Sisman, B., Gunay D. & Kucuk S. (2018). Development and validation of an educational robot attitude scale (ERAS) for secondary school students, Interactive Learning Environments, doi: 10.1080/10494820.2018.1474234

Stockwell, G., & Reinders, H. (2019). Technology, Motivation and Autonomy, and Teacher Psychology in Language Learning: Exploring the Myths and Possibilities. Annual Review of Applied Linguistics, 39, 40-51. doi:10.1017/S0267190519000084

Subramanian, Ramesh (2017) "Emergent AI, Social Robots and the Law: Security, Privacy and Policy Issues," Journal of International Technology and Information Management: Vol. 26 : Iss. 3 , Article 4.

Tolksdorf, N. F., Crawshaw, C. E. & Rohlfing, K. J. (2021). Comparing the Effects of a Different Social Partner (Social Robot vs. Human) on Children's Social Referencing in Interaction. Front. Educ. 5:569615. doi: 10.3389/feduc.2020.569615

van den Berghe, R., de Haas, M., Oudgenoeg-Paz, O., Krahmer, E., Verhagen, J., Vogt, P., ... & Leseman, P. (2021). A toy or a friend? children's anthropomorphic beliefs about robots and how these relate to second-language word learning. Journal of Computer Assisted Learning, 37(2), 396-410., doi: 10.1111/jcal.12497

Velentza, A. M., Fachantidis, N., & Lefkos, I. (2021). Learn with surprize from a robot professor,

Computers & Education, (2021), doi: 10.1016/j.compedu.2021.104272

Yang, K.-H., Chen, H.-H. (2021): What increases learning retention: employing the prediction-observation-explanation learning strategy in digital game-based learning, Interactive Learning Environments, doi: 10.1080/10494820.2021.1944219

Yueh, H. P., Lin, W., Wang, S. C., & Fu, L. C. (2020). Reading with robot and human companions in library literacy activities: A comparison study. British Journal of Educational Technology, 51(5), 1884-1900., doi: doi:10.1111/bjet.13016

Enhancing Learning Design through User Experience Research: Insights from a Survey in Four European Countries

Blaženka Divjak, Josipa Bađari, Darko Grabar, Damir Horvat, Petra Vondra
University of Zagreb
Faculty of Organization and Informatics
{blazenka.divjak, josipa.badjari, darko.grabar, damir.horvat1, petra.vondra}@foi.unizg.hr

Bart Rienties
The Open University
bart.rienties@open.ac.uk

Abstract. *In this paper, we present preliminary results from a survey conducted in four European countries, focusing on learning design, specifically on the Balanced Design Planning (BDP) concept and tool. Our objective is to showcase the further development of the BDP concept and tool, based on user experience research. The survey involved 53 higher education educators and aimed to identify the key elements of effective learning design. Our findings emphasize the importance of learning outcomes and constructive alignment as crucial components of learning design. Additionally, we highlight the importance of design analytics and training opportunities in supporting educators in the preparation of sound learning designs. Through this research, we aim to contribute to the advancement of learning design practices and the improvement of the BDP concept and tool.*

Keywords. learning design, learning analytics, design analytics, learning outcomes, constructive alignment

1 Introduction

Learning design (LD) emerged in the early 2000s as a field that recognized the potential of the Internet for documenting and sharing effective educational practices (Lockyer, 2013). The rapid development of educational technology and the pressing challenges posed by societal shifts and the COVID-19 pandemic have placed a greater emphasis on digital education and resource sharing, making LD a highly relevant and vital topic (Divjak et al., 2022).

LD lacks a single definitive definition, but Koper and Olivier (2004) describe it as the application of a pedagogical model that aligns with learning objectives, while considering learner specifics, the learning context, and the subject of study (p. 98). Lockyer et al. (2013) further define LD as a practice that involves sequencing teaching practices, resources, and support developed by educators, while Conole (2012) views LD as a methodology that guides educators in making informed decisions about designing learning activities. Similarly, Bennet et al. (2015) emphasize the learner-centered nature of LD, highlighting the need for educators to design learning experiences based on students' needs and feedback.

Moreover, recent research by Rienties et al. (2023a) underscores the importance of investigating the conceptual and technological aspects of LD and their connections to learning analytics (LA). To address these considerations, Divjak et al. (2022) have introduced a new LD concept called Balanced Design Planning (BDP), supported by a freely available tool. BDP places a strong focus on contemporary learning theories, constructive alignment, aligning learning outcomes with teaching and learning activities in accordance with Biggs' constructive alignment theory (2006) and leveraging design LA.

This paper aims to present a portion of the preliminary findings from a survey amongst 53 educators conducted in four European countries (Croatia, Germany, Finland, the UK) during 2023, which sought to explore LD, particularly the BDP concept and tool and their further development based on user experiences. Through this research, we endeavor to contribute to the advancement of LD practices and provide insights into how educators can enhance their LD approaches to create effective and engaging learning experiences. The research was developed around the following research questions:

RQ1: What are the perspectives and experiences of different user groups in diverse contexts with LD and particularly the BDP concept and tool?

RQ2: What are the essential needs and areas for further improvement of the BDP concept and tool, depending on these different user groups and diverse contexts?

2 Background

The Balanced Design Planning (BDP) concept and tool (https://learning-design.eu/) have been developed through a rigorous process utilizing the design science methodology, incorporating insights from the Open University LD and contemporary research (Divjak et al., 2022; Rienties et al. 2023b). The initial phases of this work encompassed a comprehensive needs analysis, an extensive literature review and an exploration of the existing LD concepts and tools.

Subsequently, the development and initial enhancements of the BDP concept and tool were carried out in the second phase. To ensure its effectiveness and relevance, the validation activities were conducted primarily by higher education (HE) educators within sever.al Erasmus+ projects (RAPIDE, eDESK, Teach4Edu4 and iLed). Figure 1 illustrates the progression of this process, highlighting the piloting phase where the BDP concept and tool were put into practice. This study represents a step in the treatment validation phase of the design science cycle (Divjak et al., 2022).

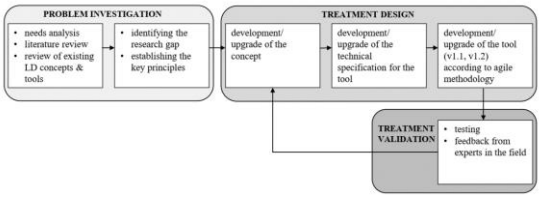


Figure 1. The BDP concept and tool design process. Source: Divjak et al. (2022). Balanced Learning Design Planning: Concept and Tool.

Currently, the BDP tool boasts of its impressive user base including over 1200 individuals from more than 30 countries worldwide. These educators represent diverse backgrounds, including schools, higher education institutions, lifelong learning providers, and industry professionals. The widespread adoption of the BDP tool underscores its versatility and applicability across various educational contexts and sectors.

In general, each course LD in BDP tool has four parts: Course Details, Planning, Analysis and Export (Figure 2).

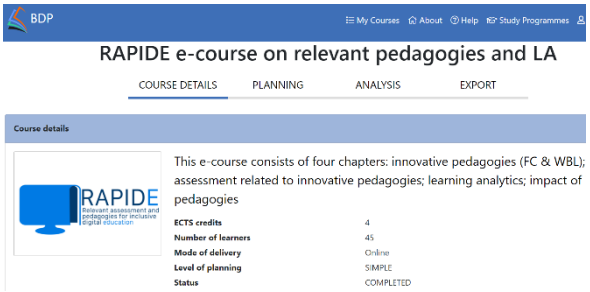


Figure 2. First page of a LD of a course in the BDP tool

3 Research design/Methodology

As part of the Erasmus+ project iLed (Learning Design in Higher Education), we conducted a comprehensive evaluation and collected feedback from a diverse range of users, including educators, instructional designers, technical experts, researchers, curriculum developers and educational decision-makers. This research aimed to address the specific needs of different international educational contexts. A wealth of studies have shown that implementing new EdTech in different cultures is not a one-size-fits-all approach (Reinecke & Bernstein, 2013; Rienties et al., 2023b; Rizvi et al., 2022). Indeed, recent preliminary findings from one study (Rienties et al., 2023b) using BDP showed subtle differences between how 86 Croatian, German, and UK educators reflected on the BDP tool and their subsequent intentions to implement it in their own context (Rienties et al., 2023b). Therefore, a research protocol was developed, approved by the relevant ethical committee, implemented and made publicly available on the iLed project website (iled-project.eu/).

While LD is not inherently tied to any specific learning theory or teaching approaches, the BDP concept and tool encourage educators to incorporate contemporary learning theories, such as constructivism, social constructivism, and connectivism, as well as innovative learning and teaching approaches. Therefore, the overarching goal of this study within the innovative iLed project was to further implement learning outcomes (LOs) and student-centered curricula through the development of an innovative LO-based LD concept and tool.

The study, led by the University of Zagreb's Faculty of Organization and Informatics (FOI) from Croatia, in collaboration with the Open University (UK), Goethe University (Germany), the University of Oulu (Finland) and the University of Zagreb's School of Medicine, began with the preparation of a research protocol and questionnaire.

The next step involved engaging evaluators from various user groups at each partner institution. These evaluators were introduced to the BDP concept and tool, within a so-called iLed onboarding e-course (https://learn.foi.hr/course/index.php?categoryid=9). There is some emerging evidence that onboarding could provide an effective approach to help educators to learn a new concept or approach (Frögéli et al., 2023). The evaluators were defined as users who had engaged with the BDP learning materials (videos and other learning materials in the iLed onboarding e-course or RAPIDE e-course) and/or had utilized the BDP tool to prepare LD of a course. Additionally, evaluators could also be users who had already used the BDP tool for at least two courses or for four hours. Once familiarized with the concept and tool, the evaluators completed the questionnaire and the data were collected and analyzed.

The evaluation focused on the diffusion of the LD concept and the acceptance of the BDP tool. To

measure technology acceptance, we utilized the Unified Theory of Acceptance and Use of Technology (UTAUT) framework (Venkatesh et al., 2003), incorporating specific questions related to the BDP concept and tool, as well as several open-ended questions.

The questionnaire preparation involved three stages (Figure 3), with the initial draft prepared by FOI in line with the conclusions from the iLed partners' meeting in December, 2022. Feedback was received from the project partners, leading to the preparation of the second version. The final questionnaire consisted of 36 questions, including 11 questions from UTAUT, 8 LD BDP tool-specific questions, seven open-ended questions, and questions related to socio-demographic data.

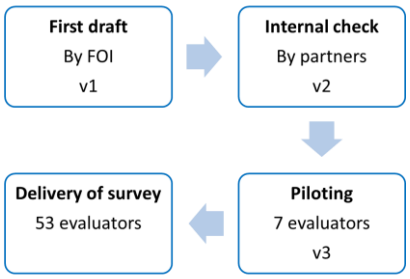


Figure 3. Stages of the preparation and delivery of the questionnaire

Prior to the survey delivery, the Ethical approvals were sought and obtained from the Ethical Committee at FOI and the OU Human Research Ethics Committee, encompassing the research protocol and questionnaire. Informed consent forms were added to the BDP tool, along with terms of service and Data Privacy Notice.

4 Results

In the survey conducted from February 2023 to May 2023, a total of 53 evaluators participated. Figure 4 provides an overview of their affiliations. The evaluators represented a diverse range of user groups, ensuring comprehensive coverage. While the majority of evaluators were educators and researchers, there was also active participation from instructional designers, learning designers, educational decision-makers, curriculum developers and technical experts.

As indicated in Figure 5, most educators were in general positive about the BDP tool and were broadly supportive in terms of technology acceptance. Overall, on average 46% of educators indicated a positive perspective about the BDP, while 6 of educators were neutral, and only 1 of educators was negative. It is noticeable that most educators (93%) believed that planning of assessment based on learning outcomes contributed to the quality of their work (LD_BDP3). Relatively the least support was given to the notion of weighted learning outcomes (LD BDP 1) whereby only 73% of educators supported this notion.

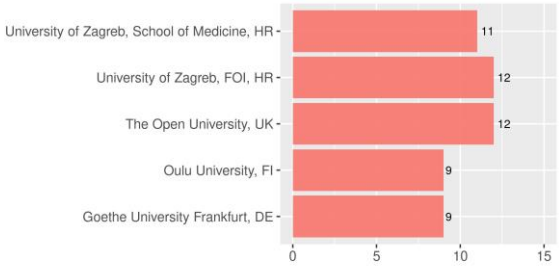


Figure 4. Number of evaluators per iLed partner institution

Here we will present a range of preliminary results from the survey that are directly related to upgrading of the BDP concept and tool (8 questions and open-ended questions).

The questionnaire included the following questions:

1. LD BDP 1: In my opinion, a learning design that is based on WEIGHTED learning outcomes improves the quality of my work as an educator.
2. LD BDP 2: I believe that planning of teaching and learning activities based on learning outcomes contributes to the quality of my work as an educator.
3. LD BDP 3: I believe that planning of assessment based on learning outcomes contributes to the quality of my work as an educator.
4. LD BDP 4: I believe that the choice of analysis available in the BDP Analysis (part of the tool) contributes to the quality of my work.
5. LD BDP 5: I find planning of teaching and learning activities in the BDP tool useful.
6. LD BDP 6: I find data presentation and visualization in the BDP Analysis (part of the tool) understandable.
7. LD BDP 7: I find data presentation and visualization in the BDP Analysis (part of the tool) useful.
8. LD BDP 8: I find the export possibilities (to pdf and to excel) useful for productivity of my work.

Open-ended questions:

1. FUNC USEF+: Which functionalities of the BDP tool do you find the most useful?
2. FUNC USEF-: Which functionalities of the BDP tool do you find the least useful?
3. AVAR -: If you would imagine the average user, what do you think they will struggle most with?
4. MISS <6: Which functionality/ies do you really miss in the BDP tool that we should prioritize in the next 6 months of development?
5. MISS >6: What are your suggestions for further improvements of the BDP tool beyond the 6 months period?
6. What type of evidence will you seek or consider when deciding whether to continue using your BDP learning design in future years?
7. Please, add any other comment about the BDP concept and tool.

The analysis of the qualitative answers on open-ended questions was performed as follows: Review of

answers; Defining categories in line with reviewed answers (Course details, Planning, Analysis, Export, Constructive alignment, Learning outcome (defining, weighting), Sharing/Collaboration, Advanced option, Concept (understanding, terminology, structure), Help (tooltips, literature, videos…), Learning types, User interface and technical implementation, General) and Classification of answers into categories (Appendix: Table 1).

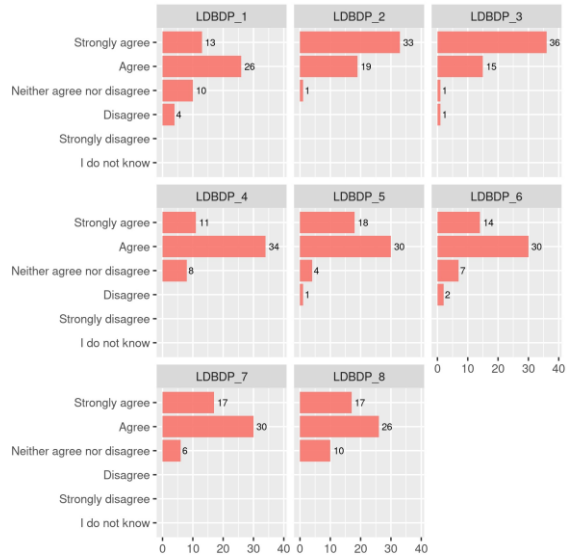


Figure 5. Frequencies of answers on the BDP tool related questions

Comments on FUNC USEF+ "Planning takes time, but quickly shows me the results." "Visualization which enables connection of learning outcomes with the weight of TLA." (Participant 46, Professor Croatia), "The possibility of linking learning outcomes to both activities and assessments." (Participant 2, Researcher, UK), "Planning the TLA units and comparing to the analysis dashboard." (Participant 37, Educational decision-maker, Finland), "Workload visualizer - succinct breakdown of activity types and timings (Participant 34, Instructional designer, learning designer, UK)".

FUNC USEF- "I think they're all useful, but I do wonder how we estimate the timings for activities. (Participant 42, Instructional Designer, UK) ", "I am confused by the concept of difficulty of learning outcomes"(Participant 10, Professor, Croatia), "For me personally, so far I'm not really sure how I could effectively use the analysis part" (Participant 30, Researcher, Finland), "Difficult to comment as I did not understand all of the training"(Participant 31, Researcher, UK).

AVAR- "Creating and weighting learning outcomes" (Participant 34, Instructional Designer, UK), "There's clearly a learning curve, to get into the model of working initially, understand the activity types and what type of data to input. This may differ to the user's mental model of course creation or their past experience. A possible way to provide additional support could be to provide examples or case studies." (Participant 12, Educational decision-makers, UK)," Drawing conclusions from the graphs to adapt the teaching" (Participant 40, Educational decision-makers, Germany), "constructive alignment with LOs and estimation of student workload" (Participant 17, Researcher, Croatia).

MISS<6 "It should be possible in the BDP Tool at the TLA level to assign specific tools or media that are used to implement the TLA.... Ideally, these should be activities that can then also be found in the Moodle …." (Participant 11, Instructional designer, Germany) , "Collaboration and feedback from students" (Participant 25, Curriculum Developer, UK) , "The alignment of the BDP tool with the constructive alignment approach (this is only somewhat achieved in the advanced option, but this is not the default option when creating the course)" (Participant 4, Curriculum developer, Croatia), "A wizard or chatbot that guides you in formulating LOs would be very useful. For example, with instructive verbs to choose from that correspond to elements from Bloom (Participant 49, Researcher, Germany)", "... to have also teacher workload, which depends on learning activities that are chosen for the course." (Participant 29, Researcher, Finland).

MISS>6: "Indication if there is a bigger inconsistency in learning design - e.g. weight of learning outcome is very different from the investment of student workload or assessment points. (Participant 9, Researcher, Croatia)", "Automated creation of a Moodle course with the planned activities from the BDP tool (Participant 47, Instructional designer, Germany)".

Other comments: "It's a great looking tool, and I like the way it channels you towards considering activities and constructive alignment." (Participant 8, Instructional Designer, UK), "The BDP concept is good and doing the course has made me aware of the importance of careful alignment of the learning outcomes to the learning design. It also adds evidence to the efficacy of chosen learning activities and where changes may be needed. Also, it helps to evaluate the student experience." (Participant 13, Educator, UK), "Ease of the new course design by utilizing the templates available with the tool." (Participant 46, Educator, Croatia), "evidence for better efficiency of student learning outcome achievement compared to standard classical teaching methods" (Participant 32, Curriculum Developer, Croatia), "The aspect of supervision by the teacher should be given greater consideration (not just: teacher is present vs. is not present). Especially in online settings… (Participant 11, Instructional Designer, Germany)".

5 Discussion

Diversity of evaluators enriched the study by incorporating various perspectives and expertise from different educational roles and domains.

RQ1: What are the perspectives and experiences of different user groups in diverse contexts with LD and particularly the BDP concept and tool?

In general, educators of different user groups and contexts were satisfied with the BDP concept and tool (RQ1). Even though they emphasized that the planning takes time they are happy with the functionalities, analysis, visualizations and collaboration and enhancement possibilities. Nevertheless, there are areas for further improvement of the BDP concept and tool (RQ2). Quantitative analysis revealed that while the concept of LD based on weighted LOs is accepted, it is not fully implemented in practice. Some educators do not utilize the LO-based approach, despite it being the foundation of student-centered teaching. They often struggle with prioritizing learning outcomes. Although many evaluators expected the LO-approach to be challenging for their colleagues, they recognized the value of planning teaching and learning activities (TLA), including assessment, based on learning outcomes in enhancing the quality of their work. However, some evaluators expressed concerns about the time-consuming nature of TLA planning and suggested the introduction of shortcuts or pre-defined templates.

Some evaluators pointed out the difficulty with shifting from educator-workload mode and scheduled activities such as lectures, seminars and tutorials to student workload mode and estimation of time an average student needs to fulfill the task.

Regarding learning analytics, specifically the design analytics available in the BDP tool, evaluators largely believed that the choice of analysis contributes to the quality of their work. However, some faced difficulties in understanding data presentation and visualization, as well as the perceived usefulness of design analytics. In open-ended responses, evaluators frequently mentioned the Analysis part as one of the most useful functions, expressing their concerns about their own and their colleagues' ability to utilize analytics for constructive alignment and assessment improvements, and seeking improvements in clarity and additional support. Here we recognize need for further development of educators' data and assessment literacy.

It is important to emphasize that aligning an assessment program with weighted intended learning outcomes is crucial for ensuring its validity. Assessment literacy is essential for educators and other stakeholders in education to competently evaluate the acceptability of assessments and effectively utilize assessment learning analytics for improving learning and teaching. Finally, evaluators found the export capabilities (to PDF and Excel) useful for enhancing their productivity.

Some evaluators pointed out that they share the LD with students and that it can be useful to collect students' feedback directly in the BDP tool.

Finally, the export possibilities (to pdf and to excel) they found useful for productivity of their work. In this context, some mentioned quality assurance and accreditation processes.

In answers in open-ended questions, evaluators often suggested that training opportunities need to be broaden and essential concepts related to learning design (LOs, constructive alignment, validity of assessment etc.) further explained.

6 Concept of the BDP LD

The BDP LD is built around several main concepts: learner-centeredness, constructive alignment, feedback to learners, validity and acceptability of assessment as well as technology supported learning (Table x).

Basic building elements include learning outcomes as a focal point, content and structural notions such as topics, units and TLA, types of TLA (acquisition, discussion, practice, investigation, production, assessment), type of assessment, duration and time on tasks as well as technology and delivery plans.

We define enablers as tools, skills and support that enable implementation of main concept for designing learning by using the basic building elements. Enablers include design analytics based on prioritization of LOs, educators pedagogical and digital skills, necessary resources for implementation of meaningful LD, assessment literacy of educators and learners, use of credits (e.g. ECTS) for measuring student workload as well as data, digital and AI literacy of all educational stakeholders. Stakeholders encompass learners, educators, educational leaders and authorities as well as industry and socially at large. They need to co-create quality learning based on LOs by using LA, adequate training means and resources supported by technology.

Finally, for the quality assurance of learning design as well as learning and teaching all stakeholders and interfaces to other systems need to be assured.

The COVID-19 pandemic has emphasized the significance of innovative pedagogical approaches such as Inquiry-Based Learning and Flipped Classrooms (FC) (Divjak et al., 2022c).

Therefore, learning design in digital era demands highly competent educator that can integrate innovative pedagogies with technologies (e.g. AI) and work in the international environment. Therefore, Rienties et al. (2023a) highlight the significance of cross-institutional and cross-national online professional development initiatives. Collaborative efforts among educators facilitate the sharing of best practices and innovative approaches, leading to continuous improvement in academic teaching.

Students' acceptance of learning design is crucial and therefore student feedback on learning design need

to be assured. For that, different means can be used such as direct students' access to LD tool providing feedback or integration of LD into learning analytics dashboards. Namely, students highly value features that aid short-term planning and organization of learning, (Divjak et al., 2023b) and therefore could be motivated to provide feedback. The upgraded BDP learning design concept is presented on Figure 6.

Table 2. Learning design concept elements

Basic building elements	**Main concepts**	**Enablers**	**Stakeholders' activities and interfaces**
LO	Learner-centeredness	Design analytics Prioritization of LOs	Educators' collaboration with peers
Topic	Constructive alignment	Pedagogical skills Digital pedagogies	Learners' feedback on learning design
TLA Type of TLA	Feedback to learners	Resources	"Interoperability" with LMS (e.g. Moodle)
Assessment	Validity and acceptability of assessment	Assessment literacy	Learning analytics integrated with design analytics
Time on task	Student workload	Credits (e.g. ECTS)	Training possibilities and support (e-course, OER, FAQ)
Technology and delivery models	Technology supported learning (TSL)	Data, digital and AI literacy	QA of teaching and learning (TL) in cooperation with educational leaders and society

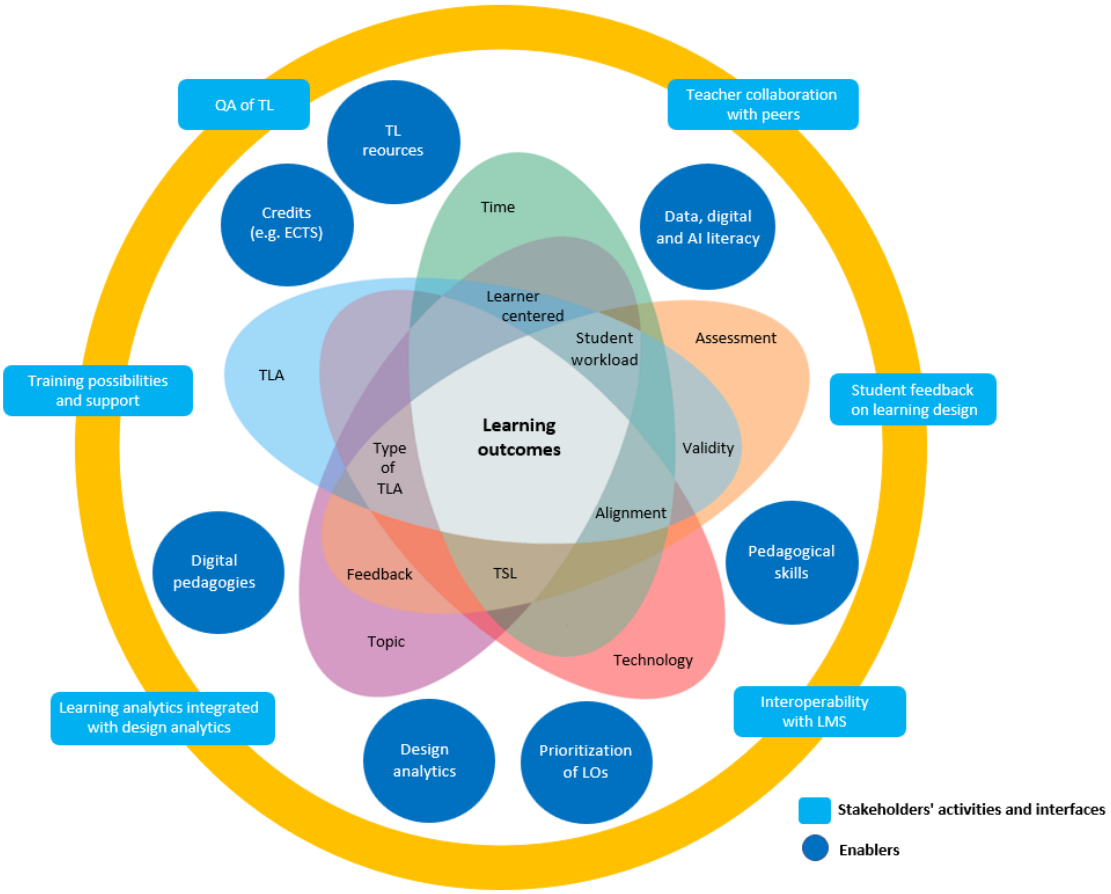


Figure 6. The upgraded BDP learning design concept

7 Conclusion

The concept of learning design (LD) lacks universal acceptance, as there are various approaches and perspectives within the field. However, in our study, we have developed a specific LD concept and tool called Balanced Design Planning (BDP) based on contemporary learning theories, learner-centered approaches, and the constructive alignment of intended learning outcomes (LOs) with teaching and assessment. The BDP concept emphasizes the importance of designing learning experiences that align with intended LOs and promote effective instructional strategies.

In our study, within Erasmus+ iLed project, we conducted a thorough user experience analysis of the BDP concept and tool. This report presents preliminary findings regarding the feedback received from 53 educators across four European countries. The aim of the analysis was to identify potential areas for improvement in the BDP based on their input.

Findings pointed out that more clarity in learning design elements, concepts and enablers are needed.

It also highlights the role of design analytics and resources in supporting educators in their LD practices. Continuous educator training, especially in areas such as LD and assessment literacy, is essential for informed decision-making and enhancing student learning outcomes.

The upgraded BDP learning design concept is proposed.

Acknowledgement

This work was supported by the project “iLed - Innovating Learning Design in Higher Education”, financed from the Erasmus+ Programme of the European Union, within KA220-HED - Cooperation partnerships in higher education (2022-1-HR01-KA220-HED-000085818). The sole responsibility for the content of this article lies with the authors. It does not necessarily reflect the opinion of the European Union.

References

Busulwa, H. S., & Bbuye, J. (2018). Attitudes and coping practices of using mobile phones for teaching and learning in a Uganda secondary school. Open Learning: *The Journal of Open, Distance and e-Learning*, 33(1), 34–45. https://doi.org/10.1080/02680513.2017.1414588

Chiu, T. K. F. (2017). Introducing electronic textbooks as daily-use technology in schools: A top-down adoption process. *British Journal of Educational Technology*, 48(2), 524–537. https://doi.org/10.1111/bjet.12432

Crook, S. J., & Sharma, M. D. (n.d.). Bloom-ing Heck! The Activities of Australian Science Teachers and Students Two Years into a 1:1 Laptop Program Across 14 High Schools. 16.

Divjak, B., Rienties, B., Iniesto, F., Vondra, P., & Žižak, M. (2022). Flipped classrooms in higher education during the COVID-19 pandemic: Findings and future research recommendations. International *Journal of Educational Technology in Higher Education*, 19(1), 9. https://doi.org/10.1186/s41239-021-00316-4

Dudaitė, J., & Prakapas, R. (2019). Influence of use of Activinspire interactive whiteboards in classroom on students’ learning. *Digital Education Review*, 299–308. https://doi.org/10.1344/der.2019.35.299-308

Fouh, E., Breakiron, D.A., Hamouda, S., Farghally, M.F., Shaffer,C.A., (2014) Exploring students learning behavior with an interactive etextbook in computer science courses. *Computers in Human Behavior*, 41. 478-485. https://doi.org/10.1016/j.chb.2014.09.061.

Gerard, L., Wiley, K., Debarger, A. H., Bichler, S., Bradford, A., & Linn, M. C. (2022). Self-directed Science Learning During COVID-19 and Beyond. *Journal of Science Education and Technology*, 31(2), 258–271. https://doi.org/10.1007/s10956-021-09953-w

Grimalt-Álvaro, C., Ametller, J., & Pintó, R. (2019). Factors Shaping the Uptake of ICT in Science Classrooms. A Study of a Large-Scale Introduction of Interactive Whiteboards and Computers. *International Journal of Innovation in Science and Mathematics Education*, 27(1). https://doi.org/10.30722/IJISME.27.01.002

Grönlund, Å., Wiklund, M., & Böö, R. (2018). No name, no game: Challenges to use of collaborative digital textbooks. Education and Information Technologies, 23(3), 1359–1375. https://doi.org/10.1007/s10639-017-9669-z

Håkansson Lindqvist, M. (2019). Talking about digital textbooks. The teacher perspective. The International *Journal of Information and Learning Technology*, 36(3), 254–265. https://doi.org/10.1108/IJILT-11-2018-0132

Hermita, N., Wijaya, T. T., Yusron, E., Abidin, Y., Alim, J. A., & Putra, Z. H. (2023). Extending unified theory of acceptance and use of technology to understand the acceptance of digital textbook for elementary School in Indonesia. *Frontiers in Education*, 8, 958800. https://doi.org/10.3389/feduc.2023.958800

Joo, Y. J., Park, S., & Shin, E. K. (2017). Students' expectation, satisfaction, and continuance intention to use digital textbooks. *Computers in Human Behavior*, 69, 83–90. https://doi.org/10.1016/j.chb.2016.12.025

Bennett, S., Agostinho S. and Lockyer L., "Technology tools to support learning design: Implications derived from an investigation of university teachers' design practices", Computers & Education, 81, 211–220, 2015. Available: https://doi.org/10.1016/j.compedu.2014.10.016.

Biggs J., "What the Student Does: teaching for enhanced learning", Higher Education Research & Development, 18(1), pp. 57–75, 1999. Available: https://doi.org/10.1080/0729436990180105. G.

Conole, R., Designing for Learning in an Open World, Springer New York, 2013. Available: https://doi.org/10.1007/978-1-4419-8517-0.

Koper, R., Olivier, B., Representing the Learning Design of Units of Learning, Educational Technology & Society, Vol. 7, No. 3 (July 2004), pp. 97-111.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13 (3): 319–340. doi:10.2307/249008, JSTOR 249008, S2CID 12476939

Divjak, B., Kadoić, N. & Žugec, B. (2021) The Use of Decision-Making Methods to Ensure Assessment Validity. U: 2021 IEEE Technology & Engineering Management Conference - Europe (TEMSCON-EUR) (TEMSCON-EUR 2021).

Divjak, B., Grabar, D., Svetec, B. & Vondra, P. (2022) Balanced Learning Design Planning: Concept and Tool. Journal of information and organizational sciences, 46 (2), 361-375 doi:10.31341/jios.46.2.6. Available at: https://jios.foi.hr/index.php/jios/article/view/1742

Divjak, B., Svetec, B., Horvat, D. & Kadoić, N. (2023) Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design. British journal of educational technology, 54 (1), 313-334 doi:10.1111/bjet.13290.

Divjak, B., Rienties, B., Iniesto, F., Vondra, P., & Žižak, M. (2022). Flipped classrooms in higher education during the COVID-19 pandemic: Findings and future research recommendations. International Journal of Educational Technology in Higher Education, 19(1), 9. doi:10.1186/s41239-021-00316-4.

Divjak, B., Svetec, B., & Horvat, D. (2023). Learning analytics dashboards: What do students actually ask for? In I. Hilliger, H. Khosravi, B. Rienties, & S. Dawson (Eds.), 13th International Learning Analytics and Knowledge Conference (LAK2023). doi:10.1145/3576050.3576141.

Frögéli, E., Jenner, B., & Gustavsson, P. (2023). Effectiveness of formal onboarding for facilitating organizational socialization: A systematic review. PLOS One, 18(2), e0281823. https://doi.org/10.1371/journal.pone.0281823

Koper, R., Olivier, B., Representing the Learning Design of Units of Learning, Educational Technology & Society, Vol. 7, No. 3 (July 2004), pp. 97-111

Lockyer L., Heathcote, E. and Dawson, S. "Informing Pedagogical Action: Aligning Learning Analytics with Learning Design", American Behavioral Scientist, 57(10), 2013. Available: https://doi.org/10.1177/0002764213479367.

Reinecke, K., & Bernstein, A. (2013). Knowing What a User Likes: A Design Science Approach to Interfaces that Automatically Adapt to Culture. MIS Quarterly, 37(2), 427-453. http://www.jstor.org/stable/43825917

Rienties, B., Balaban, I., Divjak, B., Grabar, D., Svetec, B. & Vondra, P. (2023a) Applying and Translating Learning Design and Analytics Approaches Across Borders. U: Viberg, O. & Grönlund, Å. (ur.) Practicable Learning Analytics. Advances in Analytics for Learning and Teaching. Cham, Springer, str. 35-53 doi:10.1007/978-3-031-27646-0_3.

Rienties, B., Divjak, B., Eichhorn, M., Iniesto, F., Saunders-Smits, G., Svetec, B., Tillmann, A., & Zizak, M. (2023b). Online professional development across institutions and borders. International Journal of Educational Technology in Higher Education, 20(30), 1-16. doi:10.1186/s41239-023-00399-1.

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. (2022). Beyond one-size-fits-all in MOOCs: Variation in learning design and persistence of learners in different cultural and socioeconomic contexts. Computers in Human Behavior, 126, 106973. https://doi.org/https://doi.org/10.1016/j.chb.2021.106973

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478.

Appendix

Table 1. Classification of answers into categories

	Course details	**Planning**	**Analysis**	**Export**	**Constructive alignment**	**Learning outcomes**	**Collaboration**	**Advanced option**	**Concept**	**Help**	**Learning types**	**User interface / Technical implementation**	**General**
FUNC USEF+	2	9	**26**	1	6	8	1	0	4	1	1	4	1
FUNC USEF-	0	**4**	**4**	2	1	3	0	**4**	**4**	0	1	0	1
AVAR-	0	2	2	0	10	**16**	0	0	**12**	3	2	4	3
MISS<6	0	0	4	1	1	0	1	2	**16**	5	0	**14**	0
MISS>6	0	4	0	0	1	1	0	3	5	**8**	0	**7**	2

How Programming Helps Early-Age Learners Tackle Complex Concepts

Josipa Lipovac, Dorotea Potoč, Tomislav Jagušt
University of Zagreb
Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb
{josipa.lipovac, dorotea.potoc, tomislav.jagust}@fer.hr

Abstract. *In this paper, we introduce a novel approach to facilitate the acquisition of complex learning concepts among early-age learners, which are typically designed for older age groups, through programming workshops. Our research demonstrates how 3rd-grade elementary school students can effectively comprehend and apply material typically taught in the 6th-grade curriculum, specifically related to the coordinate system and its navigation. We conducted a series of three iteration tests, with the first two being identical tests administered three months apart. The third test was intentionally designed to be more challenging and intricate to evaluate the student's ability to handle more difficult tasks after an additional three months. Our findings indicate that younger students can successfully grasp the fundamental concepts of the coordinate system and its navigation by utilizing Scratch*[1] *programming and engaging in enjoyable exercises, such as homework assignments.*

Keywords. complex learning concepts, programming, scratch, computational thinking

1 Introduction

In today's rapidly evolving digital world, programming has become one of the most important skills of the 21st century. As a result, more and more schools are introducing programming into their curriculum, and research shows that even young children can develop computational thinking skills through early exposure to programming languages and screens (Papadakis, 2022). Programming can be introduced as a standalone subject, or its foundations can be integrated into fun preschool games, music, dance, and other activities, which can help children develop the thinking skills necessary for success in the digital age (Macrides et al., 2022).

Learning programming can help children stay up-to-date with the latest technology trends, ensuring that they are not left behind in the ever-evolving tech landscape. In addition to learning how to program, children can develop other educational outcomes such as mathematical problem-solving, critical thinking, social skills, self-management, and academic skills (Popat and Starkey, 2019). By learning to code, children can quickly advance and understand and solve problems they may have never encountered before. Although the initial results may be poor, after the learning process, the material becomes clear, and they are able to solve tasks with increased accuracy (Pérez-Marín et al., 2020).

There are several reasons why children are eager to learn coding. First and foremost, it provides them with a sense of empowerment and a feeling of control over the technology that they use on a daily basis. Sometimes children become demotivated by failure if it is more difficult for them to master a programming task or to correct a programming error. Therefore, it is necessary to encourage children in fun ways to adopt a way of thinking, and not to be afraid of trial and error, because this is the main way of learning. Additionally, coding is a creative and dynamic process that encourages children to think critically, problem-solve, and experiment with different solutions. By mastering programming, children can achieve a wide range of outcomes, such as enhancing their mathematical and computational thinking skills, developing their social and emotional intelligence, and gaining a better understanding of how technology works and how it can be used to solve real-world problems (Ryan, 2021). With the ever-increasing importance of technology in modern society, the ability to code has become a highly sought-after skill that can open up a world of possibilities for children and set them on a path toward future success.

This paper shows how programming can help early-age learners understand complex learning concepts that are typically intended for older age groups, through programming workshops in Scratch. Specifically, our study focuses on 3rd and 4th-grade elementary school students and how they can effectively comprehend and apply material related to the coordinate system and navigation that is usually taught to 6th-grade students.

[1] https://scratch.mit.edu/

All of the students that has been on the workshops are 9 or 10 years old, and are fluent in reading.

The paper discusses the benefits of teaching computational thinking to young children and how it can help develop mathematical skills. We first review related work in this area in the section Related Work, then describe the workshops we conducted, and the terminology the children have learned. In the section Results, we present the exam results from three iteration tests: before any topics were introduced, after three months of learning, and at the end of the workshops more challenging tests to evaluate their understanding. The first two tests were identical, administered three months apart, and the third was more challenging and intricate. We discuss and interpret the results before concluding with a summary of our findings.

2 Related work

Computational thinking is considered a desirable competency for adapting to the future, and it is essential to introduce it to children at an early age so that they can adapt and grow alongside the technology. Schools use a variety of methods to promote computational thinking, such as introducing activities such as robotics, program design, and game-based learning. Children learn to cooperate and think differently, connecting concepts more easily. The most widely accepted concepts are project-based learning, problem-based learning, cooperative learning, and game-based learning. Less common but still valuable approaches include storytelling, design-based learning, and aesthetic experience (Hsu et al., 2018).

Research studies have shown that young children can benefit from learning programming and computational thinking skills in various ways. For example, using tablets or robots in storytelling activities can promote the development of skills needed in the 21st century, such as creativity, problem-solving, and digital competence, as well as enhance computational thinking skills (Yang et al., 2023). ScratchJr is also a tool that can promote computational thinking skills among young children (Stamatios, 2022).

Moreover, some studies have demonstrated that young children can effectively solve complex problems by breaking them down into smaller tasks. Using the Bee-Bot robot, children can develop computational thinking skills (Angeli and Valanides, 2020). It has also been found that children between the ages of 2 and 4 can learn thinking skills with Bee-Bot, perhaps with the help of parents or educators (Critten et al., 2022).

Introducing programming in early childhood education using robotics kits can foster better communication, collaboration, and creativity, and promote confidence in children to try new things without fear (Bers et al., 2019). Research also suggests that children as young as 8 years old can learn programming skills, and there are ongoing studies on whether programming should be introduced at an earlier age and when it should be taught (Duncan et al., 2014).

Scratch, as a programming tool, can help young children develop problem-solving skills through game design activities. Children can identify problems more easily, think more creatively, and engage in problem-solving attempts (Erol and Çirak, 2022).

Furthermore, the introduction of programming and robotics in schools can promote STEM education, and introducing programming into elementary education can foster computational thinking skills (Alam, 2022). Humanoid robots can also help children learn computational thinking naturally if included in the curriculum (Chen et al., 2017).

Additionally, studies have shown that using Scratch to teach programming can enhance computational thinking and improve mathematical understanding. Students' perception of computer science as a discipline changes, and they can switch to more challenging programming languages because they have already developed thinking and mathematical skills (Grover et al., 2015).

Through workshops, simple and fun activities, games, and the use of various tools such as LEGO robotic sets, Kodu environments, and mechanical calculators, children as young as 8 years old can learn abstract concepts typically taught to older children or even 18-year-olds. These activities provide a hands-on and interactive approach to learning, enabling children to explore and experiment with concepts in a playful and engaging way. By incorporating tools and technologies, children are exposed to real-world applications of abstract concepts and can develop a deeper understanding of their relevance and practical use. This approach to education has the potential to foster a lifelong love of learning and a curiosity for the world around them (Sović et al., 2014).

3 Workshop description

3.1 Scratch environment

Scratch is an innovative programming language and environment developed by the Lifelong Kindergarten Group at the MIT Media Lab. It is designed as a block-based system that empowers users to create interactive, media-rich projects (Maloney et al., 2010). The visual programming interface of Scratch allows beginners to concentrate on programming logic without being overwhelmed by syntax errors. By utilizing a drag-and-drop approach with vibrant and colorful coding blocks, Scratch becomes accessible and user-friendly, particularly for young children who can easily comprehend and manipulate it.

At its core, Scratch aims to provide a platform that encourages children to actively create interactive stories, games, and animations. It follows a structured learning approach to introduce programming concepts

gradually. Starting with simple commands, users can gradually progress to more advanced ideas, building a solid foundation in programming.

Moreover, Scratch extends beyond programming and serves as a tool for the intuitive exploration of various mathematical concepts. It fosters a deeper understanding of mathematical principles, allowing children to engage with mathematics in a creative and interactive manner.

Furthermore, Scratch boasts an online community comprising both students and educators. This community serves as a supportive environment where children can share their projects, receive feedback, and collaborate with others.

3.2 Workshops organization

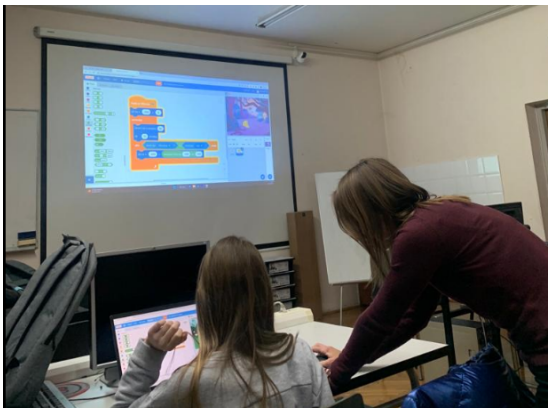


Figure 1: Image from workshop

The workshop program aims to familiarize young children in the early years of elementary school with the fundamental principles of programming through the Scratch environment. Students from the 3rd and 4th grades participated in a series of workshops spanning 22 weeks.

Table 1: Weekly workshops plans

WEEK	DESCRIPTION
1.	Getting to know the Scratch environment
2.	Coordinate system
3.	Directions
4.	Directions + Motion
5.,6.	Motion, Looks, repeat, forever and if-then
7.,8.	Variables, Operators and if-then-else
9.	Quiz
10.	Advanced motions and operators
11.,12.,**13.**	Project: merging learned knowledge
14.,15.	Clones
16.,17.,18.	Advanced programming concepts practice
19.,20.	Messages
21.,**22.**	Advanced programming concepts practice

Table 1 provides a breakdown of the concepts covered each week. Each workshop session lasted 45 minutes. Following certain workshops, students were assigned additional homework, which would then be thoroughly discussed in the subsequent workshop. On the link[2], it is possible to find projects that were worked on in certain weeks. Bolded in the table are the weeks during which survey tests were conducted, assessing the children's advancements in comprehending complex mathematical concepts.

Over the course of 22 weeks, the children were introduced to the Scratch environment, fundamental programming concepts, and various mathematical concepts that children usually encounter at school at an older age. As early as the second week, they were taught about the number line, negative numbers, and the coordinate system. They further expanded their understanding through assigned homework. In the third week, they explored the concept of a complete circle, learning the measurement of degrees and how they determine direction.

This knowledge of coordinates and directions was then applied in the creation of interactive games, providing a practical understanding of basic programming concepts. The workshop focused heavily on comprehending conditions, loops, and variables, while also delving into advanced Scratch programming topics like clones and messages.

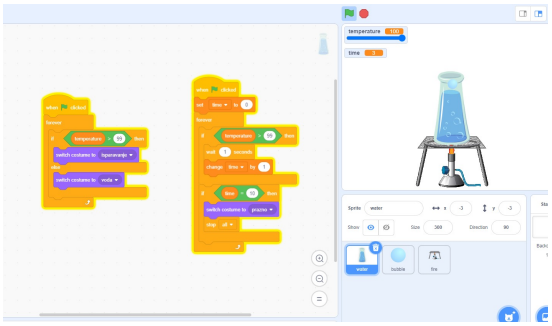


Figure 2: Example of the task on the workshop

4 The research results

4.1 The first test

At the start of the second workshop, prior to any lectures, children were administered the initial test. Its purpose was to evaluate the children's prior knowledge. The initial test consisted of three informative questions designed to check the children's familiarity with Scratch or programming. Throughout the entire semester, a total of 10 children participated in the workshops, but the comprehensive test results are available for only 6 of them.

The questions were:

1. Have you ever programmed in Scratch before these workshops?

[2]https://scratch.mit.edu/studios/33361306

2. Have you ever participated in STEMALICA's workshops before?
3. What is your knowledge of mathematics in school? What grade did you receive at the end of last school year?
4. Draw a number line in the space below this question.
5. If you have managed to draw the number line below question 4, mark the numbers 1.5 and 3.5 on it.
6. Write one negative number.
7. Determine the coordinates of points A and B marked on the coordinate system.

Based on the test results, five participants had prior programming experience with Scratch, and four had attended previous workshops by STEMALICA. Out of the group, five had an A grade in math, while one had a B. Only one participant successfully drew the correct number line, another attempted it, and the remaining participants left it blank. None of them managed to solve the fifth task, which involved marking decimal numbers. The same participant who drew the correct number line also correctly wrote a negative number. One participant wrote a positive decimal number, another wrote a positive integer, and the rest left it blank. None of the participants even attempted to solve the final task.

4.2 The second test

After 11 workshops, we conducted a follow-up test similar to the previous one, but without informative questions. This time, five participants successfully drew a correct number line, but only three of them accurately marked decimal numbers on it. All participants correctly wrote one negative number. Four of them successfully completed the final task. One participant correctly determined one point, but struggled with the other, while another participant didn't manage to complete the task accurately.

4.3 The third test

The examination took place on May 4th, 2023. The test consisted of seven tasks covering material on the coordinate system, number line, and negative numbers. An example task from the exam is shown in the **Figure 3**, where children had to determine the coordinates of treasure on the coordinate system.

The test questions were:

1. Please round off the negative numbers from the given list of numbers: 1.5, -9, 7, 2.25, -1, 0.5, -1.5, 2, 6.2, -5.3.
2. Please mark the following points on the coordinate system: A(1,1), B(6,2), C(3,6), D(4.5,2), E(1,3.5).
3. We are in the coordinate system, draw it. The point (-2, 4) is given, move 4 units to the right, and then 2 units down. What are the coordinates of your new point?
4. Please draw a coordinate system. The point (1,2) is given, move 3 units in the direction of -90°. What are your new coordinates?
5. Points A(2,2), B(2,5), and C(5.5,2) are marked on the coordinate system. To complete the rectangle ABCD, we need the coordinates of point D. What are its coordinates?
6. Determine the coordinates of points A and B marked on the coordinate system.
7. What are the coordinates of the treasure?

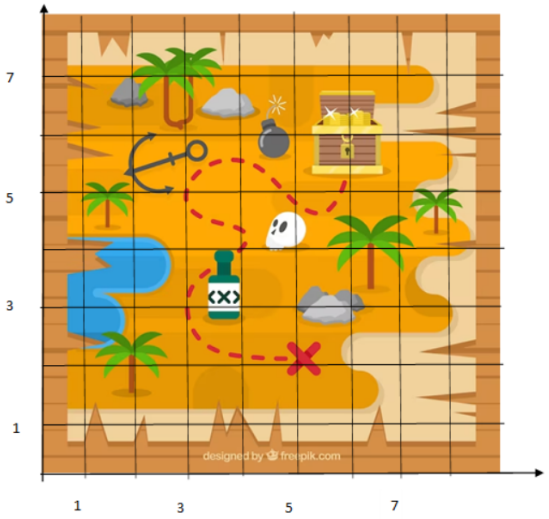


Figure 3: Example of the task on the test

All participants successfully completed the 1st, 6th, and 7th questions. The majority of participants faced difficulties with the 2nd task, but five of them managed to correctly mark at least one point, with two participants making only one mistake involving decimal numbers. Only one participant successfully completed the 3rd question, while three participants mistakenly interchanged the x and y coordinates, leading to incorrect end point. Similar issues were observed in the 4th task, with three participants accurately determining the direction despite the confusion in coordinate swapping. In the 5th task, four participants correctly marked the point, but only one of them successfully determined the point's coordinates, while others continued to struggle with decimal numbers. One participant misunderstood the task, and another participant correctly determined the decimal coordinate but not the integer one.

Table 2: Percentage of test results and participant's attendance

	1st results [%]	2nd attendance [%]	2nd results [%]	3rd attendance [%]	3rd results [%]
P1	0	85,62	100	86,36	36,67
P2	0	92,31	50	95,45	66,67
P3	50	100	75	100	73,33
P4	0	100	100	90,91	73,33
P5	0	100	62,50	68,18	46,67
P6	0	92,31	62,50	90,91	33,33

4.4 Attendance and test results

Table 2 presents individual participants' results for each test and their attendance records. Our goal was to examine the relationship between these two features.

Table 3: Average scores by tasks

	Test 1 [%]	Test 2 [%]	Test 3 [%]
1.	-	-	100
2.	-	-	43,33
3.	-	-	25
4.	16,67	83,33	25
5.	0	50	54,17
6.	16,67	100	100
7.	0	83,33	100

Table 3 displays the average test scores categorized by tasks. It is important to note that the 7th task in tests 1 and 2 is identical to the 6th task and very similar to the 7th task in test 3.

5 Discussion

Looking at results from Table 2 and evaluating the understanding of materials designed for older children it can be concluded that the participants exhibited varying levels of comprehension and skills. The analysis of their results highlights the importance of an individualized approach and adaptation of instructional materials to promote optimal learning.

Overall, certain participants demonstrated significant progress and exhibited a deep understanding of the concepts. These individuals successfully completed tasks, applied correct concepts, and accurately utilized mathematical notations. Their results indicate abilities in critical thinking, logical reasoning, and the application of learned knowledge.

On the other hand, some participants encountered challenges and made various errors. These individuals lacked an understanding of the materials and struggled to apply the concepts appropriately. Mistakes included coordinate substitutions, incorrect point labeling, and difficulties with decimal numbers. Additionally, a lack of concentration and impulsiveness were observed, resulting in imprecise responses.

These findings highlight the need for adapting instructional materials to provide participants with practical challenges tailored to their understanding. An individualized approach and careful guidance could enhance material comprehension among participants facing difficulties. It is also important to encourage participation and active engagement to improve concentration and reduce impulsiveness. This emphasizes that the understanding and acquisition of mathematical concepts among children can vary, and the teaching approach is a significant factor in achieving success. Adapting instructional materials, employing an individualized approach, and promoting active participation can be crucial and effective methods for enhancing mathematics understanding among children.

Furthermore, the student who attended all the lectures performed very well on all the tests, indicating a strong understanding. The errors made, such as swapping x and y coordinates or miscounting cells, may suggest the earlier mentioned tendency toward impulsiveness.

Also, participants P1 and P6 who performed poorly on the third test demonstrated a partial understanding of the given requirements. They consistently interchanged the order of x and y coordinates and mislabeled the initial coordinate in the third and fourth tasks. However, they displayed adequate displacement and a close approximation to the correct solution. The confusion was caused by the picture of the coordinate system in the second and fifth tasks, which led to incorrect labeling and rounding errors. Despite some areas of confusion, the students demonstrated an understanding of rotation and the coordinate system. We would attribute these errors to a lack of concentration and impulsiveness.

We would like to single out participant P5 who had a high number of absences in the workshops, which affected his test results. His performance was slightly poorer, indicating a lack of understanding of the basic concepts, which could also be noticed on the last test where his mistakes cannot be called clumsiness, but rather misunderstanding, because he did not understand some tasks that needed to be done, he also did not understand how to independently draw a coordinate system, so he left most of the test blank.

In the beginning, children had little to no knowledge about the subject matter, and on the first test, they encountered questions such as: "What is a line?" or "What does a negative number mean?" However, in the final results, we can see that the children made signifi-

cant progress and learned the basic concepts of the coordinate system.

Table 3 illustrates the overall group performance on the test tasks. After 11 weeks, it is evident that participants showed significantly greater success in solving the repeated first test. It is important to note that test 3 posed a significantly higher level of difficulty compared to tests 1 and 2. Nonetheless, progress is clearly evident. The most notable advancement is observed in the final task of the 1st and 2nd tests, which corresponds to tasks 6 and 7 in the 3rd test. In the last test, participants achieved a perfect success rate in solving these two tasks. On the other hand, the lowest success rate was observed in tasks 3 and 4 of the third test. The primary challenge in these tasks appeared to be related to the reordering of coordinate axes. Additionally, the infrequency of drawing a coordinate system independently during the workshops may have contributed to this issue.

Children with higher attendance demonstrate better understanding, although some may have achieved lower overall scores. Their responses indicate comprehension of the required tasks and approximate accuracy in their solutions.

6 Conclusion

This paper presents how Scratch programming workshops can help younger children grasp complex mathematical concepts and basic programming fundamentals typically intended for older children. By introducing children to fun block-based programming in Scratch, such as using games or everyday life scenarios like locating objects in a room using a coordinate system, children are able to overcome abstract and unfamiliar terms like number lines, coordinate systems, and angles.

Furthermore, the study reveals the positive impact of workshop attendance on understanding. Although this may not be fully evident in the final results, participants who attended most or all of the workshops demonstrated a better understanding of the subject matter and the task requirements. On the other hand, the participant with the lowest workshop attendance showed difficulties in understanding and did not attempt to solve certain tasks.

The test results indicate that the children have learned the basic concepts, although more challenging tasks were not fully mastered. However, progress can be seen when they actually understand the task at hand and what is expected of them, but make minor mistakes, such as rotating incorrectly in the coordinate system or miscounting steps for movement. This highlights the importance of consistent workshop attendance for improving comprehension and successful task completion.

As for future work, it would be interesting to explore how to further improve the Scratch programming curriculum for young children, perhaps by designing more challenging exercises that build upon the concepts they have already learned. Additionally, it could be beneficial to investigate how Scratch programming can be used to supplement other areas of early childhood education, such as creative expression. Finally, it would be worth exploring the potential impact of introducing Scratch programming workshops to children from different socio-economic backgrounds or cultural contexts.

Acknowledgments

We would like to thank the association STEMALICA[3], for the held workshops and help with the research.

References

Alam, A. (2022). Educational robotics and computer programming in early childhood education: A conceptual framework for assessing elementary school students' computational thinking for designing powerful educational scenarios. In *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, pages 1–7.

Angeli, C. and Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. *Computers in Human Behavior*, 105.

Bers, M. U., González-González, C., and Armas–Torres, M. B. (2019). Coding as a playground: Promoting positive learning experiences in childhood classrooms. *Computers Education*, 138:130–145.

Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., and Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers Education*, 109:162–175.

Critten, V., Hagon, H., and Messer, D. (2022). Can pre-school children learn programming and coding through guided play activities? a case study in computational thinking. *Early Childhood Education Journal*, 50(3):969–981.

Duncan, C., Bell, T., and Tanimoto, S. (2014). Should your 8-year-old learn coding? In *Proceedings of the 9th Workshop in Primary and Secondary Computing Education*, WiPSCE '14, page 60–69, New York, NY, USA. Association for Computing Machinery.

[3]https://stemalica.com/

Erol, O. and Çirak, N. S. (2022). The effect of a programming tool scratch on the problem-solving skills of middle school students. *Education and Information Technologies*, 27:4065–4086.

Grover, S., Pea, R., and Cooper, S. (2015). Designing for deeper learning in a blended computer science course for middle school students. *Computer Science Education*, 25(2):199–237.

Hsu, T.-C., Chang, S.-C., and Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers Education*, 126:296–310.

Macrides, E., Miliou, O., and Angeli, C. (2022). Programming in early childhood education: A systematic review. *International Journal of Child-Computer Interaction*, 32:100–396.

Maloney, J., Resnick, M., Rusk, N., Silverman, B., and Eastmond, E. (2010). The scratch programming language and environment. *ACM Transactions on Computing Education (TOCE)*, 10(4):1–15.

Papadakis, S. (2022). Apps to promote computational thinking and coding skills to young age children: A pedagogical challenge for the 21st century learners. *Educational Process: International Journal*, 11(1):7–13.

Pérez-Marín, D., Hijón-Neira, R., Bacelo, A., and Pizarro, C. (2020). Can computational thinking be improved by using a methodology based on metaphors and scratch to teach computer programming to children? *Computers in Human Behavior*, 105.

Popat, S. and Starkey, L. (2019). Learning to code or coding to learn? a systematic review. *Computers Education*, 128:365–376.

Ryan (2021). Coding for kids: Reasons kids should get started, and how they can find success. `https://www.idtech.com/blog/5-reasons-your-child-should-learn-to-code`. Accessed: May 17, 2023.

Sović, A., Jagušt, T., and Seršić, D. (2014). How to teach basic university-level programming concepts to first graders? pages 1–6.

Stamatios, P. (2022). Can preschoolers learn computational thinking and coding skills with scratchjr? a systematic literature review. *International Journal of Educational Reform*.

Yang, W., Ng, D. T. K., and Su, J. (2023). The impact of story-inspired programming on preschool children's computational thinking: A multi-group experiment. *Thinking Skills and Creativity*, 47:101–218.

Learning Analytics Frameworks: A Review of Challenges and Practices in Higher Education

Ivan Kekez, Diana Šimić
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, Varaždin
ikekez@student.foi.hr, diana.simic@foi.unizg.hr

Abstract. *Despite substantial attention, adoption of learning analytics by higher education institutions remained limited. The goal of this reseach was to analyze common challenges and practices for development and adoption of learning analytics in higher education present in learning analytics frameworks. In this context, framework is defined as a structured conceptualization that can help describe, understand, and/or implement learning analytics. The study combines systematic literature review with inductive content analysis. Categories of challenges were related to Stakeholders, Data, Infrastructure, and Ethics. The most common practice categories were Assessment and feedback; LA tools & deployment; Curriculum development & learning design.*

Keywords. learning analytics, framework, higher education, challenges and practices

1 Introduction

Learning analytics is an interdisciplinary research field that draws upon various disciplines and utilizes exponentially growing data pools generated by a variety of platforms and devices (Ferguson, 2012). Over the past decade, there has been a growing interest in learning analytics among researchers and practitioners in higher education, supported by the trend of mining and analyzing educational data provided by stakeholders to enhance teaching and learning processes in interactive environments (Chatti et al., 2012). These environments encompass different learning platforms, such as learning management systems (LMS), intelligent tutoring systems (ITS), and personal learning environments (PLE), all of which generate vast quantities of educational data (Greller & Drachsler, 2012). While there is an expectation that learning analytics will have exclusively positive effects on higher education, many institutional policy frameworks fail to adequately address the potential ethical implications and practices associated with tracking, collecting, and analyzing personal data (Slade & Prinsloo, 2013). Additionally, numerous studies tend to prioritize the implementation of various predictive analytics models, for example, to identify potential student performance (Azcona et al., 2018; Mussida & Lanzi, 2022; L. N. Singelmann & Ewert, 2022), giving limited attention to the educational and ethical principles of the research field. In other words, while learning analytics undoubtedly offers technical opportunities for higher education, there are also various challenges and issues regarding different perspectives on learning analytics adoption in higher education (Clark et al., 2020). This diversity in practices represents a significant challenge in the field, making it complex to create adequate models and frameworks in an attempt to aid learning analytics adoption in higher education (Dawson et al., 2018).

In recent review, Khalil et al. (2022) provided an overview of learning analytics frameworks focusing on development or application focus, ethics/privacy, representation, evidence of application, data types and sources, focus, and context. In contrast, the goal of this study was to investigate and highlight the established practices and challenges associated with the development and adoption of learning analytics frameworks in higher education. To achieve this goal, this study aimed to answer the following research questions:

RQ1. What are the main challenges in developing learning analytics frameworks in higher education institutions?

RQ2: What are the existing practices in implementing a framework for learning analytics adoption in higher education institutions?

In order to achieve the described goal, a systematic review was conducted using the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) framework (Page et al., 2021). Over the last few years, PRISMA has been widely recognized and used within the field of learning analytics in general, thus it is used in this review to answer research questions. The full procedure and its steps are described later in section 3 below. The contribution of this review lies not only in its analysis of non-

experimental and experimental studies related to the topic but also in its focus on addressing and analyzing specific issues within the context of learning analytics adoption. In other words, this approach enables a comprehensive examination of the challenges and practices associated with the adoption of learning analytics in diverse higher educational contexts.

2 Theoretical Basis

During the initial stages of learning analytics development phases, the pioneers of learning analytics discussed the importance of designing guidelines that would support the adoption of learning analytics in higher education. Specifically, they emphasized that the implementation of learning analytics requires a carefully crafted design to ensure positive outcomes, such as enhancing the teaching and learning processes in higher education (Chatti et al., 2012; Greller & Drachsler, 2012). For that matter, Greller & Drachsler (2012) proposed a generic framework consisting of six critical dimensions: (1) Stakeholders; (2) Objectives; (3) Data; (4) Instruments; (5) External limitations – privacy and ethics related issues, institutional norms; (6) Internal limitations – required competencies. Similarly, Chatti et al. (2012) reviewed existing publications on the subject and mapped them into a four-dimensional reference model: (1) Data and Environment (What?); (2) Stakeholders (Who?); (3) Objectives (Why?); (4) Methods (How?). A resource-based capability model presented by Knobbout & Van der Stappen (2020) placed a greater emphasis on data and business analytics, aiming to provide specific operational steps for developing the capabilities required to adopt learning analytics. The model encompassed five capability categories: (1) Data, (2) Management, (3) People, (4) Technology, and (5) Privacy and Ethics. Slade & Prinsloo (2013) put greater emphasis on ethics and privacy guidelines that would serve as an ethical framework that higher education institutions can utilize to develop context-appropriate solutions aimed toward enhancing the quality and effectiveness of teaching and learning. It is not surprising that there is overlap among discussed frameworks, as they share similar objectives and views in achieving learning analytics goals. These frameworks, as well as process areas of the Higher Education Reference Model (*Higher Education Reference Model*, 2022) were the theoretical basis for classification of challenges and practices.

3 Methodology

3.1 Selection of papers for review

This paper adopted the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) framework for its structure. While PRISMA was primarily developed for review papers in healthcare, its applicability extends to other fields such as higher education, offering new perspectives to the research field (Khalil et al., 2023; Page et al., 2021). The search was performed in May 2023. The research process is illustrated in **Figure 1.** and described in detail below.

The databases included were Web of Science (WoS Core Collection), Scopus, and IEEE Explore Digital Library. The search terms were "learning analytics," "framework," and "higher education" structured as follows: 1) Web of Science Core Collection: TS=(learning analytics) AND TS=(framework OR maturity model OR MM) AND TS=(high* education*); 2) Scopus: TITLE-ABS-KEY ("learning analytics") AND (TITLE-ABS-KEY ("high* education*") AND (TITLE-ABS-KEY ("framework")) OR (TITLE-ABS-KEY ("maturity model")) OR (TITLE-ABS-KEY ("MM")))); 3) IEEE Explore: "Abstract": learning analytics AND ("Abstract": framework OR "Abstract": maturity model) AND "Abstract": high* education*. A total of 646 papers were identified from the three databases and filtered in each database according to the following inclusion/exclusion criteria: (1) Paper types – IN: Journal articles and conference papers; OUT: Books, book chapters, editorial materials, dissertations,

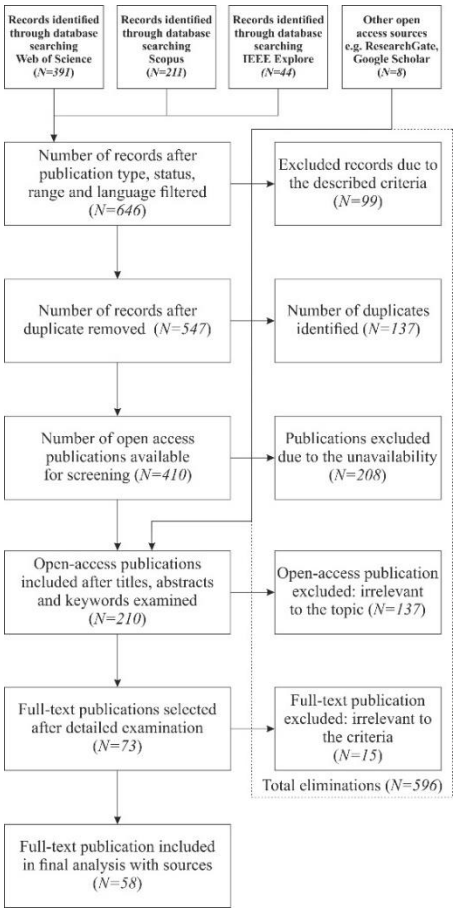


Figure 1. PRISMA flow diagram (Page et al. 2021)

workshop papers; (2) Paper status – IN: Peer-reviewed journal articles and conference papers, published papers; OUT: Non-peer-reviewed journal articles and conference papers, early access articles; (3) Publication date range – IN: 2011/01/01–2023/05/15 (yyyy/mm/dd); OUT: Outside of the proposed time interval; (4) Paper language: IN: English; OUT: Other languages. After applying the inclusion/exclusion criteria, 99 records were excluded, resulting in a total of 547 records. Among the resulting papers, 137 were identified as duplicates, and 208 were unavailable due to closed access. The final number of open-access papers available for further screening was 202.

Figure 2 presents the break-down of these papers by data source.

Title, abstract and authors' keywords of the remaining 202 papers were screened to include only papers that reported on: (1) the development of learning analytics frameworks in higher education; or (2) practices of learning analytics adoption in higher education. The first group of papers referred to an evaluation of existing frameworks and challenges related to the adoption of learning analytics in higher education. The second group addressed issues and approaches concerning learning analytics practices in higher education. Both experimental and non-experimental approaches were included. The experimental approach included case studies, action research in classrooms, model testing, etc. The non-experimental approach involved an analysis of known conceptual learning analytics frameworks, proposed guidelines and principles, evaluation of learning analytics dimensions, critical factors, and so on. Systematic reviews were excluded from further observation. By following the described process, a total of 137 papers were discarded as they were not relevant to the research questions. Additionally, 8 papers that met the criteria, but were not available in open access directly from the searched databases, were found in other repositories (Google Scholar, ResearchGate, publisher) and included in the analysis, resulting in a final number of 73 papers, after screening the titles, abstracts, and keywords.

After conducting a final full-text review, 15 of the remaining papers were excluded due to their lack of relevance or importance to the topic. As a result, the final analysis included 58 papers, 25 were in the first group, and 33 in the second group, based on the dominant characteristics of the papers. The final distribution of the data sources for the analysed papers is depicted in **Figure 3**.

3.2 Content analysis

An inductive content analysis was performed in each group of papers separately (Vears & Gillam, 2022). The unit of analysis was the whole paper. In the first group of papers the focus was on coding different types of challenges.

The second group of papers were classified according to the types of challenges identified in the first group of papers. Next, the papers were classified according to their focus on pedagogical or technological practices. Pedagogical practices referred to the processes by which stakeholders approach teaching and learning. They also dealt with ethical considerations and student privacy by developing policies for the ethical use of student data in pedagogical contexts. Likewise, pedagogical practices involved motivating and educating stakeholders to utilize learning analytics tools effectively. Furthermore, they encouraged discussions on learning analytics practices to enhance institutional policies and decision-making and promote the use of data-driven insights to improve the teaching and learning experience. On the other hand, technological practices included implementing tools and defining measures to effectively collect, analyze, and protect data in compliance with privacy regulations. They included selecting and integrating appropriate solutions for learning analytics, establishing data collection and analysis processes, ensuring data security, and guarding student privacy. Technological practices also involved using visualization tools to present analytics insights in a user-friendly manner for stakeholders and leveraging technology for iterative improvement of the learning analytics performance. These papers were also classified regarding the research design into 5 groups: 1) Pilot testing, 2) Focus groups, interviews, open discussions, 3) Case studies, 4) Reports, evidence-based recommendations, empirical analysis, 5) EDM, data collection, data analysis, predictive modeling and ML, ETL, etc. Finally, inductive coding was used to extract keywords from research questions, hypotheses,

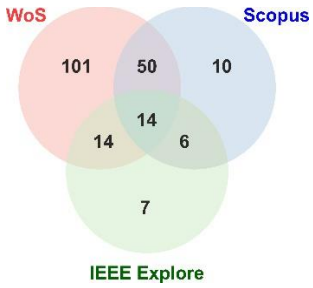


Figure 2 Number of open access papers satisfying inclusion/exclusion criteria, after deduplication, by source.

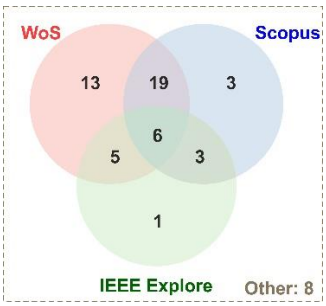


Figure 3 Number of papers included in the final analysis by source

Table 1. Patterns of challenges addressed by the first group of papers, indicating the challenge addressed by a star, and the number of papers for each combination of challenges.

Number of papers (#)	Challenges			
	(1)	(2)	(3)	(4)
5	*	*	*	*
5	*	*		*
1	*	*	*	
12	*	*		
2	*			*
25	25	23	6	12

and conclusions, indicating the type of practice or its area of application. These keywords were finally grouped into broader categories.

Excel table with references and results of content analysis is available as an open research data set (Kekez & Šimić, 2023).

4 Results

RQ1. What are the main challenges in developing learning analytics frameworks in higher education institutions?

Content analysis of the first group of papers identified four broad categories of challenges.

(1) *Stakeholder engagement in adopting learning analytics* - developing a learning analytics framework requires collaboration from various stakeholders in the process, including teachers, students, faculty administrators, decision-makers, IT staff, etc (Greller & Drachsler, 2012; Knobbout & Van der Stappen, 2020). Engaging teachers and students in learning analytics training and addressing their concerns, such as data analysis, privacy, and security, can be overwhelming and concerning, creating resistance among them and hindering learning analytics adoption (Gray et al., 2021) or can include different perspectives regarding the exploitation and use of learning analytics in higher education (Alenezi et al., 2018). In the teaching process, teachers are involved in designing learning activities, motivating students, and encouraging active participation (Joshi et al., 2020), often based on their subjective perceptions. Therefore, teachers may find it challenging to transition from diagnostic to practical data-driven interventions (Van Leeuwen, 2019) as, on average, they lack the technical knowledge to push the boundaries toward new approaches to teaching and learning (Scheffel et al., 2019; Schmitz et al., 2017). Furthermore, Sahni (2023) acknowledges that there are still challenges related to technology adoption and institutional support in effectively utilizing learning analytics tools, while highlighting the value and support that learning analytics tools have on students' learning experience resulting in higher levels of engagement, better results, and overall satisfaction.

(2) *Data availability, integration, and interoperability* - higher education institutions generate vast amounts of data, but it is scattered across different systems and platforms and in various formats. These data can provide a richness of information capable of improving the learning process and formal education in general (Clark et al., 2020). Enabling data consistency, accuracy, and availability is a significant challenge for higher education institutions when developing a comprehensive learning analytics framework. Different data formats, incompatible systems, and data privacy concerns can limit data integration efforts (Freitas et al., 2020). Establishing interoperability standards and ensuring data flow between systems and stakeholders present critical challenges in developing a framework (Chatti et al., 2012). Using educational data mining techniques can help the development of a learning analytics framework as it allows researchers to keep data consistency and allows complex methods for data analysis (L. Singelmann et al., 2021). Although educational data mining focuses more on extracting insights and value from data, it is complementary to learning analytics strategies to enhance educational processes based on valuable knowledge gained from students' data (Soltanpoor & Yavari, 2017).

(3) *Building an adequate infrastructure* - developing an effective learning analytics framework requires technological infrastructure and resources (Sanagustín, et al., 2019; Tsai et al., 2018). The LALA framework (Sanagustín, et al., 2019) defines the technological dimension as a set of manuals and steps that offer an overview of the technological needs and capacities that institutions must evaluate for the adoption of learning analytics. Institutions may face challenges by acquiring, implementing, and maintaining the necessary tools, platforms, and analytics capabilities. Furthermore, implementing new technologies and their integration into the existing systems and infrastructure is a follow-up challenge for higher educational institutions (Tsai et al., 2018) as many technologies and business models are commercial and do not directly apply to the educational processes (Alenezi et al., 2018).

(4) *Ethical and privacy concerns* - learning analytics involves the collection and analysis of sensitive student data. It is crucial to ensure compliance with privacy regulations and ethical guidelines when utilizing data for analytics. Balancing the need for data-driven insights with privacy protection is essential for building trust and maintaining ethical practices. By prioritizing privacy, institutions can establish a foundation of trust and transparency in using student data for learning analytics (Chatti et al., 2012; Greller & Drachsler, 2012; Slade & Prinsloo, 2013). Slade & Prinsloo (2013) discussed in detail ethical issues and dilemmas regarding data management, privacy, and security but also about potentially negative aspects of

learning analytics implementation in higher education institutions that could lead to a problem of viewing students primarily as data "producers" neglecting the importance of students' educational experiences. Greller & Drachsler (2012) stated that ethics extend beyond data gathering and integration, therefore ethical and privacy concerns overlap with other learning analytics categories. For example, ensuring transparency regarding data use may involve implementing protection measures, establishing clear communication channels with data 'producers' (e.g., students), and complying with relevant privacy regulations. These actions are part of different categories, as they require building a technological infrastructure, various operations on data, and stakeholder engagement.

As shown in **Table 1**, all papers address the Challenge (1) (stakeholder engagement), and only two papers do not address the Challenge (2) (data integration). Half of the papers address Challenge (4) (ethics), and only a quarter address Challenge (3) (infrastructure). The first mention of the technology infrastructure challenges appears in 2018.

RQ2: What are the existing practices in implementing a framework for learning analytics adoption in higher education institutions?

The second group of papers comprised 33 papers. Breakdown by the addressed challenges, practice type and category is presented in **Table 2**. Practice themes were grouped in eight broad practice categories. The most frequent category was *Assessment and feedback* (11 papers), followed by *LA tools & deployment* (8), *Curriculum development & learning design* (7), *Learning experience* (5), *Stakeholders* (4), *Ethics, privacy, security* (4), *Learning innovation* (4), and *Learning process* (2).

Similar to the papers on challenges, all papers on practices address the challenges of *Stakeholder engagement* and *Data availability, integration, and interoperability*. Papers usually address both *Building an adequate infrastructure* and *Ethical and privacy concerns* challenges.

Practice categories appearing most frequently together are *Assessment & feedback* and *Learning experience*. *Learning experience* category also appears frequently with the *Stakeholders* category. *LA tools & deployment* practices are often connected to *Learning process*. *Learning innovation* appears by itself, or in combination with *Assessment & feedback* or *Curriculum development & learning design*. Finally, *Ethics, privacy, security* category was not combined with other practice categories in the analysed papers.

Analysis of research designs showed that *Pilot testing* and *EDM, data collection, data analysis, predictive modeling and ML, ETL, etc*. were used in papers on Technological practices, while papers on Pedagogical practices employed various research designs, including both qualitative (*Focus groups, interviews, open discussions* and *Case studies*), and quantitative research (*Reports, evidence-based recommendations, empirical analysis*).

Even though these papers analyse a wide selection of LA practices and propose or discuss implementation of an LA framework, most of them do not provide evidence of fully successful framework implementation.

Table 2. Breakdown of the number of papers on LA practices by the challenge addressed, type and categories of practices.

Challenges					Practice type		
Stakeholders	Data	Infrastructure	Ethics	Practice categories	Pedagogical	Technological	Total
*	*	*	*	Ethics, privacy, security	2		2
*	*	*	*	Learning innovation	1		1
*	*	*	*	LA tools & deployment	1		1
*	*	*	*	LA tools & deployment Stakeholders Learning process	1		1
*	*		*	Ethics, privacy, security	2		2
*	*			LA tools & deployment	5		5
*	*			Assessment & feedback Learning experience	3		3
*	*			Learning experience Stakeholders	2		2
*	*			LA tools & deployment Learning process	1		1
*	*			Assessment & feedback Curr. dev. & learning design	1		1
*	*			Curr. dev. & learning design Learning innovation	1		1
*	*			Assessment & feedback	3	3	6
*	*			Curr. dev. & learning design	3	1	4
*	*			Assessment & feedback Learning innovation		1	1
*	*			Learning innovation		1	1
*	*	*		Curr. dev. & learning design		1	1
				Total	**26**	**7**	**33**

5 Discussion

Lee, Cheung & Kwok (2020) maintain that identification of practices provides helpful indicators for educational stakeholders to adopt learning analytics and can also motivate stakeholders to seek practical solutions or for researchers to find valuable research directions. Investigation of practical approaches offers insights and guidance to overcome challenges and maximize the benefits of learning analytics adoption in higher education institutions (Leitner et al., 2019). We hope that results of this study will be a valuable

addition to previous research on challenges and practices in LA adoption.

This synthesis revealed that most of the analysed studies discuss approaches for learning analytics adoption, but do not provide specific procedures and steps for achieving successful implementation. Although higher education institutions are mostly aware of the benefits of learning analytics adoption, many struggle to implement theoretical concepts into practice. Higher education institutions recognize the issues, but lack the specific practices that would serve as a solution.

There are several limitations to this research. First, we limited the review to open access publications and publications available through public repositories. Still, even the limited scope of these papers offered a rich overview of different challenges addressed through LA frameworks and approaches to their implementation. Second, classification of the papers on practices into those on pedagogical or technological approaches may have introduces a degree of uncertainty. There were papers that used both approaches in different measure. The authors' decision on the prevalent approach may lack objectivity. Finally, inductive content analysis is always to some degree subjective, thus this research should be extended with reliability analysis, that could not fit the extent of this paper.

6 Conclusion

This study highlights the interdisciplinary nature of learning analytics and its potential to enhance the learning process through data collection and analysis. Despite its growing attention, the adoption of learning analytics in higher education institutions remains quite limited. The study analyzed papers focused on learning analytics frameworks and practices in higher education and revealed several findings. Papers on LA framework development focused on four types of challenges: 1) Stakeholder engagement in adopting LA, 2) Data availability, integration, and interoperability, 3) Building an adequate infrastructure, and 4) Ethical and privacy concerns. The first two challenges were addressed by almost all the papers, and a quarter of the papers address the fourth challenge. The third challenge appears only in the most recent papers (since 2018), and is addressed by the least number of papers.

Papers on implementing LA frameworks followed the similar distribution of challenges, all papers addressed challenges 1) and 2). The third and the fourth challenge were only addressed in papers focusing on pedagogical aspects of LA adoption. Eight broad categories of LA practices were identified. These were, in decreasing frequency of appearance, *Assessment and feedback; LA tools & deployment; Curriculum development & learning design; Learning experience; Stakeholders; Ethics, privacy, security; Learning innovation;* and *Learning process.*

Taking into account study limitations, we may conclude that it provides a first glimpse into challenges of developing and implementing LA frameworks, and approaches used to overcome these challenges. By managing these challenges, higher education institutions can harness the full potential of learning analytics to enhance teaching and learning outcomes. Further research is needed to paint a more complete picture of the ever-changing landscape of LA adoption.

Acknowledgments

This work has been fully supported by the Croatian Science Foundation under the project IP-2020-02-5071.

References

Alenezi, A., Emmanouilidis, C., & Al-Ashaab, A. (2018). Knowledge Management to Support Learning Analytics in Higher Eduction. *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, 362–367. https://doi.org/10.1109/CIST.2018.8596653

Azcona, D., Hsiao, I.-H., & Smeaton, A. F. (2018). Personalizing Computer Science Education by Leveraging Multimodal Learning Analytics. *2018 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE.2018.8658596

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318. https://doi.org/10.1504/IJTEL.2012.051815

Clark, J.-A., Liu, Y., & Isaias, P. (2020). Critical success factors for implementing learning analytics in higher education: A mixed-method inquiry. *Australasian Journal of Educational Technology*, *36*(6), 89–106. https://doi.org/10.14742/ajet.6164

Dawson, S., Poquet, O., Colvin, C., Rogers, T., Pardo, A., & Gasevic, D. (2018). Rethinking learning analytics adoption through complexity leadership theory. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge LAK'18*, 236–244. https://doi.org/10.1145/3170358.3170375

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304. https://doi.org/10.1504/IJTEL.2012.051816

Freitas, E., Fonseca, F., Garcia, V., Ferreira, R., & Gasevic, D. (2020). Towards a Maturity Model for Learning Analytics Adoption An Overview of its Levels and Areas. *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, 122–126. https://doi.org/10.1109/ICALT49669.2020.00059

Gray, G., Schalk, A., Rooney, P., & Lang, C. (2021). A Stakeholder Informed Professional Development Framework to Support Engagement with Learning Analytics. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 237–247. https://doi.org/10.1145/3448139.3448162

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, *15*(3), 42–57.

Higher Education Reference Model (2.6.1). (2022). CAUDIT.

Joshi, A., Desai, P., & Tewari, P. (2020). Learning Analytics framework for measuring students' performance and teachers' involvement through problem based learning in engineering education. *Procedia Computer Science*, *172*, 954–959. https://doi.org/10.1016/j.procs.2020.05.138

Kekez, I., & Šimić, D. (2023). *Learning Analytics Frameworks: A Review of Challenges and Practices in Higher Education / Data* [dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/OITRGE

Khalil, M., Prinsloo, P., & Slade, S. (2022). A Comparison of Learning Analytics Frameworks: A Systematic Review. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 152–163. https://doi.org/10.1145/3506860.3506878

Khalil, M., Slade, S., & Prinsloo, P. (2023). Learning analytics in support of inclusiveness and disabled students: A systematic review. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-023-09363-4

Knobbout, J., & Van der Stappen, E. (2020). A Capability Model for Learning Analytics Adoption: Identifying Organizational Capabilities from Literature on Learning Analytics, Big Data Analytics, and Business Analytics. *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI)*, *2*(1), 47. https://doi.org/10.3991/ijai.v2i1.12793

Lee, L.-K., Cheung, S. K. S., & Kwok, L.-F. (2020). Learning analytics: Current trends and innovative practices. *Journal of Computers in Education*, *7*(1), 1–6. https://doi.org/10.1007/s40692-020-00155-8

Leitner, P., Ebner, M., & Ebner, M. (2019). Learning Analytics Challenges to Overcome in Higher Education Institutions. In D. Ifenthaler, D.-K. Mah, & J. Y.-K. Yau (Eds.), *Utilizing Learning Analytics to Support Study Success* (pp. 91–104). Springer International Publishing. https://doi.org/10.1007/978-3-319-64792-0_6

Mussida, P., & Lanzi, P. L. (2022). A computational tool for engineer dropout prediction. *2022 IEEE Global Engineering Education Conference (EDUCON)*, 1571–1576. https://doi.org/10.1109/EDUCON52537.2022.9766632

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Sahni, J. (2023). Is Learning Analytics the Future of Online Education?: Assessing Student Engagement and Academic Performance in the Online Learning Environment. *International Journal of Emerging Technologies in Learning (IJET)*, *18*(02), 33–49. https://doi.org/10.3991/ijet.v18i02.32167

Sanagustín, M. P., Hilliger, I., Maldonado, J., Pérez, R., Ramírez, L., Muñoz-Merino, P. J., Tsai, Y.-S., Ortiz, M., Broos, T., Zúñiga-Prieto, M., Sheihing, E., & Whitelock-Wainright, A. (2019). *LALA Framework (Technical Report): Building Capacity to Use Learning Analytics to Improve Higher Education in Latin America* (p. 138). https://researchmgt.monash.edu/ws/portalfiles/portal/324504028/324503712_oa.pdf

Scheffel, M., Tsai, Y.-S., Gašević, D., & Drachsler, H. (2019). Policy Matters: Expert Recommendations for Learning Analytics Policy. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming Learning with Meaningful Technologies* (Vol. 11722, pp. 510–524). Springer International Publishing. https://doi.org/10.1007/978-3-030-29736-7_38

Schmitz, M., Van Limbeek, E., Greller, W., Sloep, P., & Drachsler, H. (2017). Opportunities and Challenges in Using Learning Analytics in Learning Design. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data Driven Approaches in Digital Education* (Vol. 10474, pp. 209–223). Springer International Publishing. https://doi.org/10.1007/978-3-319-66610-5_16

Singelmann, L. N., & Ewert, D. L. (2022). Leveraging the Innovation-Based Learning Framework to Predict and Understand Student Success in Innovation. *IEEE Access*, *10*, 36123–36139. https://doi.org/10.1109/ACCESS.2022.3163744

Singelmann, L., Striker, R., Vazquez, E. A., Swartz, E., Pearson, M., Ng, S. S., & Ewert, D. (2021). Creation of a Framework that Integrates Technical Innovation and Learning in Engineering. *2021 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE49875.2021.9637179

Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Soltanpoor, R., & Yavari, A. (2017). CoALA: Contextualization Framework for Smart Learning Analytics. *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, 226–231. https://doi.org/10.1109/ICDCSW.2017.58

Tsai, Y.-S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gašević, D. (2018). The SHEILA Framework: Informing Institutional Strategies and Policy Processes of Learning Analytics. *Journal of Learning Analytics*, *5*(3). https://doi.org/10.18608/jla.2018.53.2

Van Leeuwen, A. (2019). Teachers' perceptions of the usability of learning analytics reports in a flipped university course: When and how does information become actionable knowledge? *Educational Technology Research and Development*, *67*(5), 1043–1064. https://doi.org/10.1007/s11423-018-09639-y

Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. *Focus on Health Professional Education - A Multidisciplinary Journal*, *23*(1), 111–127.

ICT Entrepreneurship and Innovation

Daria Jergan, Ivana Đunđek Kokotec, Marina Klačmer Čalopa
Inflationary Expectations of Consumers in the Republic of Croatia

Inflationary expectations of consumers in the Republic of Croatia

Daria Jergan, Ivana Đunđek Kokotec, Marina Klačmer Čalopa
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42 000 Varaždin, Croatia
djergan@student.foi.hr, {idjundjek, marina.klacmer}@foi.unizg.hr

Abstract. *The main goal of this paper is to examine respondents' views on inflation and their understanding of the concept of inflation, considering its very large impact on the everyday life of citizens. From the above, consumers must understand the concept of inflation, given that the annual inflation rate was on January 13, 2022—13.1% for the Republic of Croatia. The work is based on primary research. Therefore, the main research instrument is a survey questionnaire. The questionnaire was modelled after research conducted by The University of Michigan to report on consumer attitudes about inflation, monetary policy, and money's value over several years. Finally, based on the collected answers, appropriate statistical analyses will be made to determine the understanding of inflation and its effect on consumers and identify guidelines for possible improvements in the financial literacy of the citizens of the Republic of Croatia.*

Keywords: inflation, personal consumption, financial literacy, value of money

1 Introduction

The main problem of research on the topic "Inflationary expectations of consumers in the Republic of Croatia" is the growing rate of inflation in the Republic of Croatia in the last five years, from 2018-2022, with a record inflation rate of 13.1% in December 2022, which caused concern among Croatian citizens. The media are increasingly following inflation and the concept of inflation. Also, the media emphasize that the prices of the consumer basket are rising. Understanding consumer expectations and opinions about future events is extremely important. Also, one of the problems is the poor financial literacy of Croatian citizens. Today, various institutions are trying to expand initiatives on strengthening financial literacy. One of the ideas of this work is also to see if citizens understand the concept of "inflation," which is one of the basic concepts related to financial literacy. Accordingly, a survey deals with the specific questions of the citizens of Croatia related to inflation. The entire research is narrowly focused on the attitudes and opinions of citizens, that is, consumers. The research aims to see what consumers expect in the future (short and long-term). The work also includes sections where citizens could express themselves freely about inflation-related problems.

The aim of the paper is to determine whether citizens understand the concept of inflation and to determine what knowledge and attitudes they have about the current situation, which refers to the short term, but also to determine citizens' expectations about inflation in the long term and to analyse citizens' opinions about price increases on a personal level. The same is based on analyses that will provide conclusions related to the expected inflation rates of citizens, their opinions on the Government's measures or the growth of certain components of the consumer basket. By the defined field of work and highlighted research goals, several research questions were asked: RQ1: What is the consumer's perception of inflation trends in Croatia?; RQ2: Do consumers in the Republic of Croatia think that the future is uncertain?; RQ3: Is the financial literacy of consumers in the Republic of Croatia at a satisfactory level?

The paper is designed in such a way as first to describe the concept of inflation in general, that is, inflation and the components of the Harmonized Index of Consumer Prices over several years. After dealing with the theoretical aspect of inflation, several previous types of research, both abroad and at home, were singled out. These served as a reference for the initial conduct of this research and the survey preparation. In the next chapter, the research methodology is briefly explained, and only the most important results of the answers obtained from a total of 286 respondents collected by the survey are presented. Finally, the Cronbach alpha coefficient measure is used for the research, which is used as a measure of internal consistency where one wants to see the data connection within each section. The paper concludes with a conclusion and certain guidelines for improving the current situation.

2 Literature overviews

Inflation refers to the situation in the economy when the general price level of goods and services consumers buy in a certain period increases. The rise in the price level leads to a decrease in the purchasing power of the domestic currency, so consumers have to pay more month after month for the same basket of goods and services than before (CNB, 2021).

Babić (1996) states that inflation results from an imbalance in the national economy, i.e., inflation manifests itself in the rise of prices in the country, which various price indices can monitor. He also notes that any stronger inflation, regardless of its triggers or causes, has consequences, namely the impact on production, consumption, savings, investments, budget revenues and expenditures, the balance of payments, and social and political problems. Higher inflation affects the population's behaviour, spending more and saving less.

Author Olivier Blanchard (2011) notes that inflation is a permanent increase in the general price level, while the inflation rate is the rate at which prices rise. Macroeconomists often look at the consumer price index (CPI) to measure the average consumption or the cost of living - consumer price index, hereafter, CPI.

Product prices may change over time, depending on supply and demand for those products. For example, when it comes to the supply of corn, a good crop increases the supply of corn, which can cause its price to fall, but a bad crop will reduce the supply, so the price of corn will rise. On the other hand, on the demand side, for example, some expensive brands cause their prices to rise, while the price of less popular products will fall if their demand is lower. In addition to supply and demand, inflation is also affected by the amount of money in circulation. The greater the amount of money in circulation in an economy, the greater the risk of price growth, that is changes and increases in the general price level. Inflation means that an increase in the general level of prices means a decrease in the value of money; in other words, it is necessary to pay more to buy the same amount of products and services as before. High inflation is usually bad for individuals or entire households because it reduces their "purchasing power." Also, high inflation is bad for the economy because it can damage economic activities. Households with a low or fixed income are often affected by inflation, as it is unlikely that they have protected themselves from inflation (European Central Bank, 2023). Figure 1 shows the annual and average annual inflation rate, based on which the nominal price growth can be seen. Moreover, one of the reasons for the same in 2012 is the increase in fuel prices, as Poslovni.hr (2012) noted, because it is noted that the increase in the price of gasoline began in 2000 when the policy of banning the formation of the market price of gasoline was interrupted and suspended. A big problem was the increase in the price of fuel, so in 1999, the average price of a litter of Euro super 95 was HRK 4.03, and in March 2012, the price of Euro super 95 jumped to HRK 10.77.

In 2021, according to the data of the Croatian National Bank, inflation amounted to 5.5%, which manifested itself in the bills for fuel and food products and caused increased attention from the media and the general public. The increase in inflation in 2021 is the result of several reasons, namely: the increase in imported inflation, i.e., the rise in the prices of raw materials and industrial products on the world market, then the recovery of domestic personal consumption, which encouraged manufacturers and traders to pass on the increase in costs to end customers, i.e., consumers. The opening of the economy due to the pandemic in 2019 and 2020 stimulated the recovery of the global system, which led to an increase in the price of energy on the world market. From the middle of 2020, the prices of most other raw materials on the world market increased: cereals, oilseeds, wood and metal. In addition, oil prices have affected some other services over time, such as the increase in transportation services. In practice, the increase in the price of raw materials such as oilseeds increases the price of oil, and this indirectly affects the increase in the price of restaurants, considering that they use oil to prepare food. In 2021, many changes in habits took place, so people avoided public transport, which indirectly led to more work from home or buying more IT equipment, which indirectly led to an increase in the price of the same. The increase in the prices of food raw materials was also influenced by unfavourable weather conditions (droughts, floods, frost), which affected the yields of certain crops, such as the cultivation of cereals and wheat (CNB, 2022).

The increase in the price of crude oil on the global market resulted in an increase in the prices of oil derivatives on the Croatian market, given that Croatia is an importer of oil as well as most other raw materials such as metal for semiconductors or grains for certain products, i.e., the Croatian economy is based on imports and is dependent about prices at the world level and the increase in the prices of raw materials on the global market causes an increase in the prices of raw materials on the Croatian market, as was visible in 2021. In 2022, inflation reached a record 13.1%, and one of the indirect reasons is the war between Ukraine and Russia that took place in the first half of 2022. Namely, the Croatian economy was not particularly exposed to the direct impact of Russian aggression on Ukraine because it had limited trade and financial ties with those countries compared to other European countries with direct ties. However, the restrictions on production and production due to the war events that the war events caused other European members caused a weaker supply of certain raw materials. It is known that Russia is the world's gas exporter, so the gas prices on the European market also rose as a reflection of lower supplies. The prices of certain metals and food raw materials, which Russia and Ukraine used to

supply the world market, also rose. Russia and Ukraine played an important role in the global market as they were large producers and exporters of sunflower (sunflower oil), wheat, barley, corn, coal and oil as energy and artificial fertilizers. The mentioned raw materials were used as input for producing many food products in many countries. Furthermore, the supply of these products was reduced due to the events of the war, which increased the prices of food raw materials on the world market. Also, Russia exploited metal ore deposits and supplied many countries with iron, steel and non-ferrous metals. At that time, the supply of these raw materials was reduced due to sanctions, which made trade and payment transactions with Russia more difficult, as well as bans on the export of certain raw materials from Russia, which strongly increased metal prices on the world market after the start of the war. Croatia's food price increase was higher than the average of EU member states (CNB, 2022). It is mentioned that the war had no direct connection with the price rise in Croatia. However, Croatia is predominantly an import economy, meaning it imports most of its food raw materials, oil raw materials, and processed products from European countries. Russia and Ukraine were large exporters, large producers and exporters of sunflowers (sunflower oil), wheat, barley, corn, coal and oil as energy sources and artificial fertilizers, and this directly influenced the increase in the prices of raw materials for European countries, and after that European countries increased raw materials that they export, for example, to Croatia. After that, in response to the increase in the prices of raw materials, Croatia increased consumer prices, which is one of the reasons why the increase in prices in Croatia was higher than the average of EU member states.

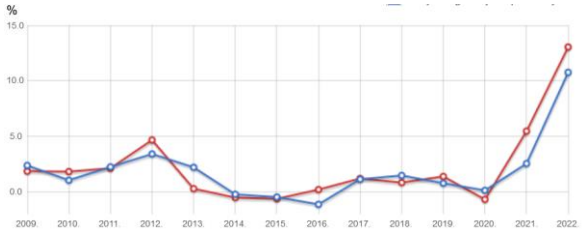


Figure 1. Annual and average annual inflation rate (Croatian Bureau of Statistics, 2023)

2.1 Projections of inflation rates until 2025

Macroeconomic inflation projections were made from September 2022 to September 2024, i.e., for three years. In 2021, Croatia and the Eurozone were affected by the energy crisis and increased inflation. The consequences of the war in Ukraine led to strong inflationary pressures for Croatia and the Eurozone. One of the scenarios for 2023, according to Figure 2, "Inflation Projections of Eurosystem Experts and European Central Bank Experts," is a negative scenario with inflation of 7.4% and 3.4% in 2024, and the reason for this is the possibility of supply disruptions in energy sources in Europe, which would consequently lead to an increase in energy prices and a decrease in production. One of the expected positive outcomes is a reduction in gas prices due to the financial support measures of the Government of the Republic of Croatia. Fiscal measures that compensated for and reduced high prices and inflation improved the inflation outlook in the projection period. HIPC-measured food price inflation is also expected to decrease in the later years of the projection period due to the underlying assumption of falling commodity prices (European Central Bank, 2023). Namely, in 2022, a package of measures to mitigate the rise in energy prices in Croatia was presented, worth HRK 4.8 billion. An example of the measures is that the VAT on pellets, briquettes, wood chips and firewood was reduced from 25% to 13%. Also, the VAT on gas and thermal energy was permanently reduced by changing the law from 25% to 13% (Government of the Republic of Croatia, 2022). Figure 2. Inflation projections of Eurosystem experts and European Central Bank experts (European Central Bank, 2023)

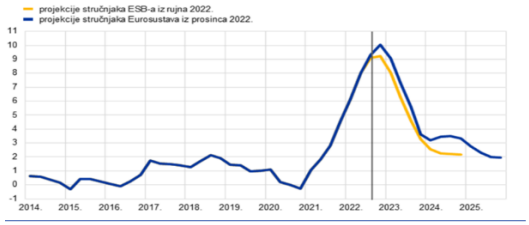


Figure 2. Eurosystem and European Central Bank experts' inflation projections (European Central Bank, 2023)

2.2 Measuring inflation expectations

In the article "Measuring Inflation Expectations," the authors Oliver Armantier, Wändi Bruine de Bruin, Simon Potter, Giorgio Topa, Wilbert van der Klaauw and Basit Zafar (2013) review and explain an initiative by the "Federal Reserve Bank of New York" which sought to improve the understanding of inflation expectations. Pointing out that inflation expectations play a key role in the functioning of modern monetary policy. Also, Armantier et. al. (2013) stated that expectations drive consumers and lead them to make a wide range of decisions, such as savings and investments, purchases of durable goods and salary negotiations. These decisions affect real economic activity and inflation. However, inflation should follow shocks; for example, if energy prices suddenly rise due to supply disruptions, short-term inflation expectations should rise simultaneously. On the other hand, long-term expectations should not change after the end of supply disruptions. Long-term inflation expectations concerning short-term inflation expectations are a hallmark of stable monetary policy. Central banks rely on various indirect approaches to measuring inflation expectations in the real world, such as expert surveys.

In addition to market measures and survey evidence, professional Central Bank forecasters rely on survey measures of general inflation expectations, with households being the main drivers of activity. Furthermore, at the same time, they play a major role in inflation expectations, and future realized inflation. An important source of survey-based information on US households and their inflation expectations is the Thomson Reuters/University of Michigan Consumer Survey. The Michigan survey randomly selects 500 households each month and asks questions that have not changed for 30 years. Armantier et. al. (2013) (according to (Gali, 2008; Sims 2009, and Woodford 2005) point out that although surveys are conducted routinely around the world, there are several problems, one of which is the possibility that consumers did not act and did not respond by their beliefs, attitudes and expectations related to inflation. In the Republic of Croatia, not much research of this type has been carried out on the title topic of the paper. However, it is worth noting the research by the authors Žmuk and Bonić (2022), who, in their work, analysed the attitudes of Croatian citizens on ways to preserve the value of property in periods of high inflation. The authors conclude that the citizens of the Republic of Croatia are familiar with the concept of inflation to a significant extent. However, they conclude that they are passive in the fight against inflation and try to preserve the value of property by saving or spending less, which are inefficient means of preserving the value of money. The inflation rate during the research period (August - September 2022) was not as pronounced as it was at the end of 2021, i.e., at the beginning of 2022, and the authors, therefore, understand that citizens did not need a higher level in the area of protecting the value of money.

3 Methodology description

In order to investigate the attitudes and expectations of the citizens of the Republic of Croatia about inflation and to propose appropriate measures to improve financial literacy, research was conducted using a survey as an instrument of the same, whereby the goal was to understand how citizens act and behave under the same economic conditions. An example of the same economic conditions is the occurrence of inflation, where one wants to see how many citizens will decide on some form of savings to suppress inflation on a personal level or what percentage of citizens have not prepared for the arrival of a wave of price increases. A survey questionnaire was distributed, created using the Google Forms tool, intended for the creation of online surveys and was conducted from April 4, 2023. until April 11, 2023. years. The survey questionnaire was created according to the example of the research mentioned earlier, namely the survey conducted by the "Federal Reserve Bank of New York" and the survey conducted by "The University of Michigan," and the final number of responses collected is 286 (n=286). This research included 286 respondents who responded to the questionnaire that was available in the period from April 4, 2023, until April 11, 2023. The questionnaire consisted of closed and open-type questions divided into three sections. Before the first section, questions are asked about the demographic characteristics of the respondents. The first section is related to the respondents' opinions on the current financial situation, and the aim was to see if the respondents understood the concept of inflation and price increases. The second section is related to the increase in prices in the respondents' households and the increase in certain components of the consumer basket. Namely, we wanted to find out what, at the respondents' opinion, has become more expensive and how they prepared for inflation. The third section is related to respondents' expectations, that is, consumers, about the occurrence of certain events. It was required to specify and specify the exact percentage of price increase or decrease in different periods to understand consumer decisions. The last part of the questionnaire was related to the respondents' views and the possibilities and probabilities of certain events, such as job loss in the next few years. Descriptive statistics were used to process closed-ended questions, and content analysis was used to analyse open-ended questions, intending to determine the presence of certain words or concepts within the text. Specifically, a conceptual analysis was used, whereby the text was coded into categories based on selective reduction. It ultimately consisted of words, sets of words, or phrases indicative of the analysis. Using the method of comparison, the obtained results were compared with the results of previous research, and a general conclusion was reached using the induction method.

4 Analysis

A total of 286 respondents participated in the research, where 23.8% of the respondents were male, while 76.2% of the respondents were female. Based on the rest of the answers collected related to the demographic characteristics of the respondents, it can be concluded that most of the respondents were in the age from between 15 and 24 years old, 20.3%. Furthermore, from the analysis of the level of education from the answers, it is evident that the largest share of respondents have completed a four-year high school (40.90%) and completed higher school or college (35.31%). Most respondents stated that they currently work full-time, and most were households with four people living in them (51.05%). In the context of analysing the financial situation of respondents in the analysed period, it is clear that most of the respondents live the same as 12 months ago; however, if the current situation was analysed with the situation five years ago, then it is clear that most of the respondents still live better or the same, and even because of all

macroeconomic shocks that were present in the analysed period. A comparison of the respondents' current financial situation with the financial situation one and five years ago is shown in Figure 3.

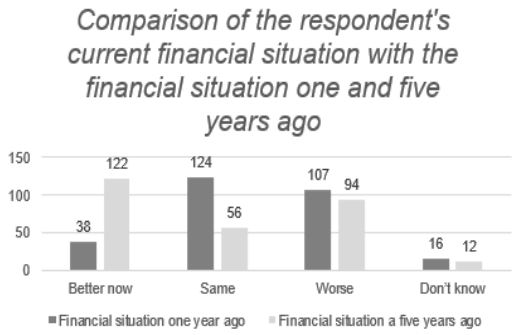


Figure 3. Comparison of the respondent's current financial situation with the financial situation one and five years ago (authors' representation according to the results of the survey)

Such results can be attributed to a rapid price increase for those who declared that they currently live worse than one or five years ago and to an increase in wages or employment for those who declared that they live better than one or five years ago. Below is the annual inflation for the previous five years. In their opinion, food and non-alcoholic beverages, as well as housing, water, electricity, gas and other fuels, furniture, home equipment and regular household maintenance, have become more expensive, significantly affecting the consumer basket. In addition, the arrival of a child is cited as a reason for the negative household budget, which specifically means higher salary charges and an increase in the prices of diapers, food for children, supplementary feeding, etc. The respondents pointed out that they believe food, non-alcoholic beverages, housing, and water have become more expensive. Electricity, gas and other fuels, furniture, home equipment and regular household maintenance, supported by the HCPI consumer price index in Croatia shows that these three components experienced the greatest price increase. According to the respondents ' opinions, figure 4 shows the elements that will increase in price the most in 2022.

On the other hand, the respondents who declared that they live better as a reason cite an increase in wages in their work sectors. For example, they cited the banking sector, which significantly eased price differences due to inflation. In addition to higher salaries, some respondents also cited career advancement and the employment of other household members as a reason for a better life, which meant additional income that could cover the increase in prices. Respondents' answers suggest a conclusion of relatively low financial literacy, given that most respondents did not think long-term about the household's financial position, given that the majority of respondents did not prepare for possible negative scenarios, such as a general increase in prices, and only reflexively reacted to the newly created situation in such a way that they sequentially tried to solve the problem mostly by reducing consumption. In connection with this, the majority believe that now is not the right time for investments or buying real estate, and they expect that the inflation rate will continue to rise at a higher rate than their incomes will grow and that some additional shocks with record high-interest rates are possible.

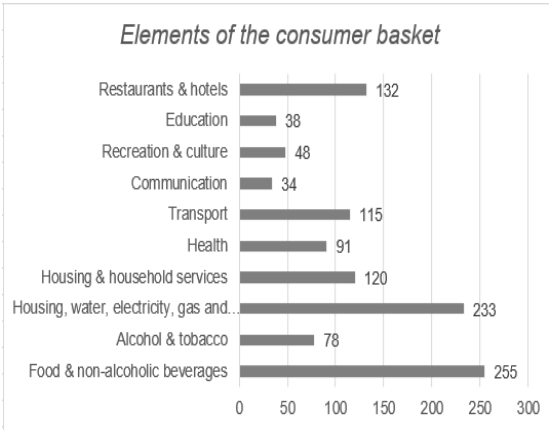


Figure 4. Elements of the consumer basket that will increase in price the most in 2022 (authors' representation according to the results of the survey)

5 Conclusion

The research problem of this paper was related to the examination of inflationary expectations of consumers in the Republic of Croatia. It was primarily initiated by the rising inflation rate in the Republic of Croatia in the last five years, from 2018-2022, with a record inflation rate of 13.1% in December 2022, which caused concern among Croatian citizens. The aspiration was to investigate whether citizens are aware of the long-term nature of consumer decisions in making financial decisions related to the household budget, which indirectly leads to certain conclusions about financial literacy. Therefore, it is very important to understand consumer expectations and their opinions about events in the future in order to be able to give certain guidelines for their improvement. Accordingly, a survey questionnaire was defined based on two American surveys, namely the survey conducted by the "Federal Reserve Bank of New York" and the survey conducted by "The University of Michigan."

The obtained results gave an insight into the consumer's perception regarding the movement of the inflation rate, thus answering the set RQ1. Namely, consumers in the Republic of Croatia mostly understand the very concept of inflation; they point out that they are aware of and have recognized the negative trends in the market, along with the identification of price increases of key elements that influence the

increase in the cost price of their consumer basket, and state that they expect further price increases, which is identical to existing research. What is additionally emphasized is that consumers believe that price growth does not, nor will it, sufficiently follow the general rise in prices, where there is room for the introduction of additional measures by legislative bodies and institutions that should act in the direction of establishing a better social balance for the working population as well, students, but especially vulnerable groups such as pensioners. With this, consumers point out quite a lot of concern and uncertainty in the coming period (RQ2). Certain developments are visible in the context of increasing wages for completed student work, raising wages in the state sector, and increasing pensions, but also the introduction of special measures in conditions of great uncertainty, such as the COVID-19 pandemic, and later certain measures during the expansive rise in the prices of oil derivatives due to geopolitical instability in Europe. However, certain problems were also identified, especially in the segment of citizens' still low financial literacy (RQ3), which can be read from the respondents' answers. Namely, the answers indicate that consumers did not think enough about the possible negative outcomes in moments of positive developments in the markets (before the pandemic) in terms of savings and investments, so in crises, they had to react relatively spontaneously and reactively in terms of short-term solutions such as direct reduction spending, giving up, for example, buying real estate or any investment.

However, positive developments in this context are visible, given that consumers are aware of upcoming events in the financial sector, have recognized the risks and understand the concepts relatively well. The same is confirmed by the national examination of citizens' financial literacy conducted by Croatian Financial Services Supervisory Agency (HANFA) and CNB in 2023, which indicates that financial literacy in Croatia is continuously growing slightly. The overall rating of financial literacy in 2023 was 12 out of 20 points, i.e., the respondents, on average, answered 60% of the questions correctly. A comparison with two earlier surveys (from 2015 and 2019) shows that citizens' financial knowledge is continuously growing: from 60% in 2015 and 65% in 2019 to 70% in 2023 (HANFA, CNB, 2023). However, there is certainly room for progress. Hence, the Government of the Republic of Croatia adopted the National Strategic Framework for Consumer Financial Literacy for the period from 2021 to 2026, in which the priority is financial education in two pillars, formal and informal education, which will include those individuals who are not included in any form of formal education. It is intended as support for formal education in the context of which a series of measures will be implemented that include various workshops, seminars, the creation of educational materials and the use of various forms of informing the public, taking into account the specific financial needs of each target group (children and youth, younger adults, people of working age who earn income and older people) (National strategic framework of consumer financial literacy for the period from 2021 to 2026, NN 68/21 (NN 68/21)).

References

Armantier, O., Bruine de Bruin, W., Potter, S., Topa, G., Van Der Klaauw, W. & Zafar, B. (2013). Measuring inflation expectations. Annu. Rev. Econ., 5(1), 273-301.

Babić, M. (1996). Makroekonomija. Zagreb: Mate.

Blanchard, O. (2011). Makroekonomija. Zagreb: Mate.

Croatian Bureau of Statistics (CBS) (2023) Inflation calculator. Retrieved from https://web.dzs.hr/calcinfl.htm

Croatian Bureau of Statistics (CBS) (2023). Consumer price indices in October 2022. Retrieved from https://podaci.dzs.hr/2022/hr/29198

Croatian Financial Services Supervisory Agency, Croatian National Bank (2023). Measuring financial literacy and financial inclusion in Croatia – 2023. Retrieved from https://www.hanfa.hr/media/9662/istra%C5%BEivanje-financijske-pismenosti-2023.pdf

Croatian National Bank (CNB) (2021). Retrieved from https://www.hnb.hr/en/-/sto-je-inflacija

Croatian National Bank (CNB) Effect of the War in Ukraine on Consumer Price Inflation (2022). Retrieved from https://www.hnb.hr/en/javnost-rada/rat-u-ukrajini/html/-/asset_publisher/lwOTLx5fWZDC/content/utjecaj-rata-u-ukrajini-na-inflaciju-potrosackih-cijena

Croatian National Bank (CNB) GDP growth to decelerate to 1.4%, inflation to drop to 7.5%?. (2022). Retrieved from https://www.hnb.hr/en/-/rast-bdp-a-u-2023-usporava-se-na-

Croatian National Bank (CNB) Why has inflation increased significantly in recent months?. (2022). Retrieved from https://www.hnb.hr/javnost-rada/aktualno-o-inflaciji/html

European Central Bank (ECB) Euro area statistics (2023) Retrieved from https://www.euro-area-statistics.org/digital-publication/statistics-insights-inflation/bloc-1.html?lang=hr

European Central Bank (ECB) Macroeconomic projections (2023). Retrieved from https://www.ecb.europa.eu/pub/projections/html/ecb.projections202303_ecbstaff~77c0227058.hr.html#toc2

European Central Bank (ECB) Measuring inflation – the Harmonised Index of Consumer Prices (HICP) (2023). Retrieved from https://www.ecb.europa.eu/stats/macroeconomic_and_sectoral/hicp/html/index.en.html

Galí, J. (2015). Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications. Princeton University Press.

National strategic framework of consumer financial literacy for the period from 2021 to 2026 NN 68/21 (NN 68/21). Retrieved from https://narodne-novine.nn.hr/clanci/sluzbeni/2021_06_68_1316.html

Poslovni.hr (2012). 12 years ago, we had HRK 2,000 less salary, but we could buy more gasoline than today. Retrieved from https://www.poslovni.hr/vijesti/prije-12-godina-imali-smo-2000-kn-manju-placu-ali-smo-mogli-kupiti-vise-benzina-nego-danas-198799

Sims, C. A. (2009). Inflation expectations, uncertainty and monetary policy.

Woodford, M. (2005). Central bank communication and policy effectiveness.

Žmuk, B. & Bonić, A. (2022). Analiza stavova hrvatskih građana o načinima očuvanja vrijednosti imovine u razdobljima visoke stope inflacije. Zbornik Ekonomskog fakulteta u Zagrebu, 20(2), 15-31.

Intelligent Energy

Stefanie Hatzl, Katharina Moitzi
Identification of technology adoption paths for smart technologies

Identification of technology adoption paths for smart technologies

Stefanie Hatzl, Katharina Moitzi
CAMPUS 02 University of Applied Sciences
IT & Business Informatics
Körblergasse 126, 8010 Graz, Austria
{stefanie.hatzl, katharina.moitzi}@campus02.at

Abstract. *Smart technologies are a key technology to ensure a transformation towards a decarbonized, decentralized and digital energy market. Therefore, we explored the technology adoption paths of smart home and smart energy technologies. Based on a mixed-methods approach, both qualitative and quantitative data were collected on the chronological order of adoption of different smart technologies in households. The results aim to provide insights on how to design and promote new products and services. Thus, by knowing which technologies are already adopted, co-adoption concepts can be established between the offered products and the existing technologies.*

Keywords. Energy transition, Smart home technologies, Smart energy technologies, Technology adoption path

1 Introduction

With the Green Deal, the European Union strives to strengthen economic resilience by transforming into a climate-neutral continent. The decarbonization is also called a "twin transition" because this overarching goal is to be achieved on the one hand through digitalisation (i.e., development of a fully integrated, interconnected, and digitized EU energy market) and on the other hand through decentralization (i.e., diffusion of renewable energy sources) (European Commission, 2022). At the household level, decentralization and digitalization can be seen, among other things, through the adoption of renewable energy technologies such as photovoltaics (PV) in combination with digital technologies, which are mentioned in the context of the concepts smart home or smart energy (Paukstadt, 2019).

Smart home technologies (SHT) include sensors, networked devices, and integrated systems (i.e., ICT) that provide services to residents (e.g., monitoring, control from a distance, automatic control) with the purpose of satisfying different needs to enhance quality of life (Balta-Ozkan, et al., 2014). These needs include comfort, health, safety, support, entertainment, or efficiency (Marikyan et al., 2019; Aldrich, 2003). In contrast Smart Energy Technologies (SET) are defined as technologies "[…] serving either one of the two common system goals of 1) increasing energy efficiency or 2) increasing the integration of renewable energy sources" (Gimpel et al., 2020). In respect to the Green Deal, households are seen as active contributors in the energy transition who, with their adoption and decision to use SET, can significantly hinder or promote the mentioned potential (BMNT, 2019).

The smart home as buzzword for the digitalization of the domestic environment holds a functional, instrumental and socio-technical perspective according to the literature. While the functional perspective describes various services that a smart home fulfils (e.g., enhancing security, support elderly, improve comfort, manage energy), the instrumental perspective takes a focus on the smart home as instrument to achieve energy efficiency. The socio-technical understands the smart home as an overall transformation of living and thus, refers to a future perspective (Wilson et al., 2015). The functional perspective is more related to definitions of SET in a broader sense, i.e., functions of smart home to meet user needs such as cost efficiency, comfort, emotion, security, or health (Marikyan et al., 2019). In contrast, the instrumental perspective relates more to the definition of SET, i.e., individual devices to system solutions that provide feedback on energy use, manage energy consumption, and control home appliances and applications (Karlin et al., 2015).

Based on these two perspectives for SHT and SET the question arise what do technology adoption paths (TAP) of households look like? There are hardly any studies that have examined in more detail when, which type of technology found its way into a household and for what reason. Sanguinetti et al. (2018) examined, based on quantitative survey, the path to smart home adoption by clustering user segments in respect to awareness, interest, and ownership of smart technologies. Thus, they considered SHT as well as SET along the decision-making process, but did not question the chronological order of adoption in respect to only the already owned technologies.

An investigation of the chronological order of adopting different SHT or SET (e.g., smart voice assistant, smart heating, smart TV) can provide additional understanding of TAP. The results aim to guide companies in designing new product bundles to successfully enter or penetrate the market.

The rest of the paper is organized as follows. Section 2 gives an overview of the related research. Here classifications and characterizations of SHT and SET are presented, and more insight is given to the adoption literature. Section 3 explains the empirical research design – a mixed methods approach – to elaborate the research question. Findings are described in section 4, while section 5 discusses them finally.

2 Related Research

2.1 SHT & SET

In Europe, the penetration rate, i.e., the number of smart homes as a proportion of the total number of European households, is (estimated) at 18% in 2022. Austrian households are above this at 26.6%. A subcategory of smart home is energy management (excluding lighting and smart home appliances), which is only 7% in Europe and 11.5% in Austrian households. Future penetration rates are expected to increase. (Statista, 2022). Thus, a variety of different technologies will characterize the smart home.

In the literature, there are different categorisations and characterisations of this multitude of different products, services, and system solutions for a smart household. Sovacool and Del Rio (2020) assign 267 SHT on the UK and EU markets in 2019 to 13 categories: household appliances, lighting, energy & utilities, entertainment, health & wellness, safety & security, baby & pet monitors, clothes & accessories, vehicles & drones, home robots, gardening, integrated solutions, and "others". In addition, SHT are grouped into different categories as part of market analyses. Thus, the Smart Home Study conducted by GfK (2016) grouped the variety of smart home devices in the categories: smart domestic appliances, smart entertainment & connectivity, smart energy & lighting, smart security & control and smart health. In comparison Statista (2022) groups SHT in the six segments: smart appliances, home entertainment, control & connectivity, energy management, comfort & lighting and security. Pritoni et al. (2018) focus directly on SET – i.e., smart home technologies with the purpose of energy management – and characterise 308 technologies on the US market in 2015-2016 according to categories such as product components, hardware, software, communication, information - feedback. In their database they divide the technologies into the following components: smart appliances, smart thermostats, smart lighting, smart plug/switch, smart hub, in home display, energy portal, load monitor and embedded display (Pritoni et al., 2018). In contrast Marikyan et al. (2019) characterise smart homes, suggesting a grouping by type of service (control/monitor, energy management, support and assistance, anticipate and respond), and addressed user needs (cost efficiency, comfort, emotion, safety, health, quality of life, sustainability), in addition to a purely technology-based classification into sensors, devices and integrated systems.

Studies on the classification of individual products in the smart home sector revealed that all these different technologies contribute to different levels of smartness. These levels of smartness range from traditional homes without the installation of smart devices, to stand-alone separate smart applications with varying degrees of interconnectedness, to fully integrated system solutions that adapt to the context. The degree of smartness is related to the degree of human interaction with smart technologies. While energy savings and efficiency with separate smart devices or smaller device bundles are mostly based on feedback mechanisms and subsequent behavioral decisions, more connected system solutions can implement this via automation and ultimately AI-supported control (Sovacool & Del Rio, 2020; Marikyan et al., 2019). Thus, there are two important characteristics of smart devices in terms of energy management on the one hand the level of smartness - i.e., the degree of connectivity - and on the other hand the energy saving potential.

A positive correlation between the degree of smartness and the savings potential is postulated in existing studies (Strother et al., 2013; Williams & Matthews, 2007). Ford et al. (2017) argue that feedback mechanisms, and thus behaviour-based energy savings, have less potential for energy savings than automation or load-shifting mechanisms. Technologies out of the categories of lighting, household appliances, integrated systems or energy management systems have a good energy saving potential on behavioral and/or operational level. Technologies providing feedback with a display, or a connected app help to save energy on behavioral side. Technology using remote, scheduled or rule-based control helps to reduce usage and enable savings. On operational level automation of control, sensor-based control or intelligent learning algorithms drive energy savings and efficiency (Ford et al., 2017).

2.2 Adoption of smart technologies

The literature contains many studies that examine factors that determine acceptance and adoption of smart technologies (e.g., Kastner et al., 2015; Mashal et al., 2023). The focus lies on the explanation of the intention or the actual adoption of separate devices (e.g., adoption of PV by Tanveer et al., 2021; smart heating and smart meter by Große-Kreul, 2022) to product categories (e.g., adoption of smart home by Shuhaiber & Mashal, 2019; of HEM by Chen et al., 2020; of green technologies by Girod et al., 2017). The

testing of existing theories (e.g., Technology Acceptance Model (TAM), Unified Theory of Acceptance and Use (UTAUT)) show, the impact of predictors such as perceived usefulness or perceived ease of use (Mashal et al., 2023).

According to the decision-making process in respect to the diffusion of innovation theory compatibility in terms of values, norms, preferences and experiences play an essential role in adoption (Rogers, 1995; Saaksjarvi, 2003). Compatibility and knowledge about a technology significantly influence the interest in considering adoption in the first place. Based on the analogical learning theory, consumers not only need knowledge in the main technology group, but also knowledge that has been made about a supplementary technology category influences adoption (Saaksjarvi, 2003). This leads to the assumption that interest in SET can also be stimulated via adoption of SHT and vice versa.

A deeper understanding of reasons and motives for using smart home is shown by qualitative research results, where energy management and saving energy by controlling appliances is named as one of the most important benefits or purposes of smart home. Others are comfort and convenience, financial savings, health benefits from mediation management to taking care of personal hygiene, entertainment and enhancing leisure by providing fun or interacting with others remotely, aesthetics and symbolic value, free services or promotional gifts (FakhrHosseini et al., 2021; Sovacool & Del Rio, 2020; Li et al, 2021; Wilson, 2017).

While many studies have characterized and categorized smart technologies and examined their drivers and barriers in terms of acceptance and adoption, there are hardly any studies that have examined TAP in respect to the chronological order of different technology types. Under the term "co-adoption," sporadic previous literature examines the adoption and use of multiple technologies at the household level in terms of product bundles. For example, Lagomarsino et al. (2023) show the potential of low-carbon technologies (i.e., PV solar system, heat pump, hybrid and electric vehicle, stationary battery) in co-adoption and the existence of co-adoption patterns over time. In the present article, a step further is taken by considering different SHT and SET in their chronological adoption order.

3 Methods

For this study we used a mixed methods approach, first having a qualitative part with interviews followed by a quantitative part with a quota-representative online survey. The aim was a triangulation of data to achieve a comprehensive understanding regarding the type of technologies adopted, the chronological order of adoption and reasons.

As basis for both, the qualitative and quantitative part a list of smart technologies that can potentially be adopted by households has been developed. Therefore, a typology of smart technologies (i.e., SHT and SET) at the household level was created based on literature review (e.g., Pritoni et al., 2018, GfK, 2016 and Sovacool and Del Rio (2020). The potential smart technologies were assigned to a categorization and degree of interconnectivity.

In the qualitative part, semi-structured interviews were conducted with 16 individuals. A convenient sample (Patton, 1990) was used. Criteria for selecting subjects were ownership of SHT or SET and heterogeneity in age, sex, education, and living situation, see table 1.

Table 1. Characteristics of the subjects interviewed. Following abbreviations are used: m = masculine, w = feminine, w/o = without higher education, degree = with at least high school degree

ID	Age	Sex	Education	Home ownership	No. of persons in the household
I1	42	m	degree	apartment	2
I2	40	m	w/o	house	4
I3	46	m	degree	apartment	3
I4	38	m	degree	house	3
I5	29	m	degree	house	2
I6	36	m	degree	apartment	2
I7	37	f	degree	apartment	4
I8	49	f	w/o	house	1
I9	20	f	degree	rent	2
I10	36	m	degree	rent	1
I11	58	f	w/o	house	2
I12	46	m	degree	house	2
I13	51	m	degree	house	1
I14	21	f	degree	rent	6
I15	51	m	degree	house	5
I16	36	m	degree	apartment	4

In the information phase, the aim of the interview was presented, and data protection aspects were explained in order to obtain informed consent. It was further explained that the interview is about smart devices in general, i.e., both SHT and SET. In the main phase, the interviewees were shown cards on which 35 different technologies were written. After that, they were asked to select those cards that included technologies that were already present in their households. The interviewees were asked in what temporal order the technologies were adopted and they should place the cards accordingly. Based on the TAP laid, individuals were asked to talk about reasons for

adopting each technology. Depending on the mentioned reasons, a further question was, if the adoption of one technology had caused the adoption of another technology. Finally, they were asked which technologies should be acquired in the future and why. The last part of the interview included a short survey about sociodemographic aspects and attitudinal characteristics according to the sampling criteria.

The interviews take place from November 2022 to January 2023 face-to-face. The interviews were recorded and then transcribed. Important visual findings, such as TAP, were captured through photos. The transcripts and photos were analysed using qualitative content analysis according to Mayring (2015) with the assistance of MAXQDA software.

Second, an online survey was conducted out among the Austrian population to complement the qualitative results. Quota sampling was chosen as the sampling method. For the characteristics age, gender and education, quotas were set according to Statistics Austria for a representative sample of the Austrian population.

The questionnaire comprised 2 parts. In the first part, sociodemographic characteristics (i.e., age, gender, education, employment status, state, type of housing, number of people in the household, income) were asked. In the second part, subjects were asked to select from a list those smart technologies that are present in their household and then to sort these selected technologies (if more than 2) according to their time of acquisition. For the technologies that have not yet been adopted, respondents were asked to what extent they were interested in purchasing them, measured on a 5-point Likert scale from (no interest to highly interested). In contrast to the interviews a limited number of technologies were asked due to usability reasons.

The survey was implemented online using the software Unipark and was conducted by a third-party provider (i.e., market research institute Norstat GmbH) with an online panel from February 13 to February 24, 2023. The results were analysed descriptively. A total of 848 people were interviewed across Austria, of which 59.2% were male, 48.9% female and 0.8% diverse. The average age is 40 years. 66% are without high school degree and 34% with high school degree.

4 Results

The list of different smart technologies (see table 2) included SHT (e.g., voice assistant, smart watch, smart fridge, smart entertainment) and SET (e.g., smart meter, smart heating, air conditioning, LED, etc.). These were assigned to 6 technology categories, that reflect different energy management potentials.

- Entertainment & Personal Technology & Health
- Safety & Security
- Appliances & Robots
- Lighting
- Energy Management
- Energy Production & Storage
- Mobility

Furthermore, we defined three integration levels: (i) enabler technology, (ii) integrated systems (not professional) and (iii) highly integrated systems (professional installation). These integration levels indicate that different degrees of interconnectedness can exist for one technology from the categories (e.g., voice assistant can stand alone, but also can control devices from the different categories as a system solution).

The results of the 16 interviews show according to table 2 that the respondents especially have acquired smart technologies in the areas of "lighting" (31) and "energy management" (35) on the one hand, and " appliances & robots" (34) and "entertainment & personal technology & health" (36) on the other. In respect to the category “appliances & robots”, there are many different mentions, with vacuum cleaners and lawn robots being more common. In the “entertainment & personal technology & health” category, smart watch and smart entertainment were most frequently mentioned. Smart meters, smart heating and smart thermostats are frequently mentioned in the “energy management” category, while smart lighting and motions sensors are common adopted technologies in the category “lighting”.

The temporal order of adoption shows that the mentioned technologies, which were acquired as first, second, or third, account for 44% of the total mentioned smart technologies. Smart heating and thermostats and motion detectors or smart lighting are particularly noteworthy, but smart entertainment (i.e., smart TV), smart watches, and voice assistants also fall into this early adoption phase of the respondents. Thus, SHT as well as SET are named, that find their way into the home at an early stage and thus function as basic technologies.

The technology adoption paths described by respondents show differences in the number of technologies adopted and the number of technologies purchased at any one time. For example, one interviewee describes the acquisition of 15 different technologies at 13 successive points in time. Another interviewee, on the other hand, describes the acquisition of 19 technologies at 5 successive points in time, whereas another person states having acquired 5 technologies at 2 successive points in time. On the one hand, this shows the extent to which the different technologies have already found their way into the household and which group of technologies is predominant.

In a few cases, a direct link between the experience with one technology and the decision to adopt another technology from the same or different technology category based on that experience was mentioned. Two interviewees got a voice assistant as a promotional gift

and mentioned that they were interested in the features of and possibilities with this technology. Subsequently both invested in smart lights. While trying out the new technologies security and privacy concerns arose. One statement was: *„The heating via heat pump has quite triggered the purchase of PV, […] and the PV can dampen the electricity costs a little over the year.”*, which represents an example for co-adoption. Another interesting finding is that, one mentioned that the smart home was intended to control blinds, lights and heating, but because having such system, future investments in technology were made in consideration to the existing system and therefore influenced the purchase decisions. The smart phone was also mentioned twice as trigger for smart watch and general interest in smart technologies.

Table 2. Absolute frequencies of technologies mentioned in total, and technologies adopted first, second and third

Category	Technology	Total	Order		
			1.	2.	3.
Entertainment & Personal Technology & Health	Smart entertainment	13	3	5	3
	Voice assistant	9	1	3	1
	Smart watch	12	2	1	2
	Smart health and assistance systems	2	-	1	-
Safety & Security	Smart security devices/systems	5	-	1	1
Appliances & Robots	Smart coffee maker	1	-	-	-
	Smart kettle	2	-	-	2
	Smart food processor	2	-	1	-
	Smart fridge	1	-	1	-
	Smart washing machine	4	-	-	-
	Smart dishwasher	1	-	-	-
	Smart oven/stove	2	-	1	1
	Smart tumble dryer	4	-	-	1
	Lawn robot	6	1	1	1
	Robot vacuum cleaner	10	1	-	2
Lighting	Motion sensor	10	3	5	-
	Dimmer	11	3	1	-
	Smart lighting / LED	10	1	2	1
Energy Management	Smart meter	7	1	-	1
	Smart air conditioner	1	-	-	-
	Smart heating	10	3	1	1
	Smart thermostat	8	3	1	-
	Smart plug	4	-	1	-
	Smart water boiler	3	-	1	1
	Smart blinds	3	-	1	-
Energy Production & Storage	PV system	6	-	-	-
	Solar thermal system	4	2	1	-
	Power storage	1	-	-	-
Integrated Systems	HEM system	2	-	-	-
	Smart home system	3	-	2	-
Mobility	E-Car	1	-	-	-
	E-Bike/E-Scooter/similar	8	-	1	-
	Total (absolute)	166	24	32	18
	% share of all technologies adopted	14	19	11	

The reasons for adopting the SHT and SET were manifold. In addition to the enthusiasm for the technology and statements like "*I bought it simply because I think it's cool*" the following reasons were broadly summarized:

- Energy saving or energy efficiency
- Practical reasons such as facilitation and time savings
- Comfort and convenience
- Monitoring
- New acquisition without a focus on smart features, as these features were included in the product
- Already present in the building at the time of house/apartment purchase
- Gift from friends/family or from the manufacturer (as part of a product test)
- Specific problem-solving like:
  - Targeted control of devices at home (e.g., aquarium pump, Christmas lighting)
  - Remote control of devices (e.g., simulating presence during vacation, controlling heating or blinds)

In the case of adoption without attention to smart functions, some respondents explained that there was no alternative/non-smart technology available, and in order they were forced to buy something smart/energy-saving. Other respondents mentioned that it was simply not consciously intended because other selection criteria were in focus during the purchase. Some interesting quotes from the interviews in this respect are: "*I didn't want that to happen*" or "*[…] with the TV it is simply a development. It was more of a necessary evil to buy a smart TV*" or "*the smart thing just came along at some point without actively wanting it*". Thus, the findings show, that a quarter of respondents have devices with smart features, but do not use them and a quarter of respondents use the smart features of the device, although it was purchased without attention to them.

Even in the case of the voice assistant, the decision to use it is not always supported by all household members or generally actively decided. When asked whether a voice assistant was owned, one respondent mentioned: "*[…] social pressure in the family. From my point of view, we wouldn't have anything like this, but kids and wife wanted an Alexa*".

Interesting thoughts regarding the purchasing because of energy saving are *"[…] because we want to produce our own energy and also try to consume this energy ourselves and to align our consumption with the production of this energy, so we can save energy"* or *"[…] not the smart features were in the focus, but the*

quality and energy consumption of the device" or *"[…] you can certainly make the energy consumption more efficient and there is potential for savings"*. Some respondents even indicated on their own that they would have purchased a smart meter without compulsory installation by energy providers.

There were also respondents indicating that they own and tried a smart device, but the features are now no longer used, based on privacy concerns or lack of interest.

The results of the survey (see figure 1) show that smart entertainment and household technologies, as well as smart lighting, emerge as pioneers in many households. In contrast e-cars and smart home solutions or energy management systems are still not present. When a technology is not yet present in a household, we inquired about the level of interest in potential adoption. In addition to the aforementioned technologies, there is a clear trend of interest in energy management technologies such as electricity production and storage. Smart heating and lighting systems are also of interest. However, there is relatively little enthusiasm for electric vehicles and smart home solutions. Interestingly, although voice assistants are already well-established in households, the interest in them is relatively low.

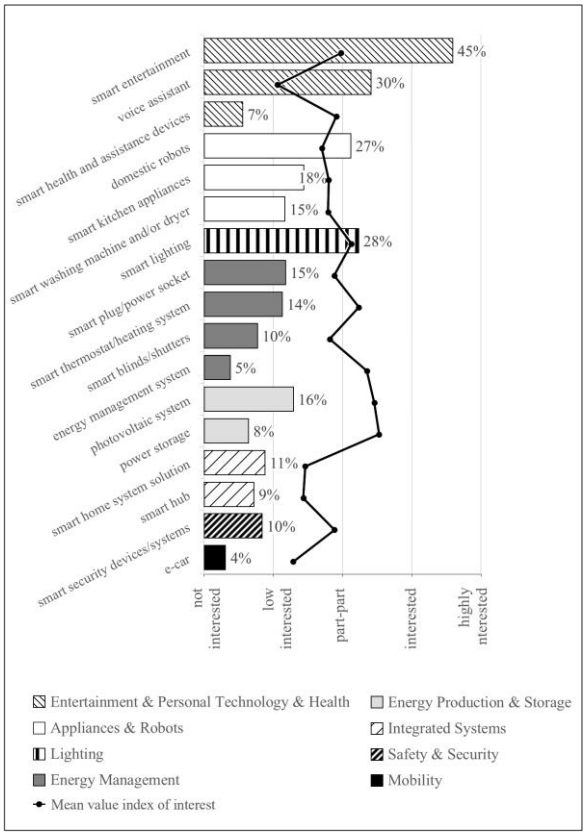


Figure 1. Percentage of people who already own technology and mean index of interest in technology

Table 3 shows that from the temporal order of adoption, it appears that respondents invested especially in smart entertainment and voice assistants first. In addition, domestic robots (e.g., lawn robot, robot vacuum cleaner) find their way into the household at an early stage. In terms of SET, smart lighting is the first item to be purchased.

Table 3. Excerpt from the chronological order of technology adoption

Technology	Frequencies Adoption order					
	1.	**2.**	**3.**	**4.**	**5.**	**6.**
smart entertainment	146	79	34	12	11	6
voice assistant	80	63	35	12	4	9
domestic robots	59	47	30	22	13	6
smart lighting	58	58	35	20	10	5
smart kitchen appliances	44	25	17	17	11	3
smart washing machine and/or dryer	30	26	18	10	8	5
photovoltaic system	30	28	9	9	9	6
smart thermostat/heating system	29	17	19	8	2	6
smart home system	29	12	10	12	9	2
smart plug	22	26	20	19	8	5
smart blinds/shutters	22	9	10	10	8	2
smart security devices/systems	20	11	14	10	6	5
smart hub	18	7	15	8	6	3
smart health and assistance devices	14	7	6	9	4	1
HEM system	12	3	5	3	2	2
power storage	11	9	9	4	7	2
e-car	3	2	5	9	1	1

5 Discussion & Conclusion

This study investigated TAP of smart technologies at the household level. The aim was to understand the chronological order of adoption and reasons for the adoption of different SHTs and SETs. The results showed that both SET and SHT are at the beginning of TAP. Smart entertainment, voice assistant and smart lighting were often ranked first, second or third in the interviews as well as in the survey. While smart heating and smart thermostats were also mentioned in the interviews, smart appliances and robots were a frequently mentioned technology at the beginning in the survey. Ultimately, the TAPs described are very heterogeneous in terms of technologies. Nevertheless, the insight into the various technology adoption paths provides a deeper understanding of households' actual decision-making behavior. In addition to the technologies already adopted by households, there is a clear trend of interest in energy management technologies.

Reasons for adoption such as comfort, energy savings, or promotional gifts are similar to the reasons given in existing studies such as that of Sovacool and Del Rio (2020). Interestingly, the adoption of smart devices was often not intended. The need for smart technologies, and thus the decision to adopt them, can be caused by the technology itself (Rogers, 1995). This can be seen in the example of the adopted smart TV. The initial trigger for buying the smart device was that the old device was broken, with the smart functions being secondary. The experience with the new smart device then created a need for smart functions.

On the one hand, the interviews show an example of co-adoption, i.e., a technology adoption path was characterized by the fact that the adoption of a heat pump prompted the adoption of PV as a complementary technology (i.e., adoption of technologies in the SET category). On the other hand, another interview revealed an example of adoption of different technology categories. Due to the voice assistant (i.e., SHT) that was introduced into the home as a promotional gift, the interest for smart lighting was aroused (i.e., SET since this technology enables more efficient energy consumption). Nevertheless, the interviews only rudimentarily showed that a conscious decision for one technology is made on the basis of another specific technology, regardless of the technology category.

By knowing which technologies are already adopted by the households, targeted points of connection can be created between the products offered and already adopted technologies. Thus, the findings can be utilized for targeted marketing strategies, enabling companies to effectively promote their offerings based on the specific technology adoption patterns observed. On the scientific side, scenarios for the diffusion of smart technologies could be developed based on the results - i.e., information on ownership and interest in smart technologies (smart home and SET).

Although a heterogeneous group of people for the interviews were recruited, there are still limitations. Additional interviews could provide better theoretical saturation and thus shed light on the trends regarding the influence of one technology (category) on another technology (category) in the TAP. In addition, future research could identify patterns based on the quantitative data on TAPs. The aim would be to use statistical analysis to determine whether the likelihood of adoption increases with the ownership of another technology.

Acknowledgments

This work was supported by the Call “Green Transformation”, a funding by Federal State of Styria.

References

Aldrich, F.K., 2003. Smart Homes: Past, Present and Future. In Harper, R. (ed.), *Inside the Smart Home* (17-39). Springer.

Balta-Ozkan, N., Boteler, B., & Amerighi, O. (2014). European smart home market development: Public views on technical and economic aspects across the United Kingdom, Germany and Italy. Energy Research & Social Science, 3, 65–77.

Bundesministerium für Nachhaltigkeit und Tourismus (BMNT). (18. Dezember 2019). Integrierter nationaler Energie- und Klimaplan für Österreich. Retrieved from https://energy.ec.europa.eu/system/files/2020-01/at_final_necp_main_de_0.pdf

Chen, C., Xu, X., Adams, J., Brannon, J., Li, F., & Walzem, A. (2020). When East meets West: Understanding residents’ home energy management system adoption intention and willingness to pay in Japan and the United States. *Energy Research & Social Science, 69*. https://doi.org/10.1016/j.erss.2020.101616

European Commission. (18. 10 2022). Digitalising the energy system - EU action plan. Retrieved from https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13141-Digitalising-the-energy-sector-EU-action-plan_en

FakhrHosseini, S., Lee, S.-H., Rudnik, J., Son, H., Lee, C., & Coughlin, J. (2021). User Needs of Smart Home Services. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65*(1), 457-461. https://doi.org/10.1177/1071181321651218

Ford, R., Pritoni, M., Sanguinetti, A., & Karlin, B. (2017). Categories and functionality of smart home technology for energy management. *Building and Environment, 123*, 543–554. https://doi.org/10.1016/j.buildenv.2017.07.020

GfK. (2016). GfK Future of Smart Home Study: Global Report.

Gimpel, H., Graf, V., & Graf-Drasch, V. (2020). A comprehensive model for individuals’ acceptance of smart energy technology – A meta-analysis. *Energy Policy, 138*. https://doi.org/10.1016/j.enpol.2019.111196

Girod, B., Mayer, S., & Nägele, F. (2017). Economic versus belief-based models: Shedding light on the adoption of novel green technologies. *Energy Policy, 101*, 415–426. https://doi.org/10.1016/j.enpol.2016.09.065

Große-Kreul, F. (2022). What will drive household adoption of smart energy? Insights from a consumer acceptance study in Germany. *Utilities*

Policy, 75. https://doi.org/10.1016/j.jup.2021.101333

Kastner, I., & Stern, P. C. (2015). Examining the decision-making processes behind household energy investments: A review. *Energy Research & Social Science, 10*, 72–89. https://doi.org/10.1016/j.erss.2015.07.008

Lagomarsino, M., van der Kam, M., Rahmani Azad, Z., Parra, D., & Hahnel, U. J. (2023). Co-Adoption Pathways Toward a Low-Carbon Energy System. Advance online publication. https://doi.org/10.2139/ssrn.4443494

Li, W., Yigitcanlar, T., Erol, I., & Liu, A. (2021). Motivations, barriers and risks of smart home adoption: From systematic literature review to conceptual framework. *Energy Research & Social Science, 80*, 102211. https://doi.org/10.1016/j.erss.2021.102211

Marikyan, D., Papagiannidis, S., & Alamanos, E. (2019). A systematic review of the smart home literature: A user perspective. *Technological Forecasting and Social Change, 138*, 139–154. https://doi.org/10.1016/j.techfore.2018.08.015

Mashal, I., Shuhaiber, A., & AL-Khatib, A. w. (2023). User acceptance and adoption of smart homes: A decade long systematic literature review. *International Journal of Data and Network Science, 7*(2), 533–552. https://doi.org/10.5267/j.ijdns.2023.3.017

Mayring, P. (2015). Qualitative Inhaltsanalyse: Grundlagen und Techniken (12., überarbeitete Auflage). Beltz.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage Publications, Inc.

Paukstadt, U. (2019). A Survey on Smart Energy Services for Private Households. *Proceedings of the 14th International Conference on Wirtschaftsinformatik*, 1448-1462. Retrieved from https://aisel.aisnet.org/wi2019/track12/papers/7/

Pritoni, M., Ford, R., Karlin, B, & Sanguinetti, A. (2018). Home energy management (HEM) database: A list with coded attributes of 308 devices commercially available in the US. *Data in Brief, 16*, 71-74. https://doi.org/10.1016/j.dib.2017.10.067.

Rogers, E.M. (1995). Diffusion of Innovations. 4th Edition, the Free Press, New York.

Sanguinetti, A., Karlin, B., & Ford, R. (2018). Understanding the path to smart home adoption: Segmenting and describing consumers across the innovation-decision process. *Energy Research & Social Science, 46*, 274-283. https://doi.org/10.1016/j.erss.2018.08.002

Saaksjarvi, M. (2003). Consumer adoption of technological innovations. *European Journal of Innovation Management, 6*(2), 90–100. https://doi.org/10.1108/14601060310475246

Shuhaiber, A., & Mashal, I. (2019). Understanding users' acceptance of smart homes. *Technology in Society, 58*, 101110. https://doi.org/10.1016/j.techsoc.2019.01.003

Sovacool, B. K., & Del Furszyfer Rio, D. D. (2020). Smart home technologies in Europe: A critical review of concepts, benefits, risks and policies. *Renewable and Sustainable Energy Reviews, 120*, 109663. https://doi.org/10.1016/j.rser.2019.109663

Statista (2022). Smart Home. Retrieved from https://de.statista.com/outlook/dmo/smart-home/oesterreich#market-shares

Strother, N. & Lockhart B. (2013). In-home Displays, Networked Hem Systems, Stand-alone Hem Systems, Web Portals, and Paper Bill Hem Reports, Global Market Analysis and Forecasts, Navigant Consulting Inc, Chicago, IL.

Tanveer, A., Zeng, S., Irfan, M., & Peng, R. (2021). Do Perceived Risk, Perception of Self-Efficacy, and Openness to Technology Matter for Solar PV Adoption? An Application of the Extended Theory of Planned Behavior. *Energies, 14*(16), 5008. https://doi.org/10.3390/en14165008

Williams, E.D. & Matthews, H.S. (2007). Scoping the potential of monitoring and control technologies to reduce energy use in homes. *Proceedings of the IEEE International Symposium on Electronics and the Environment*, 2007, 239-244.

Wilson, C., Hargreaves, T., & Hauxwell-Baldwin, R. (2015). Smart homes and their users: a systematic analysis and key challenges. *Personal and Ubiquitous Computing, 19*(2), 463-476. https://doi.org/10.1007/s00779-014-0813-0

Wilson, C., Hargreaves, T., & Hauxwell-Baldwin, R. (2017). Benefits and risks of smart home technologies. *Energy Policy, 103*, 72–83. https://doi.org/10.1016/j.enpol.2016.12.047

Intelligent Systems

Dennis Rall, Thomas Fraunholz, Bernhard Bauer
AI-Democratization: From Data-first to Human-first AI

Martina Ptiček, Jasminka Dobša
Annotation and Machine Identification of Metaphors in Croatian Newspaper Articles

Victor Mitrana, Andrei Paun, Mihaela Paun, Jose Ramon Sanchez Couso
Some Remarks on the Formal Operations Inspired by the Gene Assembly in Ciliates

Juraj Ďuďák, Gabriel Gašpar, Jakub Perička, Eva Vangelová, Daniel Gurín
Design of system for collecting and analyzing of human gait parameters

Filip Babić, Petra Grd, Igor Tomičić, Ena Barčić
Deep Transfer Learning for Detecting Face Images with Moustache

Vlatka Davidović, Sanda Martinčić Ipšić
Towards Sequence-to-Sequence Neural Model for Croatian Abstractive Summarization

Marko Horvat, Ljudevit Jelečević, Gordan Gledec
Comparative Analysis of YOLOv5 and YOLOv6 Models Performance for Object Classification on Open Infrastructure: Insights and Recommendations

Gaurish Thakkar, Marko Tadić, German Rigau
ComFac4LT – Computing Facilities for Language Technologies

Gabriel Gašpar, Juraj Dudak, Ivan Sladek, Ingrid Zuziakova, Tomas Meravy
Utilization of Open-Source in Upgrading Technological Devices for Industry 4.0 readiness

Ivana Kovačević, Mihaela Orić, Ivana Hartmann Tolić, Emmanuel Karlo Nyarko
Modelling the Fermentation Process in Winemaking using Temperature and Specific Gravity

Amila Akagić, Emir Buza, Marko Horvat
Mapping RGB-to-NIR with Pix2Pix Image-to-Image Translation for Fire Detection Applications

Ena Barčić, Petra Grd
Optimizing Point Cloud Processing for Facial Reconstruction Using Open-Source Tools: a Preliminary Study

Marko Horvat, Klara Krmpotić, Andrija Krtalić, Amila Akagić
**Bridging Blockchain Technology and Humanitarian Demining:
A Novel Concept for Decentralized Storage of Landmine and UXO Locations**

Ivan Magdalenić, Elvis Popović, Neven Vrček, Boris Tomaš, Lovro Posarić
Localization of intruder UAV operator using Unmanned Aerial Vehicles

Dijana Oreški, Dunja Višnjić, Nikola Kadoić
Comparison of Machine Learning Algorithms on Noisy Data

Miran Pobar
YOLOv7 Model for Small Object Handling in Maritime Images

Karel Križnar, Jože Rožanec, Blaž Fortuna, Dunja Mladenić
Explainable Artificial Intelligence Meets Active Learning: A Novel GradCAM-based Active Learning Strategy

Vedran Orešković, Ana Meštrović, Slobodan Beliga
Towards Computational Content Analysis of Crises-Related News in Electronic Media

AI-Democratization: From Data-first to Human-first AI

Dennis Rall, Thomas Fraunholz
WOGRA AG
Hery-Park 3000, 86368 Gersthofen
{dennis.rall, thomas.fraunholz}@wogra.com

Bernhard Bauer
University of Augsburg
Software Methodologies for Distributed Systems
Universitaetsstrasse 6a, 86159 Augsburg
bernhard.bauer@informatik.uni-augsburg.de

Abstract. *In recent years, there have been significant advances in artificial intelligence (AI) research. However, not everyone can benefit from these advancements because it requires a significant amount of expertise to apply them. The concept of democratizing AI, known as AI democratization, aims to make AI accessible to everyone. One exciting development in this field is the shift from a model-first to a data-first approach. Previously, the focus was on building and training complex models, which required a high level of expertise. However, new high-level frameworks have drastically simplified the process, making it possible to create machine learning (ML) models based on data descriptions automatically. This has paved the way for data-first AI, where the focus is on the data engineering rather than model creation. The next logical step in this progression is a human-first approach to AI. By simplifying the process of creating ML models even further and leveraging the insights gained from data-first AI, we can make AI accessible to a even broader range of users. We underpin our theoretical considerations with the development of the Open space for Machine Learning (Os4ML) platform and summarize our achievments so far in this endeavor. Our approach has the potential to significantly lower the barrier to entry for using AI and to accelerate AI adoption across a wide range of industries and domains.*

Keywords. Human-first AI, data-first AI, Artificial Intelligence, AutoML, No-Code ML, Low-Code ML

1 Introduction

The field of artificial intelligence (AI) has seen tremendous progress, with the potential for AI to be helpful in many different areas ("The AI Index Report", 2023). However, due to the high level of expertise required and the shortage of AI experts, many fields have not yet been able to benefit from this technology. To address this issue, there has been an increasing focus on AI democratization, which aims to make AI accessible to as many users as possible (Allen et al., 2019). The goal of AI democratization is to empower individuals who lack expertise in the field to leverage AI for their own benefit and to bring the benefits of AI to a wider audience.

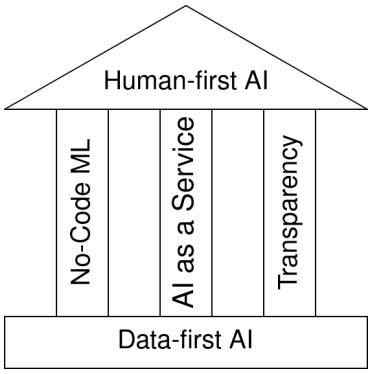


Figure 1: In order to advance from data-first AI to human-first AI, it is imperative to provide users with a no-code ML solution, make it available as a service, and ensure complete transparency throughout the development process.

One approach that has emerged in this context is data-first AI, also known as data-centric AI (Zha et al., 2023, "Data-centric Artificial Intelligence: A Survey"). Unlike traditional approaches that focused on designing complex ML models, data-first AI emphasiszes the importance of optimizing both the data's quality and quantity. Once high-quality data is obtained, low-code tools can be used to create ML models (Cabot, 2020). This approach has shown promising results, as it can lead to better accuracy and more robust models.

The data-first approach significantly reduces the need for AI expertise, but some level of technical knowledge is still required. To further democratize AI, the next step is to move towards human-first AI, which aims to make access to AI even easier. In this article, we will first discuss the data-first approach to AI and then explain how we can build human-first AI from it. Specifically, we will focus on three key components, as illustrated in Figure 1. First, we need to reduce the required coding effort even further and move from low-code to no-code ML. Second, we need to make access to this AI easier by offering it as a service. Third, the entire process needs to be transparent and trustworthy. Throughout this article, we will delve into each of these

components in detail. Finally, we will introduce our platform, Open space for Machine Learning, which incorporates these components and can help to build effective AI solutions.

2 Data-first AI

The data-first approach has shifted the focus of machine learning from designing models to data engineering tasks. This approach represents a significant improvement over the traditional model-first approach, which required developers to write low-level ML code to design models. Over time, higher-level tools and frameworks have emerged, which provide a more abstract interface for building ML models.

The data-first approach emphasizes the importance of developing and maintaining high-quality training data. This includes tasks such as data collection, data validation, data cleaning and data integration (Whang et al., 2023). By focusing on these tasks, developers can ensure that their models are based on accurate, diverse, and relevant data, which can improve their performance and accuracy.

In addition to developing the training data, data maintenance is also essential in the data-first approach. This includes tasks such as data understanding, data quality assurance, and data acceleration. Developers must ensure that the data they are working with is trustworthy and reliable, as poor data quality can lead to incorrect predictions or biased models (Zha et al., 2023, "Data-centric AI: Perspectives and Challenges").

To create ML models for high-quality data, declarative Machine Learning Frameworks (Molino & Re, 2021) are used. For instance, we can take the PetFinder dataset (Zhang & Zhang, 2019) as a case study, containing information about animals in a shelter. Figure 2 illustrates an example of a dog (type 1) named Fenny, five years old, and not adopted for 100 days (AdoptionSpeed 4).

In contrast to the model-first approach, which demands considerable knowledge and experience in machine learning algorithms and model architecture, the data-first approach requires expertise in the domain of the problem. Once the data engineering tasks are completed, describing the data is sufficient to train ML models. An description for the PetFinder dataset would entail the presence of multiple features for each data point, including an image, a numerical age value, a string name value, and categorical values indicating the animal's type and adoption speed. By providing this data description, a declarative ML tool can be used to create an ML model that is tailored to the specific problem and capable of accurately predicting the adoption speed. Overall, the data-first approach represents a significant advancement in the field of machine learning, providing a more streamlined and effective way to develop machine learning models.

Image	Type	Name	Age	AdoptionSpeed
	1	Fenny	5	4

Figure 2: Example of a pet of the PetFinder (Zhang & Zhang 2019) dataset. Fenny is a 5-year-old dog (Type 1) that has not been adopted within 100 days (AdoptionSpeed 4).

2.1 Ludwig as an example of a declarative ML framework

Declarative ML frameworks, such as Ludwig (Molino, Dudin & Miryala, 2019), Overton (Re, 2020), Fastai (Howard & Gugger 2020), and AutoGluon (Erickson et al., 2020), allow for the streamlined creation and training of machine learning models. Ludwig, for example, employs an Encoder-Combiner-Decoder model, where inputs are first processed by tailored encoders based on their data type. The encoded values are then combined using a combiner, and the decoder is used to make predictions on the target output. The specific decoder employed is dependent on the data type of the output features. This approach optimizes the model for the given dataset, resulting in precise and efficient predictions. Figure 3 shows a possible configuration for training a Ludwig model for the PetFinder dataset.

```
input_features:
- name: Image
  type: image
- name: Type
  type: category
- name: Name
  type: text
- name: Age
  type: numerical

output_features:
- name: AdoptionSpeed
  type: category

training:
  epochs: 10
  batch_size: 32
```

Figure 3: An example Ludwig configuration for the PetFinder dataset. The input and output features are listed together with some additional training information.

2.2 Limitations of the data-first AI approach

Although the data-first AI approach with its declarative nature simplifies the creation and training of ML models, a few challenges persist. First, despite reducing the required coding skills, some degree of coding is still necessary. Second, installing the framework

and setting it up to work on specialized hardware, such as GPUs, requires a certain level of ongoing technical expertise, which is typically provided as a service deliverable. Third, some data-first approaches already provide a high level of transparency by incorporating open source principles and providing additional materials, such as explanatory documentation. This inherent transparency fosters user trust, so it is critical to maintain or improve this level of openness to maintain their confidence in the system. In figure 1, we have illustrated our approach to tackle these challenges and shift from data-first to human-first AI. In the upcoming sections, we will describe this in detail.

3 Simplified Terminology for Human-first AI

In order to enhance the accessibility of AI, we consulted with professionals in technical fields who have limited or no expertise in AI but could benefit from its usage. Our initial finding was that the terminology poses a barrier. This means that technicians are unfamiliar with concepts like input and output features, preprocessing, learning rate, and batch size.

Therefore, our first step was to develop a simplified vocabulary to discuss how to address the outstanding challenges of the data-first approach. In Figure 4, we present this user-friendly language.

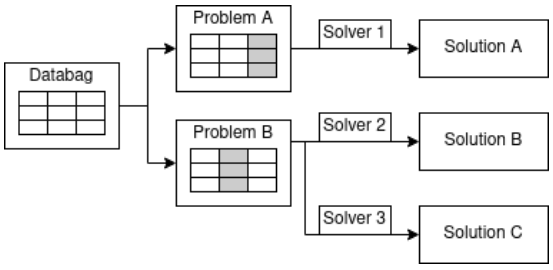


Figure 4: Simplified terminology to model AI problems: A Databag holds some multi-modal data. By specifing what you want to predict you create a Problem. A Solver is an abstraction of an ML algorithm and is used to create a Solution for the Problem.

A Databag is an abstraction of data and its associated metadata, which we suggest representing using a multi-modal data format, similar to a table or dataframe (e.g. Pandas). In this format, each column corresponds to a specific datatype, and each row represents an entry for that column. As illustrated in Figure 2, the PetFinder-Dataset is an excellent example of such a multi-modal dataset with various columns, including images, numerical, categorical, and string values. However, this concept can also be applied to represent datasets with fewer columns, such as image classification datasets, where only two columns typically exist, the image and a corresponding categorical label. The multi-modal approach offers a great abstaction to deal with very different types of data.

To define a Problem on a Databag, one or more columns can be specified for prediction. For instance, in an image classification dataset, the label column would be the output of a typical image classification task. Similarly, in the PetFinder-Dataset, AdoptionSpeed, represented as a categorical value, would be the output for a classification task. However, if AdoptionSpeed were represented as the actual days it takes for a pet to be adopted, the Problem would be a regression one. Although single-output Problems are more common, the ability to specify multiple outputs enables us to model more complex Problems, such as an image classification with two independent labels. With a textual Databag, can represent various NLP tasks, including machine translation and sentiment analysis. In essence, this straightforward abstraction enables us to model a wide range of ML problems without needing specialized terminology for each.

The Solver serves as an abstraction layer for the ML algorithm that is used to tackle a Problem. It is responsible for abstracting all the complex elements that are important for the internals of the specific ML algorithm, such as learning rate and batch size, but are not relevant for its usage. Its main function is to create and train an ML model based on the data provided in the Databag and the defined Problem.

Once the ML model is trained, the resulting data including the trained model and relevant metrics, are stored in a Solution. This Solution can then be used to predict the output columns of the Problem for unseen data.

By using this terminology, users can benefit from a user-friendly interface to utilize AI technology without requiring a deep understanding of the internal workings of the Solver. At the same time, ML experts can still maintain the necessary flexibility to design and optimize ML algorithms according to specific needs and constraints.

4 Three Pillars of Human-first AI

With this simplified terminology, we can now delve into how to transition from a data-first approach to a human-first approach in AI. Figure 1 illustrates three pillars to achieve this goal.

4.1 No-Code ML

The first pillar aims to shift from a low-code approach to a no-code approach by providing a user-friendly interface that leverages declarative ML frameworks. By doing so, users will be able to create and deploy ML models without the need to write any code.

To achieve this goal, the first step is to create a Databag. The user interface should support the uploading of data in various formats, including excel, csv, or

zip files. Once uploaded, the system should automatically inspect the data to determine the appropriate data types for each column. This information can then be stored in the Databag for later use.

To define a Problem on this Databag, the user selects the desired output columns. Based on this the system can then choose an appropriate Solver, such as a Solver based on Ludwig (Molino, Dudin & Miryala, 2019), which can be integrated seamlessly. The input and output sections for the Ludwig Solver can be derived from the multi-modal data definition of the Databag and the associated Problem.

The parameters of the training algorithms, also known as hyperparameters, can be set to appropriate default values or optimized using more expensive hyperparameter optimization strategies, depending on the available computing power. Examples of such hyperparameters include learning rate, batch size and number of hidden layers in a neural network.

Hyperparameter optimization (HPO) requires less human interaction and knowledge, but more computational power (Yu & Zhu, 2020). HPO techniques can range from simple examples such as random and grid search to more complex algorithms such as Bayesian optimization (Mockus, 1972) and Asynchronous Successive Halving Algorithm (Li et al., 2020), which have shown better performance.

This no-code ML solution enables users without coding skills or technical knowledge to benefit from AI technology, increasing its accessibility.

4.2 AI as a Service

We have demonstrated the feasibility of creating a no-code AI system using a data-first approach and simplified terminology. However, we recognize that the setup of such a system may still require expert knowledge. To address this challenge, we propose the development of a human-first AI platform that offers no-code ML as a service, with a user interface that abstracts away the underlying hardware.

To enable scalability, the platform must be deployed in the cloud. However, we understand that data privacy and security concerns may arise. To address these concerns, we suggest designing the platform in a way that allows users to host a private instance of the platform. This approach provides greater control over their data and ensures that sensitive information is not accessible to unauthorized parties. This can be achieved simply by making the deployment scripts for the platform readily available to users.

By implementing a hybrid approach that includes a public instance and an option for users to self-host the platform, we can address these two critical issues.

4.3 Transparency

In the previous section, we emphasized the importance of making the platform deployment accessible to users. However, this is just the first step towards transparency and trust in the technology. Transparency goes beyond deployment and includes making the entire platform accessible and understandable to users, independent experts, and other stakeholders. This is particularly important as artificial intelligence is often perceived as an incomprehensible and esoteric technology by non-experts.

By publicly sharing the platform's source code, we provide independent experts with the ability to scrutinize and monitor its internal workings, which enhances the platform's credibility. Furthermore, it allows users to gain a deeper understanding of the technology's mechanisms, leading to increased adoption and better outcomes. Thus, open-sourcing the platform is a fundamental step towards fostering widespread acceptance and trust in the technology.

In addition to the benefits of increased transparency and trust, open-sourcing the platform also facilitates collaboration among developers and the wider community. By making the source code available to others, developers can work together to identify and fix issues, implement new features, and enhance the overall functionality of the platform. This collaboration can result in more rapid development and deployment of new features, which in turn can lead to a more robust and adaptable platform. Therefore, open-sourcing the platform not only fosters trust and understanding, but also encourages innovation and community-driven improvement, making it a vital step towards creating a truly successful and impactful technology.

In the quest to increase trust and transparency in the technology, open-sourcing the platform is an important first step. In addition, utilizing explainable AI represents another promising approach. By leveraging explainable AI, the internal workings of ML models become more understandable to users, resulting in a greater level of transparency. This, in turn, can increase user confidence in the technology and encourage adoption. Therefore, incorporating explainable AI into the platform's development improves the overall trustworthiness and user-friendliness of the technology.

5 Comparison of existing AutoML Tools

In table 1 you can see a comparison of existing AutoML tools (Calefato et al., 2023). We have updated the entry for Ludwig (XAI features are already available in the codebase) and added additional rows for Orange Data Mining, KNIME Analytics Platform and Os4ML.

Automated machine learning (AutoML) aims to streamline the entire ML pipeline, encompassing tasks

Tool	Infrastruct.	Solution	Data types	Preparation		Analysis			Dissemination		XAI
				Data clean.	Data label.	Feat. engin.	Model train.	Model eval.	Model deploy.	Model monitor.	
AutoGluon*	On prem.	API	img, tab, txt	Y	N	Y	Y	Y	N	N	N
AutoGoal*	On prem.	API, CLI	img, tab, txt	N?	N	Y?	Y	Y	N	N	N
AutoKeras*	On prem.	API	img, tab, txt	Y?	N	Y?	Y	Y	N	N	N
Amazon SageMaker AutoPilot	Cloud	API, Web	img, tab, txt	Y	Y	Y	Y	Y	Y	Y	Y
BigML	Cloud	API, Web	img, tab, txt	Y	N	Y	Y	Y	Y	Y	Y
DataRobot AI Cloud	Cloud	API, Web	img, tab, txt	Y	Y	Y	Y	Y	Y	Y	Y
Google Vertex AI	Cloud	API, Web	img, tab, txt	Y	Y	Y	Y	Y	Y	Y	Y
H20 Driverless AI	Cloud, on prem.	API, Web	img, tab, txt	Y	Y	Y	Y	Y	Y	Y	Y
IBM Watson AutoAI	Cloud	API, Web	img, tab, txt	Y	N	Y	Y	Y	Y	Y	Y
Ludwig AI AutoML*	On prem.	API, CLI	img, tab, txt	Y?	N	Y	Y	Y	Y	N	Y?
MS Azure AutoML	Cloud	API, Web	img, tab, txt	Y	Y	Y	Y	Y	Y	Y	Y
Rapid Miner Studio	On prem.	Desktop	img, tab, txt	Y	N	Y	Y	Y	Y	Y	N
Orange Data Mining*	On prem.	API, Desktop	img, tab, txt	Y	N	Y	Y	Y	N	N	Y
KNIME Analytics Platform*	On prem.	Desktop	img, tab, txt	Y	N	Y	Y	Y	N	N	Y
Os4ML*	Cloud, on prem.	API, Web	img, tab, txt	Y	(Y)	Y	Y	Y	(Y)	(Y)	(Y)

Table 1: Table is taken from (Calefato et al., 2023), the Ludwig entry is updated and the Orange Data Mining, KNIME Analytics Platform and Os4ML entries are added. All open source tools are marked with a * behind the name in the first column. Y? means that the feature was present in the code, but not in the documentation. N? means that is was present in the documentation, but not in the code. Features, that are currently under development, are marked with (Y).

like data preparation, data analysis, and model generation (He et al., 2021). However, unlike the human-first AI approach, its primary focus is not merely on reducing the technical expertise required to use these tools. Contrary to the notion that "automated" implies a no-code solution like the introduced human-first AI approach, AutoML tools can still demand considerable technical expertise to set up and operate. Take Ludwig for example, as mentioned in section 2.1, it simplifies certain aspects of the ML pipeline, but users still need significant technical knowledge to utilize its API or CLI interface effectively.

Therefore, it is essential to recognize that human-first AI encompasses more than just AutoML. Besides automating the ML pipeline, human-first AI is dedicated to making the entire process accessible through a simple, no-code interface that can be accessed as a service.

To our knowledge, Os4ML is currently the only tool that provides a no-code web interface, supports both cloud and on-premise infrastructures, and remains open source. This unique combination makes Os4ML an outstanding solution in the field of human-first AI. In the following section we will look at the details of Os4ML.

6 The Open space for Machine Learning Platform

We are excited to provide an update on the progress of the Open space for Machine Learning (Os4ML) platform, which represents our commitment to human-first AI. We have achieved significant milestones in the development of the platform, including the successful implementation of a frontend with our proposed terminology and the integration of a Ludwig Solver, as outlined in section 4.1.

To illustrate, let's consider the PetFinder Dataset. Creating a Databag for this dataset is a seamless process. All you need to do is upload your data in one of our supported formats, such as an excel file with tabular data and a zip file containing images, onto our system. Our platform automatically detects the datatype of each column, although you can adjust the selection if necessary. Once you've selected the output column, simply choose a Solver (currently, only the Ludwig Solver is available, but we plan to add more options in the future) and let the platform train your ML model.

Once the training process is finished, you have the freedom to either download the model in the onnx format ("Open Neural Network Exchange", 2023) or directly upload new, unseen data for prediction. As we look ahead, we are actively enhancing the platform's capabilities to allow you to serve the model and grant straightforward access to it via API. This feature will enable you to seamlessly incorporate the model into your current systems, unlocking its full potential for your business or organization.

Additionally, we are proud to offer our platform to the public through a public instance[1] than can handle the high scalability demands of a cutting-edge AI platform. Moreover, in line with our commitment to transparency and openness, we have made the platform's deployment scripts publicly available along with the complete codebase of the platform[2]. These efforts reflect our dedication to create a trustworthy and accessible platform, and we look forward to continue to build on this initial progress to achieve our goals.

Although the integration of explainable AI into the human-first AI approach holds great promise for increasing transparency and trust in the technology, it is still an area that requires further research and development. We acknowledge that there is still much work to be done in this regard, and we remain committed to explore this avenue in future efforts. We will continue to report on our progress as we work towards incorporating explainable AI into our platform, with the ultimate goal of creating a technology that is both trustworthy and user-friendly.

[1] https://www.os4ml.com/

[2] https://github.com/WOGRA-AG/Os4ML

7 A Vision for the Future: Chatbot Interface

The emergence of large language models and chat systems (Fan et al., 2023), such as the GPT models (Eloundou et al., 2023), has opened up exciting possibilities for user interface design. We propose a vision for leveraging such systems as an interface to the Os4ML platform. However, it is important to acknowledge that this is a rapidly evolving field and the ideas presented here are intended to spark further discussion and exploration rather than to offer a definitive solution.

By leveraging a chat system, such as a standalone Os4ML chatbot or a plugin for an existing system like ChatGPT ("Chat Plugins", 2023), users can easily describe their data in natural language. The system would then automatically convert the data into a format compatible with Os4ML. If the user has not specified outputs, they would be prompted to do so. Once the Problem is defined, a Solver could be utilized to generate a Solution. The final model would be presented to the user in the same way as with traditional user interfaces, providing options to download the model, access it through API, or directly upload new data for evaluation.

A good starting point for this can be the GPT4All project, which provides a chatbot that is privacy aware and uses minimal computing ressources (Anand et al., 2023).

Utilizing a chat system as the user interface for the Os4ML platform offers several advantages over a traditional interface. Firstly, users can input information using natural language through various methods such as typing or speech-to-text tools like whisper (Radford et al., 2022), making the platform more accessible to users with varying preferences and abilities. Secondly, the chat system can automatically format data to the Os4ML standard, relieving users from the need to learn how to format data themselves.

Moreover, the simplicity of natural language makes it one of the easiest user interfaces, ensuring that the platform is user-friendly and easy to use for everyone. An additional benefit of using a chat system is that users can receive immediate answers to their questions, eliminating the need to search through documentation and enabling them to obtain the information they need quickly and efficiently.

In summary, the chat system streamlines the interface, making it more flexible and accessible for all users, while also simplifying the process of data formatting and enabling users to receive real-time support.

8 Summary & Outlook

In this article, we have demonstrated how the principles of data-first AI can be leveraged to create human-first AI. Our approach involved building a no-code ML solution around the data-first AI, offering it as a service, and making the codebase open source to promote transparency and collaboration. The Open space for Machine Learning platform represents a practical example of our approach in action.

As we look ahead, our focus remains steadfast on the continued development of our platform, with the goal of making AI more accessible to a broader range of users. We are thrilled to witness how users utilize our platform and how this influences their work. Through our efforts, we strive to empower individuals with the tools they need to realize their full potential and achieve their goals.

Acknowledgments

Os4ML is a project of the WOGRA AG research group in cooperation with the German Aerospace Center and is funded by the Ministry of Economic Affairs, Regional Development and Energy as part of the High Tech Agenda of the Free State of Bavaria.

References

Allen, B., Agarwal, S., Kalpathy-Cramer, J. & Dreyer, K. (2019) Democratizing AI. *Journal Of The American College Of Radiology*, 961-963

Anand, Y., Nussbaum, Z., Duderstadt, B., Schmidt, B. & Mulyar, A. (2023) *GPT4All: Training an Assistant-style Chatbot with Large Scale Data Distillation from GPT-3.5-Turbo.* https://github.com/nomic-ai/gpt4all

Cabot, J. (2020) Positioning of the low-code movement within the field of model-driven engineering. *Proceedings Of The 23rd ACM/IEEE International Conference On Model Driven Engineering Languages And Systems: Companion Proceedings*, https://doi.org/10.1145/3417990.3420210

Calefato, F., Quaranta, L., Lanubile, F. & Kalinowski, M. (2023) *Assessing the Use of AutoML for Data-Driven Software Engineering.*

Chat Plugins (2023). Retrieved from https://platform.openai.com/docs/plugins/

Eloundou, T., Manning, S., Mishkin, P. & Rock, D. *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.* (2023)

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M. & Smola, A. *AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data.* (2020)

Fan, L., Li, L., Ma, Z., Lee, S., Yu, H. & Hemphill, L. A *Bibliometric Review of Large Language Models Research from 2017 to 2023.* (2023)

He, X., Zhao, K. & Chu, X. (2021) *AutoML: A survey of the state-of-the-art*. *Knowledge-Based Systems* 212 pp. 106622, https://doi.org/10.1016%252Fj.knosys.2020.106622

Howard, J. & Gugger, S. (2020) Fastai: A Layered API for Deep Learning. *Information*, 11, 108, https://doi.org/10.3390

Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Hardt, M., Recht, B. & Talwalkar, A. *A System for Massively Parallel Hyperparameter Tuning.* (2020)

Mockus, J. (1972) On Bayes methods for seeking an extremum. *Avtomatika I Vychislitelnaja Technika.*, 3, pp. 53-62

Molino, P., Dudin, Y. & Miryala, S. *Ludwig: a type-based declarative deep learning toolbox.* (2019)

Molino, P. & Re, C. *Declarative Machine Learning Systems.* (2021)

Open Neural Network Exchange (2023). Retrieved from https://onnx.ai/

Radford, A., Kim, J., Xu, T., Brockman, G., McLeavey, C. & Sutskever, I. *Robust Speech Recognition via Large-Scale Weak Supervision.* (2022)

Re, C. (2020) Overton: A Data System for Monitoring and Improving Machine-Learned Products. *10th Conference On Innovative Data Systems Research*, CIDR 2020, Amsterdam, The Netherlands, January 12-15, 2020, Online Proceedings., http://cidrdb.org/cidr2020/papers/p33-re-cidr20.pdf

Shi, X., Mueller, J., Erickson, N., Li, M. & Smola, A. *Benchmarking Multimodal AutoML for Tabular Data with Text Fields.* (2021)

The AI Index Report 2023 (2023). Retrieved from https://aiindex.stanford.edu/report/

Whang, S., Roh, Y., Song, H. & Lee, J. (2023) Data collection and quality challenges in deep learning: a data-centric AI perspective. *The VLDB Journal*, https://doi.org/10.1007/s00778-022-00775-9

Yu, T. & Zhu, H. *Hyper-Parameter Optimization: A Review of Algorithms and Applications.* (2020)

Zha, D., Bhat, Z., Lai, K., Yang, F. & Hu, X. *Data-centric AI: Perspectives and Challenges.* (2023)

Zha, D., Bhat, Z., Lai, K., Yang, F., Jiang, Z., Zhong, S. & Hu, X. *Data-centric Artificial Intelligence: A Survey.* (2023)

Zhang, S. & Zhang, K. *PetFinder Challenge: Predicting Pet Adoption Speed.* (2019)

Annotation and Machine Identification of Metaphors in Croatian Newspaper Articles

Anotiranje i strojno identificiranje metafore u hrvatskim novinskim člancima

Martina Ptiček, Jasminka Dobša
University of Zagreb
Sveučilište u Zagrebu
Faculty of Organization and Informatics
Fakultet organizacije i informatike
Pavlinska ul. 2, 42000, Varaždin
marpticek@student.foi.hr, jasminka.dobsa@foi.hr

Abstract. *A metaphor is an integral part of communication, but it also affects the way we think and understand the world. Identifying metaphors by using machine learning methods is a challenging task, and the first step in machine identification is the creation of annotated datasets. As part of the paper, an annotated dataset of metaphors in Croatian newspaper articles was created, taken from two newspaper portals - Index.hr and Jutarnji list. The Metaphor Identification Procedure Vrije Universiteit (MIPVU) procedure was adapted for these needs. Metaphor identification through machine learning was approached as a task of classifying text sequences, whereby large language models were used, which were trained on texts in the Croatian language as well. The paper presents the challenges and observations of annotating a metaphor. The identification model achieved an F1 score of 70.75%, and a difference in identification success was observed depending on the data source. The paper provides a comparison of classification accuracy depending on the size of the dataset, as well as guidelines for further research.*

Keywords. Metaphor, metaphor annotation, metaphor identification, large language models

1 Introduction

Since Lakoff and Johnson (1980) presented the theory of the conceptual metaphor, it has become the object of study in many sciences – from psychology, philosophy and linguistics to computational

Sažetak. *Metafora je sastavni dio komunikacije no ona utječe i na način kako razmišljamo i poimamo svijet. Identificiranje metafore metodama strojnog učenja izazovan je zadatak, a prvi korak pri strojnom identificiranju je izrada anotiranih skupova podataka. U sklopu rada izrađen je anotirani skup podataka metafore u hrvatskim novinskim člancima, preuzetim s dva novinska portala -Index.hr i Jutarnji list. Za te potrebe prilagođena je MIPVU (akronim od engl. Metaphor Identification Procedure Vrije Universiteit) procedura. Identificiranju metafore strojnim učenjem pristupilo se kao zadatku klasificiranja sekvenci teksta, pri čemu su korišteni veliki jezični modeli, trenirani i na tekstovima na hrvatskom jeziku. U radu se navode izazovi i opažanja pri anotiranju metafore. Modelom identificiranja ostvaren je F1 rezultat od 70,75% te je uočena razlika u uspješnosti identificiranja ovisno o izvoru podataka. U radu je dana usporedba točnosti klasifikacije ovisno o veličini skupa podataka, kao i smjernice za daljnja istraživanja.*

Ključne riječi. Metafora, anotiranje metafore, identificiranje metafore, veliki jezični modeli

1 Uvod

Od kada su Lakoff i Johnson (1980) predstavili teoriju konceptualne metafore, ona je postala objekt proučavanja u mnogim znanostima – od psihologije, filozofije i lingvistike, pa do računalne lingvistike i računalnih znanosti (usp. Steen et al., 2010).

Metafora je sastavni i važan dio ljudske komunikacije. Lakoff i Johnson svojom su teorijom

linguistics and computer science (cf. Steen et al., 2010).

The metaphor is an integral and important part of human communication. Lakoff and Johnson demonstrated with their theory that a metaphor does not remain only in language but affects the way we think and understand the world around us. Lakoff and Johnson teach us about metaphorical concepts like ARGUMENT IS WAR (for instance, "Your claims are *indefensible*.", "*He attacked the weak points* of my argument."), where we map the more concrete, physical domain of war onto the more abstract domain of discussion, which is why we understand the discussion as a war in which someone wins and someone loses.

Lakoff and Johnson do not deal with the issue of the linguistic metaphor, but a metaphor, in addition to being conceptual, is certainly also linguistic because it is expressed by language and in language. Despite not discussing the linguistic metaphor themselves, their theory of the conceptual metaphor has also influenced how we identify the linguistic metaphor.

For instance, Steen et al. (2010), based on the MIP (*Metaphor Identification Procedure*) (Pragglejaz Group, 2007) developed the MIPVU (*Metaphor Identification Procedure Vrije Universiteit*) of tagging linguistic metaphors, which approaches metaphor identification precisely from the perspective of the conceptual metaphor (Krennmayr and Steen, 2017). Steen et al. use the MIPVU procedure to denote texts from the four registers of the BNC Baby corpus[1] – conversation, fiction, academic texts and newspaper articles. They conclude that the metaphor appears most often in academic texts (18.5%) and newspaper articles (16.4%), followed by fiction (11.9%) (cf. Krennmayr and Steen, 2017). Shutova (2017) also states that the metaphor is most common in newspaper articles, politics and a journals about language and literature.

It is precisely because of the conclusions of the previous papers that, as part of this paper, the metaphor was annotated in newspaper articles using the MIPVU procedure, adapted to the Croatian language. Also, as part of the paper, a deep learning model was developed for metaphor identification, where the problem of identification was approached as a problem of the classification of text sequences. When creating the model, contextual word embeddings, that is, large language models were used, which were trained on texts in the Croatian language, among other languages. The language models used are XLM-RoBERTa (Conneau et al., 2020), CroSloEngual (Ulčar & Robnik-Šikonja, 2020) and BERTić (Ljubešić & Lauc, 2021), and the results obtained by each of these models are compared.

pokazali kako metafora ne ostaje samo u jeziku već utječe na način kako razmišljamo i poimamo svijet oko sebe. Lakoff i Johnson uče nas o metaforičkim konceptima poput RASPRAVA JE RAT (npr. „Tvrdnje su ti *neobranjive*.“, „*Napao je slabe točke* moje argumentacije.“) gdje konkretniju, fizičku domenu rata preslikavamo na apstraktniju domenu rasprave pa tako i raspravu poimamo kao rat u kojem netko pobjeđuje a netko gubi.

Lakoff i Johnson ne bave se pitanjem lingvističke metafore, no metafora je, osim konceptualna, svakako i lingvistička jer se izražava jezikom i u jeziku. Unatoč tome što sami ne govore o lingvističkoj metafori, njihova je teorija konceptualne metafore utjecala i na to kako identificiramo lingvističku metaforu.

Tako Steen et al. (2010), na temelju MIP (akronim od engl. *Metaphor Identification Procedure*) (Pragglejaz Group, 2007) razvijaju MIPVU (akronim od engl. *Metaphor Identification Procedure Vrije Universiteit*) proceduru označavanja lingvističke metafore, koja identificiranju metafore pristupa upravo iz rakursa konceptualne metafore (Krennmayr i Steen, 2017). MIPVU procedurom Steen et al. označavaju tekstove iz četiri registra BNC Baby korpusa[1] – konverzacija, fikcija, akademski tekstovi i novinski članci. Zaključuju kako se metafora najčešće pojavljuje u akademskim tekstovima (18,5%) i novinskim člancima (16,4%) te nakon toga u fikciji (11,9%) (usp Krennmayr i Steen, 2017). Shutova (2017) također navodi kako je metafora najčešća u novinskim člancima, politici i časopisu o jeziku i književnosti.

Upravo zbog zaključaka prethodnih radova, u sklopu ovog rada metafora je anotirana u novinskim člancima i to korištenjem MIPVU procedure, prilagođene hrvatskom jeziku. Također, u sklopu rada izrađen je model dubokog učenja za identificiranje metafore, gdje se problemu identificiranja pristupilo kao problemu klasifikacije sekvenci teksta. Pri izradi modela korišteni su kontekstualni vektorski prikazi riječi (engl. *contextual word embeddings*), odnosno veliki jezični modeli, koji su, između ostalih jezika, trenirani i na tekstovima na hrvatskom jeziku. Jezični modeli koji se koriste su XLM-RoBERTa (Conneau et al., 2020), CroSloEngual (Ulčar & Robnik-Šikonja, 2020) i BERTić (Ljubešić & Lauc, 2021), te se uspoređuju rezultati dobiveni svakim od navedenih modela.

Prema našem saznanju, ovo je prvi rad koji koristi MIPVU proceduru te izrađuje anotirani skup podataka metafore u hrvatskim novinskim člancima, kao i prvi rad koji koristi velike jezične modele u identificiranju metafore u hrvatskim novinskim člancima.

[1] The corpus is available at http://www.natcorp.ox.ac.uk/corpus/babyinfo.html
[1] Korpus je dostupan na poveznici http://www.natcorp.ox.ac.uk/corpus/babyinfo.html

To our knowledge, this is the first paper that uses the MIPVU procedure and creates an annotated dataset of metaphors in Croatian newspaper articles, as well as the first paper that uses large language models to identify metaphors in Croatian newspaper articles.

The rest of the paper is organized as follows - in the section on related work, an overview of relevant metaphor annotating procedures and papers that approach metaphor identification using large language models is provided. An overview of the collected data and the annotation procedure is provided in the *Data* section, while the *Classification Model Settings* chapter provides an overview of the metaphor identification system using large language models. The results of the research are presented in the *Results* section, and the conclusion provides a summary of the research, as well as suggestions and guidelines for further research.

2 Related work

The first step in metaphor identification using machine learning methods is the creation of annotated datasets. Below is a brief overview of previous procedures in metaphor annotation that are relevant for this paper, as well as metaphor identification methods using neural networks and large language models.

2.1 Annotating Metaphors

For annotating linguistic metaphors, the currently most accepted procedures are MIP (Pragglejaz Group, 2007) and MIPVU (Steen et al., 2010), which is an upgrade of MIP. Both methods tag the metaphor on a single lexical unit, which most often coincides with a single word, but sometimes compounds and phrases also represent a lexical unit.

MIP has three steps in which it determines whether a word is used metaphorically and does not distinguish between types of metaphors. MIPVU, on the other hand, introduces a distinction between "regular" ("*MRW*"[2]), direct ("*MRW, direct*") and implicit ("*MRW, implicit*") metaphors. However, as we find in Krennmayr and Steen (2017), it was shown that direct and implicit metaphors are represented in an extremely small percentage of 0.2% of the entire corpus.

The MIPVU procedure was used to annotate the well-known VUA[3] dataset, which is used in a large number (if not in all) of papers on metaphor identification in the English language, and 186,695 words were annotated in it.

Ostatak rada organiziran je na sljedeći način - u poglavlju povezani radovi dan je pregled relevantnih procedura označavanja metafore te radova koji identificiranju metafore pristupaju korištenjem velikih jezičnih modela. Pregled prikupljenih podataka i procedure anotiranja dan je u poglavlju *Podaci*, dok je u poglavlju *Postavke modela za klasifikaciju* dan pregled sustava identificiranja metafore korištenjem velikih jezičnih modela. Rezultati istraživanja prikazani su u poglavlju *Rezultati* te je u zaključku dan sažetak istraživanja kao i prijedlozi i smjernice za daljnje istraživanje.

2 Povezani radovi

Prvi korak u identificiranju metafore metodama strojnog učenja izrada je anotiranih skupova podataka. U nastavku je dan kratki pregled za ovaj rad relevantnih prethodnih procedura u anotiranja metafore, kao i metoda identificiranja metafore korištenjem neuronskih mreža i velikih jezičnih modela.

2.1 Anotiranje metafore

Za anotiranje lingvističke metafore, trenutno najprihvaćenije procedure su MIP (Pragglejaz Group, 2007) i MIPVU (Steen et al., 2010), koja je nadogradnja na MIP. Obje metode metaforu označavaju na pojedinoj leksičkoj jedinici, što se najčešće poklapa s jednom riječi, no ponekad i složenice i fraze predstavljaju leksičku jedinicu.

MIP ima tri koraka u kojima određuje koristi li se riječ metaforički te ne razlikuje vrste metafore. MIPVU pak uvodi razlikovanje između „obične" metafore (oznaka „*MRW*"[2]), direktne (oznaka „*MRW, direct*") i implicitne (oznaka „*MRW, implicit*") metafore. No, kao što nalazimo kod (Krennmayr i Steen (2017), pokazalo se da su direktna i implicitna metafora zastupljene u izrazito malom postotku od 0.2% cijelog korpusa.

MIPVU procedurom označen je dobro poznati VUA[3] skup podataka koji se koristi u velikom broju radova (ako ne i u svim) identificiranja metafore u engleskom jeziku, te je u njemu anotirano 186.695 riječi.

2.2 Identificiranje metafore velikim jezičnim modelima

Prema našim saznanjima, prvi rad koji koristi kontekstualni vektorski prikaz riječi je Gao et al. (2018), gdje autori/ce koriste ELMo (Peters et al., 2018). Identificiranju metafore u radu se pristupa

[2] MRW is an acronym for *Metaphor Related Word*
[3] The dataset is available at http://www.vismet.org/metcor/documentation/home.html
[2] MRW je akronim od engl. *Metaphor Related Word*
[3] Skup podataka dostupan je na poveznici http://www.vismet.org/metcor/documentation/home.html

2.2 Metaphor identification by using large language models

To the best of our knowledge, the first paper using contextual word embeddings is Gao et al. (2018), in which the authors use ELMo (Peters et al., 2018). Metaphor identification in the paper is approached as tasks of (1) Sequence labeling, where each word is tagged either as a metaphor or as a literal use; (2) classification by which only the verb in the sentence is tagged as a metaphor or literal use. In both cases, the BiLSTM network is used, and a layer of the attention mechanism was added to the classification model. Using ELMo when tagging text sequences results in an F1 of 70.4% versus an F1 of 61.7% without ELMo.

Mao et al. (2019) base their paper on the hypothesis that the use of linguistic theories of the metaphor - MIP and selectional preferences (Katz & Fodor, 1963; Wilks, 1975, 1978) - will improve the success of identifying metaphors. They use GloVe (Pennington et al., 2014) and ELMo and create two architectures, where MIP is implemented in the first and selectional preferences in the second. The MIP-based model approaches metaphor classification based on the difference between the literal meaning of the word and the contextual one. For the representation of the contextual meaning, the model uses BiLSTM, while for the representation of the literal meaning, it uses GloVe and ELMo, and in this architecture the model achieves an F1 score of 74.0% for the entire VUA data set, and 70.8% for VUA verbs.

Su et al. (2020) approach metaphor identification as a reading comprehension problem and use BERT in their paper (Devlin et al., 2019) embedding layer, while they use a transformers neural network as the basis of their model (Vaswani et al., 2017). In the training process, they use the RoBERTA (Liu et al., 2019) word embeddings, with the aim of fine-tuning the transformers layers. They achieved an F1 score of 80.4% for VUA verbs and 76.9% for all words in the VUA dataset.

Gong et al. (2020) also use the RoBERTA model and approach identification as a task of tagging text sequences. The model is fed linguistic properties (POS, distribution of words in topics, concreteness of words, WordNet, VerbNet and properties from the corpus) while a Feed forward network is used for classification. The best F1 score of 77.1% was achieved for VUA verbs.

Chen et al. (2020), using the BERT fine-tuning method, performed three different experiments – correcting spelling errors in the TOEFL data set, using data from another domain in the training phase, and learning from another type of figurative language. In learning from another type of figurative language, they compose sets of phrases, with the idea that phrases are often metaphors, and in this experiment they achieved the best F1 score of 77.5% kao zadacima (1) označavanja sekvenci teksta (engl. *Sequence labeling*), gdje se svaka riječ označava ili kao metafora ili kao doslovno korištenje; (2) klasifikacije kojom se samo glagol u rečenici označava kao metafora ili doslovno korištenje. U oba slučaja koristi se BiLSTM mreža te je u klasifikacijskom modelu dodan sloj mehanizma pažnje (engl. *Attention*). Korištenjem ELMo-a pri označavanja sekvenci teksta ostvaruje se F1 od 70,4% naspram F1 61,7% bez ELMo-a.

Mao et al. (2019) svoj rad temelje na hipotezi kako će korištenje lingvističkih teorija metafore - MIP i selekcijskih preferenci (Katz & Fodor, 1963; Wilks, 1975, 1978) - poboljšati uspješnost identificiranja metafore, te koriste GloVe (Pennington et al., 2014) i ELMo. Izrađuju dvije arhitekture, gdje je u prvoj implementiran MIP a u drugoj selekcijske preference. Model temeljen na MIP-u pristupa klasificiranju metafore na osnovi razlike između doslovnog značenja riječi i kontekstualnog. Za reprezentaciju kontekstualnog značenja model koriste BiLSTM dok za reprezentaciju doslovnog značenja koriste GloVe i ELMo te u toj arhitekturi ostvaruju F1 vrijednost 74,0% za cijeli VUA skup podataka, odnosno 70,8% za VUA glagole.

Su et al. (2020) identificiranju metafore pristupaju kao problemu čitanja s razumijevanjem te u radu koriste BERT (Devlin et al., 2019) u sloju za ugradnju, dok kao osnovu svog modela koriste transformers neuronsku mrežu (Vaswani et al., 2017). U procesu treniranja koriste RoBERTa (Liu et al., 2019) vektorski prikaz riječi, s ciljem finog podešavanja transformers slojeva. Ostvaruju F1 mjeru 80,4% za VUA glagole i 76,9% za sve riječi u VUA skupu podataka.

Gong et al. (2020) također koriste RoBERTa model te identificiranju pristupaju kao zadatku označavanja sekvenci teksta. Modelu predaju lingvistička svojstva (POS, distribucija riječi u temama, konkretnost riječ, WordNet, VerbNet te svojstva iz korpusa) dok se za klasifikaciju koristi aciklička mreža (engl. *Feed forward network*). Najbolji F1 rezultat od 77,1% ostvaren je za VUA glagole.

Chen et al. (2020) u svom radu metodom finog podešavanja BERT-a izvršavaju tri različita eksperimenta – ispravljanje grešaka u pravopisu TOEFL skupa podataka, korištenje podataka iz druge domenu u fazi treniranja te učenje iz druge vrste figurativnog jezika. U učenju iz druge vrste figurativnog jezika sastavljaju skupove fraza, s idejom kako su fraze često metafore te u tom eksperimentu ostvaruju najbolje rezultate od 77,5% za VUA glagole i 73,4% za VUA sve vrste riječi.

Dankers et al.(2020) predlažu arhitekturu koja inkorporira širi diskurs, tako što koriste prozor konteksta od $2k + 1$ rečenica prije i poslije rečenice koja je u fokusu identificiranja metafore. U radu eksperimentiraju s dvije različite arhitekture – prva u

for VUA verbs and 73.4% for VUA all types of words.

Dankers et al. (2020) propose an architecture that incorporates the wider discourse, by using a context window of $2k + 1$ sentences before and after the sentence that is the focus of identifying the metaphor. They experimented with two different architectures – in the first, GloVe and ELMo are in the embedding layer, and BiLSTM in the encoding layer, and in the second the BERT model is in the embedding and the encoding layer. The best F1 score of 71.5% was achieved on the architecture with BERT and hierarchical attention, on the VUA dataset.

Lin et al. (2021) approach metaphor identification with the Contrastive Learning approach designed in order for the model to learn the difference between the literal and metaphorical use of a word, based on their distance in the vector space representation. It uses the RoBERTA model for the encoder and contextual display of words, and achieves an F1 score of 79% for VUA all types of words and 75.6% for VUA verbs.

3 Data

Since previous papers on metaphor identification have shown that a large percentage of metaphors are represented in newspaper articles (Krennmayr and Steen, 2017; Shutova, 2017), for the purposes of this paper, articles from two popular Croatian news portals, Index.hr and Jutarnji list, were collected. [4]

The portals were searched by the keywords *migranti* (engl. migrants), *premijer* (engl. prime minister), *predsjednik* (engl. president), *vlada* (engl. government), *Europska unija* (engl. European Union), *korona* (engl. corona). Furthermore, the criterion was that the articles were published between 1/1/2010 and 12/31/2021, while the articles themselves were chosen randomly, with 30 articles were downloaded from each newspaper portal.

In the preparation of texts for tagging, the articles were tokenized to the sentence level, using the NLTK library (Bird et al., 2009), and 1,265 sentences from Index.hr and 742 sentences from Jutarnji list were obtained, i.e. a total of 2,007 sentences.

In order to determine whether there is a difference in the classification result depending on the size of the set, assuming that there is, a comparison of the accuracy of the model was made as part of the paper, depending on the size of the dataset, and the results of the comparison are presented in the *Results* chapter.

kojoj je u sloju ugradnje GloVe i ELMo, te u sloju enkodiranja BiLSTM, te druga u kojoj je u sloju ugradnje i enkodiranja BERT model. Najbolji F1 rezultat od 71,5%, ostvaren je na arhitekturi s BERT-om i hijerarhijskom pažnjom, na VUA skupu podataka.

Lin et al. (2021) identificiranju metafore pristupaju kontrastnim učenjem (engl. *Contrastive learning*), kako bi model naučio razliku između doslovnog i metaforičkog korištenja neke riječi, na temelju njihove udaljenosti u prikazu u vektorskom prostoru. Za enkoder i kontekstualni prikaz riječi koristi RoBERTa model, te ostvaruju rezultat od F1 79% za VUA sve vrste riječi te 75,6% za VUA glagole.

3 Podaci

Budući da su prethodni radovi identificiranja metafore pokazali kako je metafora u velikom postotku zastupljena u novinskim člancima (Krennmayr i Steen, 2017; Shutova, 2017) za potrebe ovog rada prikupljeni su članci s dva popularna hrvatska novinska portala, Index.hr i Jutarnji list. [4]

Portali su pretraživani po ključnim riječima migranti, premijer, predsjednik, vlada, Europska unija, korona. Nadalje je kriterij bio da su članci objavljeni između 1.1.2010. i 31.12.2021., dok su sami članci birani nasumičnim odabirom te je sa svakog novinskog portala preuzeto 30 članka.

U pripremi tekstova za označavanje, članci su tokenizirani na razinu rečenice, pri čemu je korištena NLTK biblioteka (Bird et al., 2009), te je dobiveno 1.265 rečenica s Index.hr i 742 rečenice s Jutarnjeg lista, odnosno ukupno 2.007 rečenica.

Kako bi se utvrdilo postoji li razlika u rezultatu klasifikacije ovisno o veličini skupa, uz pretpostavku kako postoji, u sklopu rada napravljena je i usporedba točnosti modela, ovisno o veličini skupa podataka te su rezultati usporedbe prikazani u poglavlju *Rezultati*.

3.1 Procedura označavanja metafore

Kao osnova za anotiranje metafore uzeta je MIPVU procedura, koja je nadograđena i prilagođena za hrvatski jezik. Naime, MIPVU metoda je primarno rađena za engleski jezik, te samim time ne uzima u obzir specifičnosti drugih jezika (usp. Bogetić et al., 2019; Nacey et al., 2019).

Bogetić et al. (2019) primjenjuju MIPVU proceduru na srpski jezik te uočavaju kako su potrebne prilagodbe za anotiranje metafore u srpskom jeziku. Te su prilagodbe, zbog strukture jezika, primjenjive i na hrvatski jezik. Tako autorice

[4] According to Similarweb (accessed on 02/23/2022), Index.hr is the third and Jutarnji list the fifth most visited portal in Croatia.
[4] Prema Similarweb (pristupljeno 23.02.2022.), Index.hr je treći a Jutarnji list peti najposjećeniji portal u Hrvatskoj.

3.1 Metaphor Annotation Procedure

The MIPVU procedure, which was upgraded and adapted for the Croatian language, was used as the basis for annotating the metaphors. Namely, the MIPVU method was primarily developed for the English language, and therefore does not take into account the specificities of other languages (cf. Bogetić et al., 2019; Nacey et al., 2019).

Bogetić et al. (2019) apply the MIPVU procedure to the Serbian language and observe that adjustments are needed for annotating metaphors in the Serbian language. Due to the structure of the language, these adjustments are also applicable to the Croatian language. For instance, the authors state that when determining a lexical unit, (1) reflexive verbs (e.g. "smijem se" in Croatian and Serbian, meaning "I'm laughing"), (2) common expressions such as "u stvari" ("in fact"), "u redu" ("alright"), "bez obzira" ("regardless") and (3) compounds like "srednja škola" ("secondary school") should be considered one lexical unit, while (4) a preposition followed by a negative pronoun is divided into two units (e.g. "ni s kim" ("with no one") is divided into "ni s" and „kim").

In addition to determining the lexical unit, Bogetić et al. particularly draw attention to the metaphor expressed in oblique cases, which is a specificity of the Serbian but also the Croatian language, and for the purposes of this paper the MIPVU method has been expanded with instructions for tagging the metaphor expressed in oblique cases:

- For each word that is inflected with an oblique case and stands alone, without a preposition, determine which case it is and what the meaning is in the context.
- Determine whether the meaning of the case in the context is different from the basic meaning, but can be understood in comparison with it. If yes, mark the word as a metaphor expressed by case (*MRW, inflectional*)

An important step in the MIPVU procedure is determining whether a word that could potentially be a metaphor has a more basic, simpler meaning. When determining this, it is necessary to consult a monolingual dictionary, for which the *Veliki rječnik hrvatskoga standardnog jezika* (Great Dictionary of the Croatian Standard Language) was chosen (Jojić and Nakić, 2015). If, after checking this dictionary, there is uncertainty as to whether there is a more basic meaning, an additional check was made on the *Hrvatski jezični portal* (Croatian Language Portal). [5]

In addition to the MIPVU procedure and its adaptation, the annotators were instructed to try to determine the conventionality of the metaphor, on a scale from 1 to 5, where 1 is a completely new and 5 navode, a što se preuzima za potrebe ovog rada, kako je pri određivanju leksičke jedinice potrebno (1) povratne glagole (npr. „smijem se"), (2) uobičajene izraze poput „u stvari", „u redu", „bez obzira" i (3) složenice poput „srednja škola" smatrati jednom leksičkom jedinicom, dok se (4) prijedlog praćen negativnom zamjenicom dijeli na dvije jedinice (npr. „ni s kim" se dijeli na „ni s" i „kim").

Osim određivanja leksičke jedinice, Bogetić et al. posebno skreću pažnju na metaforu izraženu kosim padežima, što je specifičnost srpskog ali i hrvatskog jezika, te je i za potrebe ovog rada MIPVU metoda proširena uputama za označavanje metafore izražene kosim padežom:

- Za svaku riječ koja je deklinirana kosim padežom te stoji samostalno, bez prijedloga, utvrdite o kojem se padežu radi te koje je značenje u kontekstu.
- Utvrdite da li značenje padeža u kontekstu različito od osnovnog značenja, ali se može razumjeti u usporedbi s njim. Ukoliko da, označite riječ kao metaforu izraženu padežom (*MRW, inflectional*)

Važan korak u MIPVU proceduri je određivanje da li za riječ koja bi potencijalno mogla biti metafora, postoji osnovnije, jednostavnije značenje. Pri određivanju je potrebno konzultirati jednojezični rječnik, za što je odabran Veliki rječnik hrvatskoga standardnog jezika (Jojić i Nakić, 2015). Ukoliko bi nakon provjere u navedenom rječniku postojala nesigurnost ima li riječ osnovnije značenje, dodatno je vršena provjera na Hrvatskom jezičnom portalu[5].

Dodatno na MIPVU proceduru i njenu prilagodbu, anotatorice su upućene da pokušaju odrediti konvencionalnost metafore, na skali od 1 do 5, gdje je 1 potpuno nova a 5 potpuno konvencionalna metafora. Ovaj je korak dodan budući da kod Shutove (2017) nalazimo kako je upravo konvencionalnost bila predmet neslaganja među anotatorima/cama, budući da su neke metafore u potpunosti konvencionalne i uobičajene u jeziku pa ih neki anotatori nisu niti smatrali metaforama.

Budući da je za anotiranje metafore potrebno razumjeti i teorijske okvire poput konceptualne metafore, odabrane su anotatorice koje imaju formalno obrazovanje iz područja humanističkih znanosti.

3.2 Označeni skup podataka

Metaforu su anotirale dvije osobe, korištenjem tzv. *community* verzije alata Label Studio[6].

Cohen-Kappa slaganje među anotatoricama na razini rečenice je 82.79% za tekstove s Index.hr, dok je za tekstove s Jutarnjeg lista 90.08%. Slaganje među anotatoricama na razini riječi je nešto manje

[5] Available at https://hjp.znanje.hr
[5] Dostupno na https://hjp.znanje.hr
[6] https://labelstud.io/

a completely conventional metaphor. This step was added because in Shutova (2017) we find that it was precisely conventionality that was the subject of disagreement among the annotators, since some metaphors are completely conventional and common in the language, so some annotators did not even consider them to be metaphors.

Considering that it is necessary to understand theoretical frameworks such as the conceptual metaphor to annotate a metaphor, annotators who have formal education in the field of humanities were selected.

3.2 Annotated dataset

The metaphor was annotated by two people, using the so-called community version of the Label Studio tool. [6]

Cohen-Kappa agreement among annotators at the sentence level is 82.79% for texts from Index.hr, while for texts from Jutarnji list it is 90.08%. The agreement between the annotators at the word level is slightly less, the Cohen-Kappa for Index.hr is 67.69%, while for Jutarnji list it is 68.08%. Although the agreement here is less, it is still satisfactory. Namely, as shown by Reidsma and Carletta (2008), apart from the fact that an agreement between 67% and 80% is considered tolerable in computational linguistics, modern machine learning methods can accept less agreement if the disagreement acts as random noise in the data.

Table 1. Dataset overview. The metaphor column shows the number of sentences in which both annotators marked a metaphor and the percentage of the metaphor in the total number of sentences.

	No. of Sentences	Anotator A	Anotator B	Metaphor
Index.hr	1.265	515	576	492 38.89%
Jutarnji list	742	417	435	408 54.99%
Total	2,007	932	1011	900 44.84%
Dataset train test split				
Train	1,605			720 44.86%
Test	402			180 44.78%

It is interesting here that there are sentences for which both annotators agree that they contain a metaphor, but there is a disagreement about which word is the metaphor (e.g. "Posljednjih dana već su *stizali signali* da je "slučaj Marić" mnogo ozbiljniji nego što se u javnosti mogao dobiti dojam i da će odrediti daljnje odnose koalicijskih partnera. ", engl. "In recent days, signs are showing (in Croatian, pa je tako Cohen-Kappa za Index.hr 67.69% dok je za Jutarnji list 68.08%. Unatoč tome što je slaganje ovdje manje, i dalje je zadovoljavajuće. Naime, kao što su pokazali Reidsma i Carletta (2008), osim što se slaganje između 67% i 80% smatra podnošljivim u računarskoj lingvistici, suvremene metode strojnog učenja mogu prihvatiti manje slaganje, ukoliko neslaganje djeluje kao slučajni šum u podacima.

Tablica 1. Pregled skupa podataka. U stupcu metafora prikazan je broj rečenica u kojima su obje anotatorice označile metaforu te postotak metafore u ukupnom broju rečenica.

	Broj rečenica	Anotatorica A	Anotatorica B	Metafora
Index.hr	1.265	515	576	492 38.89%
Jutarnji list	742	417	435	408 54.99%
Ukupno	2.007	932	1.011	900 44.84%
Podjela skupa na podskupove za treniranje i testiranje				
Treniranje	1.605			720 44.86%
Testiranje	402			180 44.78%

Ovdje je zanimljivo kako postoje rečenice za koje se obje anotatorice slažu kako sadrže metaforu, no postoji neslaganje koja je riječ metafora (npr. „Posljednjih dana već su *stizali signali* da je "slučaj Marić" mnogo ozbiljniji nego što se u javnosti mogao dobiti dojam i da će odrediti daljnje odnose koalicijskih partnera.“ – anotatorica B kao metaforu označava samo *stizali*; „Naposljetku se po medijima *potegnulo pitanje* promjene samoga teksta prisege, a i upitno je postalo što činiti sa svim hrvatskim zakonima, pravnim aktima i propisima koji su pisani na isti način: u muškome rodu.“ – anotatorica B označava samo *potegnulo*; „Sasvim je sigurno da će se na *tragediji* ovih ljudi i dalje *skupljati jeftini* politički *bodovi*, a nedostatak pravog političkog programa i plana *zakrivati* populističkom retorikom.“, anotatorica A označava *skupljati*, *jeftini*, *bodovi*, *zakrivati* dok anotatorica B označava *tragediji*, *skupljati*, *zakrivati*).

Potrebno je imati na umu kako su označavanja u lingvistici, odnosno označavanja u tekstu kompleksni zadaci te anotatori/ce često odabiru različite oznake za istu stvar (usp. Reidsma & Carletta, 2008), čak i kada se radi o znatno jednostavnijim označavanjima od metafore.

U finalni skup podataka za identificiranje metafore, kao pozitivni primjeri uzete su one rečenice za koje se obje anotatorice slažu da sadrže metaforu. Sve ostale rečenice uzete su kao negativni primjeri, što uključuje rečenice za koje se obje

[6] https://labelstud.io/

literally "*signs are arriving*") that the "Marić case" is much more serious than the public could have known and that it will determine the future relations of the coalition partners." – annotator B marked only "*arriving*" as a metaphor; "Naposljetku se po medijima *potegnulo pitanje* promjene samoga teksta prisege, a i upitno je postalo što činiti sa svim hrvatskim zakonima, pravnim aktima i propisima koji su pisani na isti način: u muškome rodu. ", engl. "Ultimately, the issue of changing the text of the oath itself was discussed in the media (in Croatian, literally "*question was pulled*", and it became questionable what to do with all Croatian laws, legal acts and regulations that are written in the same way: in the masculine gender." - annotator B marked only "*pulled*"; "Sasvim je sigurno da će se na *tragediji* ovih ljudi i dalje *skupljati jeftini* politički *bodovi*, a nedostatak pravog političkog programa i plana *zakrivati* populističkom retorikom. ", engl. "It is quite certain that *cheap* political *points* will continue to be scored (in Croatian, literally "*collected*") on the *tragedy* of these people, and the lack of a real political program and plan will be *covered up* with populist rhetoric.", annotator A annotated *collected*, *cheap*, *points*, *cover up*, while annotator B annotated *tragedy*, *collected*, *cover up*).

It is necessary to keep in mind that annotation in linguistics, i.e. annotation in the text, is a complex tasks, and annotators often choose different tags for the same thing (cf. Reidsma & Carletta, 2008), even when it comes to much simpler annotations than annotating a metaphor.

The final dataset for metaphor identification as positive examples included only those sentences where metaphor was identified by both annotators. The rest of the sentences were labeled as negative examples and they include both the sentences where none of the annotators identified metaphor and the sentences where metaphor was identified by only one of the annotators (altogether 143 sentences). The final dataset contains 900 positive examples (492 Index.hr and 408 Jutarnji list) out of the total number of 2.007 sentences.

From the dataset, 20% of the examples were separated into the test set, using a stratification method to ensure an equal proportion of positive and negative examples in both the training and test sets. This means that the training set contains 1605 sentences, of which 720 are positive examples, while the testing set consists of 402 sentences, of which 180 are positive examples.

In the pre-processing of the data, all letters are converted to lowercase and all special characters are deleted.

The prepared dataset is publicly available at GitHub.[7]

anotatorice slažu kako ne sadrže metaforu ali i one gdje jedna anotatorica smatra da postoji metafora, no druga se ne slaže (takvih je rečenica 143). Finalni skup podataka sadrži 900 pozitivnih primjera (492 Index.hr te 408 Jutarnji list), od ukupno 2.007 rečenica.

Iz skupa podataka, 20% primjera je odvojeno u skup za testiranje, pri čemu je korištena metoda stratifikacije kako bi se osigurao jednak omjer pozitivnih i negativnih primjera i u skupu za treniranje i u skupu za testiranje. Tako skup za treniranje sadrži 1605 rečenica, od čega 720 pozitivnih primjera, dok se skup za testiranje sastoji od 402 rečenice, od čega je 180 pozitivnih primjera.

U pred-procesiranju podataka, sva su slova pretvorena u mala te su izbrisani svi posebni znakovi.

Pripremljeni skup podataka javno je dostupan na GitHubu.[7]

3.3 Opažanja pri označavanju metafore

Označavanje metafore kompleksan je i dugotrajan zadatak (za anotiranje je svakoj anotatorici bilo potrebno oko 70 sati rada). MIPVU procedura, na kojoj se temelji procedura primijenjena u ovom radu, na prvo čitanje djeluje jednostavna no potrebno je stalno vraćanje na upute kako bi se utvrdilo o kakvoj se metafori radi.

Nakon nekog vremena provedenog u anotiranju obje anotatorice primjećuju kako praćenjem uputa i anotiranjem, sve počinje djelovati kao metafora. Slična opažanja nalazimo i kod (Krennmayr i Steen, 2017) koji navode kako, ipak, unatoč tome što MIPVU procedura ostavlja dojam kako bi sve mogla biti metafora, u konačnici je metafora zastupljena u samo 13,6% leksičkih jedinica u VUA skupu.

Anotatorice navode i kako je teško biti konzistentan pri određivanju stupnja konvencionalnosti metafore ali i pri samom određivanju da li se neka riječ koristi metaforički, čak i u slučajevima kada se koristi u istom kontekstu. To zahtjeva stalno vraćanje na prethodno označene rečenice, što ponekad rezultira promjenom mišljenja o prethodnoj anotaciji.

U procesu anotiranja primijećeno je kako novinski tekstovi sadrže najviše metafora koje su u potpunosti ili skoro u potpunosti konvencionalne, odnosno označene su s 4 ili 5 na skali konvencionalnosti. Anotatorica A od ukupno 1.193 metafore (u 932 rečenice), njih 96% označava s 4 ili 5, dok Anotatorica B od ukupno 1.234 metafore (u 1.011 rečenica), njih 87% označava s 4 ili 5. Anotatorice primjećuju kako su neke metafore toliko konvencionalne kako je upitno da li je uopće ispravno smatrati ih metaforama (npr. „Njegov je

[7] https://github.com/pticek/Identification-of-Metaphors-Cro-Newspaper.
[7] https://github.com/pticek/Identification-of-Metaphors-Cro-Newspaper

3.3 Observations on metaphor annotation

Annotating metaphors is a complex and time-consuming task (each annotator needed about 70 hours of work to annotate). The MIPVU procedure, on which the procedure applied in this paper is based, seems simple at first, but the annotator must constantly return to the instructions in order to determine what kind of metaphor it is.

After some time spent annotating, both annotators notice that after a while of following the instructions and annotating, everything starts to seem as a metaphor. We find similar observations in (Krennmayr and Steen, 2017), who state that, despite the fact that the MIPVU procedure gives the impression that everything could be a metaphor, ultimately the metaphor is represented in only 13.6% of the lexical items in the VUA set.

The annotators state that it is difficult to be consistent when determining the degree of conventionality of a metaphor, but also when determining whether a word is used metaphorically, even in cases where it is used in the same context. This requires constantly returning to previously annotated sentences, which sometimes results in a change of mind about the previous annotation.

In the annotation process, it was noticed that the newspaper texts contain the most metaphors that are completely or almost completely conventional, that is, they are marked with 4 or 5 on the conventionality scale. Annotator A, out of a total of 1,193 metaphors (in 932 sentences), marked 96% with a 4 or 5, while Annotator B, out of a total of 1,234 metaphors (in 1,011 sentences), marked 87% with a 4 or 5. The annotators note that some metaphors are so conventional that it is questionable whether they should even be considered metaphors (e.g. "Njegov je jedini cilj zaštititi Inu kao kompaniju od nacionalnog interesa - kažu nam *izvori*.", engl. "His only goal is to protect INA as a company of national interest - *sources* tell us.")

4 Classification model settings

Metaphor identification was approached as a problem of text sequence classification, at the sentence level. In the classification, large language models were used, trained on datasets in the Croatian language (BERTić, CroSloEngual and XLM-RoBERTa base) using the Huggingface library (Wolf et al., 2020).

The same library was used for tokenization, by calling the AutoTokenizer classes when using the BERTić and CroSloEngual models, or by calling the XLMRobertaTokenizer class in the model that uses the XLM-RoBERTa language model.

The Huggingface classes AutoModelForSequenceClassification were used for classification with the BERTić and CroSloEngual

jedini cilj zaštititi Inu kao kompaniju od nacionalnog interesa - kažu nam *izvori*.“)

4 Postavke modela za klasifikaciju

Identificiranju metafore pristupilo se kao problemu klasifikacije sekvenci teksta, i to na razini rečenice. Pri klasifikaciji su korišteni veliki jezični modeli, trenirani i na skupovima podataka na hrvatskom jeziku, (BERTić, CroSloEngual i XLM-RoBERTa base) i to korištenjem Huggingface biblioteke (Wolf et al., 2020).

Ista je biblioteka korištena za tokenizaciju, pozivanjem klasa AutoTokenizer pri korištenju BERTić i CroSloEngual modela, odnosno pozivanjem klase XLMRobertaTokenizer u modelu koji koristi XLM-RoBERTa jezični model.

Za klasifikaciju su korištene Huggingface klase AutoModelForSequenceClassification pri klasifikaciji BERTić i CroSloEngual modelom, odnosno XLMRobertaForSequenceClassification za klasifikaciju XLM-RoBERTa modelom.

Za svaki je model rađeno fino podešavanje (engl. *fine-tuning*), te su tako napravljena četiri eksperimenta sa svakim modelom i to (1) broj epoha 3, stopa učenja 3e-5; (2) broj epoha 3, stopa učenja 2e-5; (3) broje epoha 5, stopa učenja 3e-5; (4) broj epoha 5, stopa učenja 2-e5. Veličina serije (engl. *batch size*) je uvijek 32 te su rezultati klasifikacije bilježeni za svaku epohu testiranja te je kao konačan rezultat uzet najbolji rezultat za epohu. Modeli su izvršavani na Google Colab okruženju, korištenjem GPU-a.

5 Rezultati

Rezultati klasifikacije na pozitivnim primjerima prikazani su u tablici 2, gdje možemo vidjeti kako je najbolji F1 rezultat od 70,75% ostvaren s BERTić modelom, pri testu s 5 epoha i stopi učenja 3e-5. Preciznost je u ovom slučaju 70,95% dok je odziv 70,56%. U tablici 3 prikazan je cijeli izvještaj klasificiranja za BERTić model, pa tako vidimo kako je točnost 73,88%, dok je ponderirani F1 rezultat 73,87%. Najbolji je rezultat ostvaren u trećoj epohi, nakon čega F1 na pozitivnim primjerima pada na 70,25% u četvrtoj epohi pa na 69% u petoj epohi.
Sljedeći najbolji F1 rezultat je 67,60% i ostvaren je s XLM-RoBERT-a modelom, u testu s 5 epoha i stopom učenja 2e-5.

Sva tri modela pravilno su klasificirala 80 rečenica, odnosno 44,4% od svih pozitivnih primjera, dok su sva tri pogrešno klasificirala 33 iste rečenice, što čini 18.3%.

models, and the XLMRobertaForSequenceClassification for classification with the XLM-RoBERTA model.

Fine-tuning was done for each model, and thus four experiments were made with each model, namely (1) number of epochs 3, learning rate 3e-5; (2) number of epochs 3, learning rate 2e-5; (3) number of epochs 5, learning rate 3e-5; (4) number of epochs 5, learning rate 2-e5. The batch size is always 32 and the classification results are recorded for each testing epoch, while the best result in an epoch is regarded as the best result of the experiment. The models were executed on the Google Colab environment, using GPU.

5 Results

The classification results on positive examples are shown in Table 2, where we can see that the best F1 result of 70.75% was achieved with the BERTić model, with a test with 5 epochs and a learning rate of 3e-5. In this case, the precision is 70.95%, while the recall is 70.56%. Table 3 shows the entire classification report for the BERTić model, and we can see that the accuracy is 73.88%, while the weighted F1 result is 73.87%. The best result was achieved in the third epoch, after which F1 on positive examples drops to 70.25% in the fourth epoch and to 69% in the fifth epoch.
The next best F1 result is 67.60% and was achieved with the XLM-RoBERT model, in a test with 5 epochs and a learning rate of 2e-5.

All three models correctly classified 80 sentences, or 44.4% of all positive examples, while all three incorrectly classified 33 of the same sentences, which is 18.3%.

Table 2. Model comparison – precision, recall and F1 measure for positive examples. The best F1 result of 70.75% was achieved with the BERTić model, in a test with 5 epochs and a learning rate of 3e-5. Bold letters indicate the best F1 results for each of the language models used.

	3 ep, 2e-5	**3 ep., 3e-5**	**5 ep., 2e-5**	**5 ep., 3e-5**
BERTić				
Precision	73.68%	65.41%	68.00%	70.95%
Recall	54.44%	67.22%	66.11%	70.56%
F1	62.62%	66.30%	67.04%	**70.75%**
CroSloEngual				
Precision	51.61%	65.17%	64.29%	70.95%
Recall	88.89%	64.44%	70.00%	58.33%
F1	65.31%	64.80%	**67.02%**	64.02%
XLM-RoBERTa				

Tablica 2. Usporedba modela – preciznost, odziv i F1 mjera za pozitivne primjere. Najbolji F1 rezultat od 70.75% ostvaren je s BERTić modelom, pri testu s 5 epoha i stopom učenja 3e-5. Masnim slovima označeni su najbolji F1 rezultati za svaki od korištenih jezičnih modela.

	3 ep, 2e-5	**3 ep., 3e-5**	**5 ep., 2e-5**	**5 ep., 3e-5**
BERTić				
Preciznost	73,68%	65,41%	68,00%	70,95%
Odziv	54,44%	67,22%	66,11%	70,56%
F1	62,62%	66,30%	67,04%	**70,75%**
CroSloEngual				
Preciznost	51,61%	65,17%	64,29%	70.95%
Odziv	88,89%	64,44%	70,00%	58,33%
F1	65,31%	64,80%	**67,02%**	64,02%
XLM-RoBERTa				
Preciznost	61,50%	55,97%	67,98%	65,26%
Odziv	68,33%	83,33%	67,22%	68,89%
F1	64,74%	66,96%	**67,60%**	67,03%

Tablica 3. Izvještaj klasificiranja za BERTić model, kojim je, u usporedbi s ostalim modelima, ostvaren najbolji rezultat.

	Preciznost	**Odziv**	**F1**
Negativni	76,23%	76,58%	76,40%
Pozitivni	70,95%	70,56%	70,75%
Točnost			73,88%
Makro prosjek	73,59%	73,57%	73,58%
Ponderirani prosjek	73,87%	73,88%	73,87%

S obzirom na ovo opažanje, napravljena su tri dodatna klasificiranja korištenjem BERTić modela (5 epoha, stopa učenja 3e-5), i to (1) zasebna klasifikacija tekstova s Index.hr odnosno Jutarnjeg lista, (2) skup za treniranje su rečenice s Index.hr a za testiranje Jutarnji list i (3) skup za treniranje su rečenice s Jutarnjeg lista a za testiranje s Index.hr.

Kao što možemo vidjeti u tablici 4, rezultati za klasifikaciju rečenica s Jutarnjeg lista su, za pozitivne primjere, značajno bolji nego za Index.hr – za Jutarnji list F1 mjera za je 67,36% dok je za Index.hr 55,62%. S druge pak strane, rezultati za klasifikaciju negativnih primjera znatno su bolji za Index.hr nego za Jutarnji - F1 42,59% za Jutarnji list, naspram 77,74% za Index.hr.

U tablici 5 prikazani su rezultati testova odvajanjem jednog izvora podataka za treniranje a drugog za testiranje. Kao što možemo vidjeti, rezultati klasifikacije su znatno bolji kada se za treniranje koristi Index.hr a za testiranje Jutarnji list

Precision	61.50%	55.97%	67.98%	65.26%
Recall	68.33%	83.33%	67.22%	68.89%
F1	64.74%	66.96%	**67.60%**	67.03%

By analyzing the misclassified sentences, we can see that there are more of these from Index.hr than from Jutarnji list. In the test dataset there are 101 positive examples from Index.hr (56%) and 79 from Jutarnji list (44%), while BERTić incorrectly classified 41 sentences from Index.hr (66%) and 21 from Jutarnji list (34%). As shown in Table 1, the texts from Jutarnji list contain a higher percentage of metaphors in the total number of sentences from that portal, but in general there are more examples from Index.hr.

Table 3. Classification report for the BERTić model, which, in comparison with other models, achieved the best result.

	Precision	**Recall**	**F1-score**
Negative	76,23%	76.58%	76.40%
Positive	70.95%	70.56%	70.75%
Accuracy			73.88%
Macro avg	73,59%	73.57%	73.58%
Weighted avg	73.87%	73.88%	73.87%

Considering this observation, three additional classifications were made using the BERTić model (5 epochs, learning rate 3e-5), namely (1) a separate classification of texts from Index.hr and Jutarnji list, (2) a training set consisting of sentences with Index.hr and for testing Jutarnji list and (3) a training set that contains sentences from Jutarnji list and for testing from Index.hr.

As we can see in Table 4, the results for the classification of sentences from Jutarnji list are, for positive examples, significantly better than for Index.hr - for Jutarnji list the F1 measure is 67.36%, while for Index.hr it is 55.62% . On the other hand, the results for the classification of negative examples are significantly better for Index.hr than for Jutarnji - F1 42.59% for Jutarnji list, compared to 77.74% for Index.hr.

Table 5 shows the test results achieved by separating one data source for training and another for testing. As we can see, the classification results are significantly better when Index.hr is used for training and Jutarnji list is used for testing (F1 for positive examples 69.92%), rather than vice versa (F1 for positive examples 58.71%).

If we compare the sentences that were misclassified in the first test with the data set of 2007 sentences from both sources with those that were misclassified in the tests where the set of one source was used for training or testing, we come to an interesting observation in the case when the sentences (F1 za pozitivne primjere 69,92%), nego li obratno (F1 za pozitivne primjere 58,71%).

Usporedimo li rečenice koje su pogrešno klasificirane u prvom testu sa skupom podataka od 2007 rečenice iz oba izvora, s onima koje su pogrešno klasificirane u testovima gdje je skup jednog izvora korišten za treniranje odnosno za testiranje, dolazimo do zanimljivog opažanja u slučaju kada se rečenice s Index.hr koriste za testiranje – 21 rečenica od 41 koje su u prvom testu pogrešno klasificirane, u ovom su slučaju ispravno klasificirane. S druge strane, kada je Index.hr korišten za treniranje a Jutarnji list za testiranje, od 21 rečenice s Jutarnjeg lista koje su u prvom testu pogrešno klasificirane, u ovom slučaju je njih 18 pogrešno klasificirano.

Ovdje je potrebno svakako imati na umu kako svaki portal tj novinari portala imaju svoj stil pisanja te bi bilo potrebno napraviti dodatnu lingvističku analizu samih primjera kako bi se pokušao pronaći uzorak u njima. Podrobnija bi nam lingvistička analiza tekstova potencijalno mogla pokazati i ograničenja samih velikih jezičnih modela, pri identificiranju metafore, što je van opsega ovog rada te se predlaže za buduće istraživanje.

Budući da su u skupu podataka pozitivni primjeri samo one rečenice gdje se obje anotatorice slažu da postoji metafora u rečenici, analizom rezultata testiranja provjereno je koliko je negativnih primjera BERTić klasificirao kao pozitivne, a da se poklapaju s primjerima koje je samo jedna od anotatorica označila kao metaforu. U skupu za testiranje nalazi se 28 negativnih primjera koji su negativni jer ih je samo jedna od anotatorica označila kao rečenice s metaforom te je od tog broja 11 rečenica klasificirano kao pozitivno. Svakako bi u narednim istraživanjima bilo zanimljivo vidjeti kakvi bi rezultati bili kada bi se kao pozitivni primjeri uključile i one rečenice koje je samo jedna anotatorica označila kao rečenice s metaforom.

Tablica 4. Rezultati zasebne klasifikacije rečenica s Jutarnjeg lista i Index.hr Tablica prikazuje F1, odziv i preciznost za pozitivne, odnosno negativne primjere.

	Preciznost	**Odziv**	**F1**
	Pozitivni primjeri		
Jutarnji list	59,26%	78,05%	67,37%
Index.hr	66,20%	47,95%	55,62%
	Negativni primjeri		
Jutarnji list	56,10%	34,33%	42,59%
Index.hr	71,98%	84,51%	77,74%

Tablica 5. Rezultati klasifikacija podjelom skupova podataka na skup za testiranje i testiranje prema izvoru podatka.

from Index.hr are used for testing - 21 sentences out of 41 that were incorrectly classified in the first test were correctly classified in this case. On the other hand, when Index.hr was used for training and Jutarnji list for testing, out of 21 sentences from Jutarnji list that were misclassified in the first test, 18 of them were misclassified in this case.

Here it is necessary to keep in mind that each portal, i.e. portal journalist, has its own writing style, and it would be necessary to do an additional linguistic analysis of the examples themselves in order to try to find a pattern in them. A more detailed linguistic analysis of the texts could potentially show us the limitations of the large language models themselves when identifying metaphors, which is beyond the scope of this paper and is suggested for future research.

Since the positive examples in this dataset are only the sentences in which metaphor was identified by both annotators, an additional analysis of the testing results was done in order to check how many negative examples were classified as positive by BERTić which at the same time were labeled as metaphor by only one annotator. In the test set there are 28 negative examples that were labeled as negative because only one of the annotators has identified metaphor in them and out of this pool 11 sentences were classifies as positive. Therefore, it would certainly be interesting to further explore this issue in future research and conduct experiments on a dataset that would include as positive, also the sentences where metaphor was identified by only one of the annotators.

Table 4. Results of separate classification of sentences from Jutarnji list and Index.hr The table shows precision, recall and F1 for positive and negative examples.

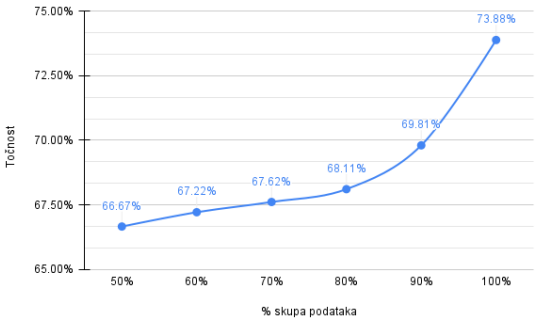
	Precision	**Recall**	**F1**
	Positive examples		
Jutarnji list	59,26%	78,05%	67,37%
Index.hr	66,20%	47,95%	55,62%
	Negative examples		
Jutarnji list	56,10%	34,33%	42,59%
Index.hr	71,98%	84,51%	77,74%

Table 5. Results of classification achieved by dividing data sets into a test set and testing according to data source.

	Precision	**Recall**	**F1**
Train set Index.hr, test set Jutarnji list			
Positive	73,51%	66,66%	69,92%
Negative	63,44%	70,65%	66,85%
Train set Jutarnji list, test set Index.hr			
Positive	52,57%	66,46%	58,71%
Negative	74,34%	61,84%	67,51%

	Preciznost	**Odziv**	**F1**
Treniranje Index.hr, testiranje Jutarnji list			
Pozitivni	73,51%	66,66%	69,92%
Negativni	63,44%	70,65%	66,85%
Treniranje Jutarnji list, testiranje Index.hr			
Pozitivni	52,57%	66,46%	58,71%
Negativni	74,34%	61,84%	67,51%

Kako bi se utvrdilo koliki je rast performansi modela ovisno o veličini skupa podataka, s BERTić modelom (5 epoha, 3e-5) napravljeni su testovi s 50%, 60%, 70%, 80% i 90% skupa, gdje je u svim skupovima omjer rečenica s Index.hr i Jutarnjeg lista bio isti kao u cijelom skupu podataka. Kao što možemo vidjeti na slici 1, točnost raste što je skup podataka veći te su najbolji rezultati ostvareni u 3 epohi, nakon čega rezultati padaju. Također možemo uočiti kako postoji veliki skok u točnosti između skupova podataka sa 90% i 100% primjera, što svakako ukazuje kako je potrebno napraviti dodatna istraživanja s većim skupovima podataka.



Slika 1. Usporedba točnosti modela ovisno o veličini skupa podataka.

6 Zaključak

U radu je izložena prilagodba MIPVU procedure za anotiranje metafore u hrvatskim novinskim člancima, kao i opažanja o anotiranju. Pokazalo se kako je, budući da je metafora kompleksan fenomen, anotiranje metafore zadatak koji zahtjeva mnogo vremena, ali i razumijevanje teorijskih okvira na kojima počivaju procedure anotiranja,.

Identificiranje metafore na razini rečenice modelom klasifikacije sekvenci teksta i korištenjem velikih jezičnih modela dalo je zadovoljavajuće rezultate. Najbolji rezultat ostvaren je korištenjem BERTić modela - F1 rezultat 70,75%. Analizom pogrešno klasificiranih pozitivnih primjera, uočeno je kako je model bolje klasificirao primjere s Jutarnjeg lista nego s Index.hr portala. Svakako je potrebna podrobnija lingvistička analiza pogrešno klasificiranih primjera, kako bi se u njima pokušao

In order to determine the increase of the model's performance depending on the size of the dataset, tests were made with the BERTić model (5 epochs, 3e-5) with 50%, 60%, 70%, 80% and 90% of the set, where in all sets, the ratio of sentences from Index.hr and Jutarnji list was the same as in the entire data set. As we can see in Figure 1, the accuracy increases as the dataset is larger and the best results are achieved in 3 epochs, after which the results drop. We can also see that there is a large jump in accuracy between the datasets with 90% and 100% examples, which certainly indicates that additional research with larger datasets is required.

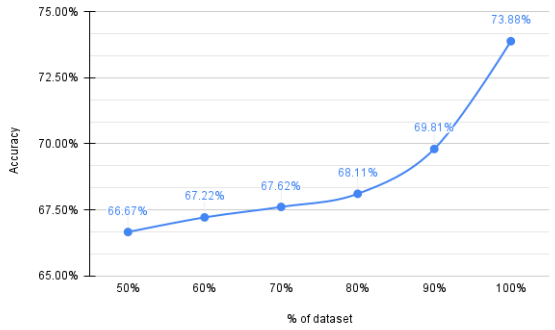


Figure 1. Comparison of model accuracy depending on dataset size.

6 Conclusion

The paper presents the adaptation of the MIPVU procedure for annotating metaphors in Croatian newspaper articles, as well as observations about the annotation. It has been shown that, since the metaphor is a complex phenomenon, annotating a metaphor is a task that requires a lot of time, but also an understanding of the theoretical frameworks on which annotation procedures are based.

Identifying metaphors at the sentence level with the text sequence classification model and by using large language models provided satisfactory results. The best result was achieved by using the BERTić model - an F1 result 70.75%. By analyzing the misclassified positive examples, it was observed that the model better classified the examples from Jutarnji list than those from the Index.hr portal. A more detailed linguistic analysis of misclassified examples is certainly required in order to try to find a pattern in them, which would potentially indicate the limitations of large language models.

A comparison of model results depending on the size of the dataset showed that the accuracy of the model increases with the size of the dataset, and a larger dataset will be prepared in future research.

Bearing in mind that metaphor is expressed on the level of a word (or a few words), the limitation of the model presented in this paper is that metaphor is identifyed on the level of a sentence. In further pronaći uzorak, a koji bi potencijalno ukazao i na ograničenja velikih jezičnih modela.

Usporedba rezultata modela ovisno o veličini skupa podataka pokazala je kako s veličinom skupa raste i točnost modela te će se u budućim istraživanjima pripremiti veći skup podataka.

Identificiranje metafore na razini rečenice je ograničenje modela predstavljenog u ovom radu, budući da se metafora izražava na razini riječi (ili nekoliko riječi) u rečenici. U daljnjim će se istraživanjima pristupiti i izradi modela koji identificiraju metaforu na razini riječi (tokena), odnosno leksičke jedinice.

Također, za naredna je istraživanja potrebno izraditi skup podataka koji sadrži akademske tekstove, posebice o književnosti i jeziku, budući da su prethodna istraživanja pokazala veliku zastupljenost metafore u tim tekstovima.

Reference

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.

Bogetić, K., Bročić, A., & Rasulić, K. (2019). Linguistic metaphor identification in Serbian. In S. Nacey, A. G. Dorst, T. Krennmayr, & W. G. Reijnierse (Eds.), Metaphor Identification in Multiple Languages (pp. 203–226). John Benjamins. https://doi.org/10.1075/celcr.22.10bog

Chen, X., Leong, C. W., Flor, M., & Klebanov, B. B. (2020). Go Figure! Multi-task transformer-based architecture for metaphor detection using idioms: ETS team in 2020 metaphor shared task. Proceedings of the Second Workshop on Figurative Language Processing. Association for Computational Linguistics, 235–243. https://doi.org/10.18653/v1/2020.figlang-1.32

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

Dankers, V., Malhotra, K., Kudva, G., Medentsiy, V., & Shutova, E. (2020). Being neighbourly: Neural metaphor identification in discourse. Proceedings of the Second Workshop on Figurative Language Processing. Association

research, the creation of models that identify metaphors at the level of words (tokens), or lexical units, will be approached.

Also, for subsequent research, it is necessary to create a dataset that contains academic texts, especially on literature and language, since previous research has shown a high prevalence of metaphors in these texts.

References

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.

Bogetić, K., Bročić, A., & Rasulić, K. (2019). Linguistic metaphor identification in Serbian. In S. Nacey, A. G. Dorst, T. Krennmayr, & W. G. Reijnierse (Eds.), Metaphor Identification in Multiple Languages (pp. 203–226). John Benjamins. https://doi.org/10.1075/celcr.22.10bog

Chen, X., Leong, C. W., Flor, M., & Klebanov, B. B. (2020). Go Figure! Multi-task transformer-based architecture for metaphor detection using idioms: ETS team in 2020 metaphor shared task. Proceedings of the Second Workshop on Figurative Language Processing. Association for Computational Linguistics, 235–243. https://doi.org/10.18653/v1/2020.figlang-1.32

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

Dankers, V., Malhotra, K., Kudva, G., Medentsiy, V., & Shutova, E. (2020). Being neighbourly: Neural metaphor identification in discourse. Proceedings of the Second Workshop on Figurative Language Processing. Association for Computational Linguistics, 227–234. https://doi.org/10.18653/v1/2020.figlang-1.31

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

Gao, G., Choi, E., Choi, Y., & Zettlemoyer, L. (2018). Neural Metaphor Detection in Context. Proceedings of the 2018 Conference on Empirical Methods in Natural Language for Computational Linguistics, 227–234. https://doi.org/10.18653/v1/2020.figlang-1.31

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

Gao, G., Choi, E., Choi, Y., & Zettlemoyer, L. (2018). Neural Metaphor Detection in Context. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 607–613. https://doi.org/10.18653/v1/D18-1060

Gong, H., Gupta, K., Jain, A., & Bhat, S. (2020). IlliniMet: Illinois System for Metaphor Detection with Contextual and Linguistic Information. Proceedings of the Second Workshop on Figurative Language Processing. Association for Computational Linguistics, 146–153. https://doi.org/10.18653/v1/2020.figlang-1.21

Jojić, L., & Nakić, A. (Eds.). (2015). Veliki rječnik hrvatskoga standardnog jezika. Školska knjiga.

Katz, J. J., & Fodor, J. A. (1963). The Structure of a Semantic Theory. Language (Linguistic Society of America), 39(2), 170–210. https://doi.org/10.2307/411200

Krennmayr, T., & Steen, G. (2017). VU Amsterdam Metaphor Corpus. In Handbook of Linguistic Annotation, by Nancy Ide and James Pustejovsky, 1053–1071. Springer.

Lakoff, G., & Johnson, M. (1980). Metaphors We Live By. University of Chicago Press.

Lin, Z., Ma, Q., Yan, J., & Chen, J. (2021). CATE: A Contrastive Pre-trained Model for Metaphor Detection with Semi-supervised Learning. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, 3888–3898. https://doi.org/10.18653/v1/2021.emnlp-main.316

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://doi.org/10.48550/arXiv.1907.11692

Ljubešić, N., & Lauc, D. (2021). BERTić - The Transformer Language Model for Bosnian, Croatian, Montenegrin and Serbian. Proceedings of the 8th BSNLP Workshop on Balto-Slavic Natural Language Processing, 37–42.

Mao, R., Lin, C., & Guerin, F. (2019). End-to-End Sequential Metaphor Identification Inspired by

Processing, 607–613. https://doi.org/10.18653/v1/D18-1060

Gong, H., Gupta, K., Jain, A., & Bhat, S. (2020). IlliniMet: Illinois System for Metaphor Detection with Contextual and Linguistic Information. Proceedings of the Second Workshop on Figurative Language Processing. Association for Computational Linguistics, 146–153. https://doi.org/10.18653/v1/2020.figlang-1.21

Jojić, L., & Nakić, A. (Eds.). (2015). Veliki rječnik hrvatskoga standardnog jezika. Školska knjiga.

Katz, J. J., & Fodor, J. A. (1963). The Structure of a Semantic Theory. Language (Linguistic Society of America), 39(2), 170–210. https://doi.org/10.2307/411200

Krennmayr, T., & Steen, G. (2017). VU Amsterdam Metaphor Corpus. In Handbook of Linguistic Annotation, by Nancy Ide and James Pustejovsky, 1053–1071. Springer.

Lakoff, G., & Johnson, M. (1980). Metaphors We Live By. University of Chicago Press.

Lin, Z., Ma, Q., Yan, J., & Chen, J. (2021). CATE: A Contrastive Pre-trained Model for Metaphor Detection with Semi-supervised Learning. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, 3888–3898. https://doi.org/10.18653/v1/2021.emnlp-main.316

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://doi.org/10.48550/arXiv.1907.11692

Ljubešić, N., & Lauc, D. (2021). BERTić - The Transformer Language Model for Bosnian, Croatian, Montenegrin and Serbian. Proceedings of the 8th BSNLP Workshop on Balto-Slavic Natural Language Processing, 37–42.

Mao, R., Lin, C., & Guerin, F. (2019). End-to-End Sequential Metaphor Identification Inspired by Linguistic Theories. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3888–3898. https://doi.org/10.18653/v1/P19-1378

Nacey, S., Dorst, A. G., Krennmayr, T., & Reijnierse, W. G. (2019). Metaphor Identification in Multiple Languages. John Benjamins. https://doi.org/10.1075/celcr.22

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543. https://doi.org/10.3115/v1/D14-1162

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 2227–2237. https://doi.org/10.18653/v1/N18-1202

Pragglejaz Group. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. Metaphor and Symbol, 1, 1–39.

Rei, M., Bulat, L., Kiela, D., & Shutova, E. (2017). Grasping the Finer Point: A Supervised Similarity Network for Metaphor Detection. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 1537–1546. https://doi.org/10.18653/v1/D17-1162

Reidsma, D., & Carletta, J. (2008). Reliability Measurement without Limits. Computational Linguistics, 34(3), 319–326. https://doi.org/10.1162/coli.2008.34.3.319

Shutova, E. (2017). Annotation of Linguistic and Conceptual Metaphor. In Handbook of Linguistic Annotation, by Nancy Ide and James Pustejovsky, 1073-1100 (pp. 1073–1100). Springer. https://doi.org/10.1007/978-94-024-0881-2_40

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). A Method for Linguistic Metaphor Identification. John Benjamins. https://doi.org/10.1075/celcr.14

Su, C., Fukumoto, F., Huang, X., Li, J., Wang, R., & Chen, Z. (2020). DeepMet: A Reading Comprehension Paradigm for Token-level Metaphor Detection. Proceedings of the Second Workshop on Figurative Language Processing. Association for Computational Linguistics, 30–39. https://doi.org/10.18653/v1/2020.figlang-1.4

Processing (EMNLP), 1532–1543. https://doi.org/10.3115/v1/D14-1162

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 2227–2237. https://doi.org/10.18653/v1/N18-1202

Pragglejaz Group. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. Metaphor and Symbol, 1, 1–39.

Rei, M., Bulat, L., Kiela, D., & Shutova, E. (2017). Grasping the Finer Point: A Supervised Similarity Network for Metaphor Detection. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 1537–1546. https://doi.org/10.18653/v1/D17-1162

Reidsma, D., & Carletta, J. (2008). Reliability Measurement without Limits. Computational Linguistics, 34(3), 319–326. https://doi.org/10.1162/coli.2008.34.3.319

Shutova, E. (2017). Annotation of Linguistic and Conceptual Metaphor. In Handbook of Linguistic Annotation, by Nancy Ide and James Pustejovsky, 1073-1100 (pp. 1073–1100). Springer. https://doi.org/10.1007/978-94-024-0881-2_40

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). A Method for Linguistic Metaphor Identification. John Benjamins. https://doi.org/10.1075/celcr.14

Su, C., Fukumoto, F., Huang, X., Li, J., Wang, R., & Chen, Z. (2020). DeepMet: A Reading Comprehension Paradigm for Token-level Metaphor Detection. Proceedings of the Second Workshop on Figurative Language Processing. Association for Computational Linguistics, 30–39. https://doi.org/10.18653/v1/2020.figlang-1.4

Ulčar, M., & Robnik-Šikonja, M. (2020). FinEst BERT and CroSloEngual BERT. In Lecture Notes in Computer Science (pp. 104–111). https://doi.org/10.1007/978-3-030-58323-1_11

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Proceedings of the 31st Annual Conferenceon Neural Information Processing Systems, 6000–6010.

Wilks, Y. (1975). A preferential, pattern-seeking, Semantics for natural language inference. Artificial Intelligence , 6(1), 53–74. https://doi.org/10.1016/0004-3702(75)90016-8

Ulčar, M., & Robnik-Šikonja, M. (2020). FinEst BERT and CroSloEngual BERT. In Lecture Notes in Computer Science (pp. 104–111). https://doi.org/10.1007/978-3-030-58323-1_11

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Proceedings of the 31st Annual Conferenceon Neural Information Processing Systems, 6000–6010.

Wilks, Y. (1975). A preferential, pattern-seeking, Semantics for natural language inference. Artificial Intelligence , 6(1), 53–74. https://doi.org/10.1016/0004-3702(75)90016-8

Wilks, Y. (1978). Making preferences more active. Artificial Intelligence, 11(3), 197–223. https://doi.org/10.1016/0004-3702(78)90001-2

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Wilks, Y. (1978). Making preferences more active. Artificial Intelligence, 11(3), 197–223. https://doi.org/10.1016/0004-3702(78)90001-2

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Some Remarks on the Formal Operations Inspired by the Gene Assembly in Ciliates

Victor Mitrana
Universidad Politecnica de Madrid and National Institute for R&D for Biological Sciences
Department of Information Systems and Bioinformatics Department
Calle Alan Turing s/n, 28031, Madrid, Spain
victor.mitrana@upm.es

Andrei Păun, Mihaela Păun
National Institute for R&D for Biological Sciences
Bioinformatics Department
296 Independenței Bd., Bucharest 060031, Romania
{apaun@gmail.com, mihaela.paun@incdsb.ro}

Jose-Ramón Sánchez-Couso
Universidad Politecnica de Madrid
Department of Information Systems
Calle Alan Turing s/n, 28031, Madrid, Spain
joseramon.sanchezcouso@upm.es

Abstract. *We continue here the theoretical study initiated approximately twenty years ago on the possibility of using living cells for computing. In this paper, we reconsider the formal operations inspired by the intramolecular DNA rearrangements in the evolution of the macronucleus from the micronucleus in a group of ciliates. After introducing the concept of a valid string, we propose an efficient algorithm for checking this property for a given string. Then we investigate which of the considered operations preserve the property of a string to be valid. We also show that just one of the operations can be simulated by a finite transducer. The important problem regarding the order of applying the operations is then investigated showing that one operation can commute with the other two. Finally, we introduce the iterated variants and investigate a few properties. A sort of a "normal form" for the gene assembly in ciliates is obtained. The paper ends by a short discussion about open problems and further directions of research.*

Keywords. Ciliate; gene assembly, valid string, intramolecular operations, finite transducer.

1 Introduction

Do the cells "compute"? This seems to be one of the main questions in the theory of computing, especially nowadays when more and more complex computational problems arise in different areas. Some directions of research are aimed to find exact algorithms using current hardware architectures, while other approaches enable approximate, uncertain, tolerant solutions which are mainly based on other architectures than the classic ones. Many directions of research (Machine learning, Fuzzy sets, Evolutionary computing, Artificial neural networks, Expert systems, etc.) have been emerged and vividly developed under the umbrella of *Soft Computing*. Other approaches have eventually formed the *Unconventional Computing* paradigm that includes: *Natural Computing (DNA and cellular (in vivo) computing, biochemical networks, etc.) and Quantum Computing*, see (Kari & Rozenberg, 2008). This work is a contribution to the area of the so-called "in-vivo" computing or the computation with living cells. If the cells really compute, are we able to understand the way in which they do it and which are the problems "solved" by them? It is commonly accepted that, most likely, they do this by reading and modifying DNA sequences all the time. Among many computational models based on the manipulation of DNA, see, eg., (Păun, et al., 1998) and (Rozenberg et al., 2012), there is a model of a rather different type that is based on the way in which some microorganisms, calls *ciliates*, are copying some protein-coding genes from their micronucleus to their macronucleus (Ehrenfeucht et al., 2004). Ciliates are unicellular eukaryotes having two types of nuclei: a macronucleus and a micronucleus (germline). The macronucleus, which is somatically active, is formed from the inactive micronucleus after sexual reproduction. It seems that ciliates have "invented" very ingenious solutions to complex problems million years ago.

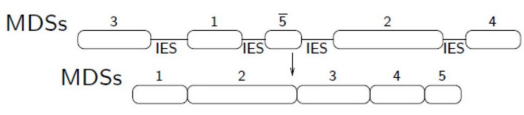


Figure 1: The arrangement of genes in the micronucleus (above) and their arrangement in the macronucleus (below).

A functional copy of the genes in the macronucleus is obtained by eliminating some DNA sequences, and/or changing some gene order, and/or reverting some inverted genes in the micronucleus. We illustrate this situation in Figure 1, where $\bar{5}$ means that the fifth gene is inverted. In this note, we consider the formal variants of three intramolecular operations which were proposed in (Ehrenfeucht et al., 2004) as formal operations for the unscrambled gene assembly in ciliates, more precisely in a group of ciliates called *Stichotrichs*. These operations called *ld, hi, dlad*, formally define the DNA recombination based on folds that are aligned by pointers, some relatively short nucleotide sequences at the intersection of MDSs (macronuclear destined sequences) and noncoding IESs (internally eliminated sequences). It is worth mentioning that the operations considered here are intramolecular operations; however, intermolecular operations have also been defined and investigated, see, e.g., (Kari & Rozenberg, 2008), (Brijder et al., 2012).

Informally speaking, we may say that what ciliates do in gene assembly is excision, inversion and interchange (sorting) of DNA sequences. Therefore, a hypothetical computation with ciliates would mean to encode the given instance of a problem in an abstract micronuclear gene, leaves the ciliate to assemble it into a macronucleus, and then filter and sequence the result. Initially, the formal computational power of gene assembly in ciliates was investigated in comparison with that of Turing machines: (Landweber & Kari, 1998), (Onolt-Ishdorj et al, 2007), yielding to computationally complete models. Later on, there were reported conceptual solutions to some hard problems, like the Hamiltonian Path Problem (Alhazov et al, 2008). Such a solution, if ever implemented in ciliates, would have the advantage that the organisms themselves implement generally time-consuming steps of the procedure.

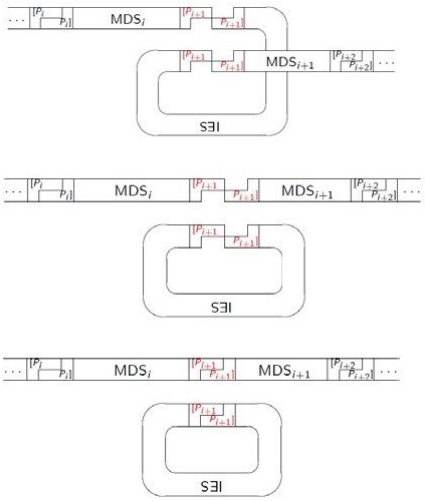


Figure 2: The ld operation.

We now explain the three operations. The operation *ld* excises one circular molecule, which contains noncoding blocks only. In this way, two consecutive MDSs are aggregated into a bigger composite MDS. More precisely, assume that the MDS_{i+1} follows MDS_i, the two being separated by an IES, and the pointer P_{i+1} flanks the IES. Then *ld* makes a fold aligned by pointer P_{i+1}, excises the IES as a circular molecule, and recombines MDS_i and MDS_{i+1} into a longer coding block, as in Figure 2.

In the case of hi, the DNA sequence makes a fold aligned by pointer P_{i+1}, reverts MDS_{i+1}, and rearrange the two MDSs in the correct order, as illustrated in Figure 3.

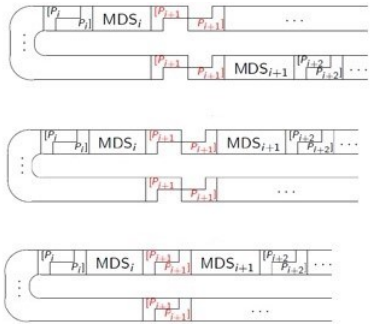


Figure 3: The hi operation.

Finally, in Figure 4, the $dlad$ is illustrated. The molecule is double folded aligned by the pointers P_{i+1} and P_{i+2}, and then a new molecule is obtained by the recombination sequence shown in Figure. 4.

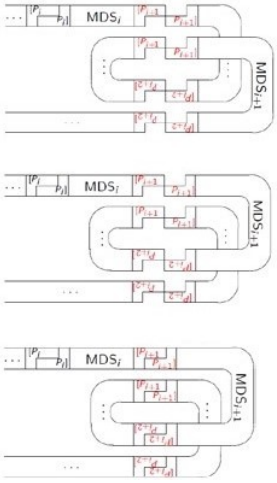


Figure 4: The $dlad$ operation.

We consider here three formal operations on words and language inspired by the operations shown in the three figures above. We first define the *gene assembly in ciliates* (gac)-scheme, the valid string with respect to a gac-scheme, and discuss the algorithm for deciding whether or not a given string is valid with respect to a given gac-scheme. We show that this algorithm is essentially based on the Aho-Corasick algorithm which can be implemented by means of a finite automaton

such that the language of all valid strings with respect to a gac-scheme is regular. Then we consider a few properties of the non-iterated versions of these operations. Thus, the ld operation can be simulated by a finite transducer while this is not possible for the other two operations. All the three operations preserves the property of a string of being valid. A natural and biologically motivated question is whether or not the order of applying these operations does matter. We solve completely the problem by showing that the order does not really matter.

The iterated variants of the three operations are defined and a few properties are investigated. A sort of a "normal form" concerning the order of applying these operations is given. Finally, a few open problems are discussed.

2 Preliminaries

We introduce here the main concepts and definitions which will be used in the rest of the paper. For undefined notions, the reader is referred to (Rozenberg et al., 2012).

The cardinality of a finite set A is denoted by $card(A)$. An alphabet V is a finite set of symbols/letters, a string over V is a finite sequence of symbols from V, the empty string is denoted by λ, the set of all strings over V is denoted V^*, while the set of all nonempty strings over V is denoted by V^+. Further on, the length of a string x is denoted by $|x|$. For a string $w \in V^*$, $Sub(w) = \{x \in V^+ \mid w = uxv, u, v \in V^*\}$ is the set of all substrings of w. If u or v is not empty in this definition, then x is a proper substring of w. Furthermore, w^R denotes the mirror image of the string w, that is $w^R = a_n a_{n-1} \dots a_1$, provided that $w = a_1 a_2 \dots a_n$, where each a_i, $1 \leq i \leq n$, is a symbol. For a set of strings (language) L we set $L^R = \{w^R \mid w \in L\}$.

A *finite transducer* is a 6-tuple $M = (Q, V, U, \delta, q_0, F)$, where Q, V, U are finite and non-empty sets called states set,input alphabet and output alphabet, respectively, $q_0 \in Q$ is the initial state, $F \subseteq Q$ is the set of final states, and δ is the transition-and-output function from $Q \times (V \cup \{\lambda\})$ to finite subsets of $Q \times U^*$. This function may be extended in a natural way to $Q \times V^*$. Each finite transducer M as above defines a *finite transduction*

$$g_M(x) = \{y \in U^* | (q, y) \in \delta(q_0, x), q \in F\}.$$

The finite transduction g_M can be extended to a language $L \subseteq V^*$ as follows

$$g_M(L) = \bigcup_{x \in L} g_M(x).$$

If M reads exactly one symbol at each transition and U is ignored (i.e. M has no output), we have a *finite automaton*. Moreover, M is called deterministic, if δ is a function from $Q \times V$ into $Q \times U^*$.

A gac-scheme (gene assembly in ciliates) is a pair $\sigma = (V, P)$, where V is an alphabet and P is a finite subset of V^+, whose elements are called pointers, such that $P^R = P$.

We now introduce the three formal operations on strings suggested by the gene assembly in ciliates illustrated in Figures 2, 3, and 4.

Given a gac-scheme $\sigma = (V, P)$ and a string $w \in V^+$, we formally define the ld operation as follows.

$$\begin{aligned} ld_\sigma(w) &= \{x\alpha z \mid w = x\alpha y\alpha z,\ x, z \in V^*, \\ &\quad y \in V^+,\ \alpha \in P, Sub(y) \cap P = \emptyset\}. \end{aligned}$$

We consider here that the circular molecule obtained in Figure 2 is no longer the micronuclear precursor for another assembly, so we keep only the linear molecule. Note that, as mentioned before, the segment y must not contain any pointer.

Given a gac-scheme $\sigma = (V, P)$ and a string $w \in V^+$, we formally define the hi operation as follows.

$$\begin{aligned} hi_\sigma(w) &= \{x\alpha y^R \alpha^R z \mid w = x\alpha y\alpha^R z, x, z \in V^*, \\ &\quad y \in V^+, \alpha \in P\}. \end{aligned}$$

The $dlad$ operation is applicable to those strings which have an alternating direct repeat pair of pointers. This string can be folded into two loops, each of which is aligned by one pair of pointers, as illustrated in Figure 4. Formally, given a gac-scheme $\sigma = (V, P)$ we define

$$\begin{aligned} dlad_\sigma(w) &= \{u\alpha y\beta x\alpha v\beta z \mid w = u\alpha v\beta x\alpha y\beta z, \\ &\quad v, x, y \in V^+, u, z \in V^*, \alpha, \beta \in P\}. \end{aligned}$$

It is worth noting that simpler variants have been considered in (Harju et al., 2006), too. This means the followings: in the case of hi and dld each fold, and consequently each rearranged sequence, contains just one MDS. It is obvious that there are sequences such that no simple operation is applicable while the variants considered here can be applied. For instance, take the string

$$w = x_0 \underline{32} x_1 \underline{10} x_2 \underline{67} x_3 \underline{89} x_4 \underline{45} x_5 \underline{67} x_6 \underline{23} x_7 \underline{45} x_8,$$

where all the pointers are underlined, and none of the substrings x_i, $0 \leq i \leq 8$ contains any digit. Clearly, w is a valid string but no simple operation is applicable to this string. On the other hand, both hi and dld can be applied. Indeed,

$$x_0 \underline{32} x_6^R \underline{76} x_5^R \underline{54} x_4^R \underline{98} x_3^R \underline{76} x_2^R \underline{01} x_1^R \underline{23} x_7 \underline{45} x_8 \in hi(w),$$

and

$$x_0 \underline{32} x_1 \underline{10} x_2 \underline{67} x_6 \underline{23} x_7 \underline{45} x_5 \underline{67} x_3 \underline{89} x_4 \underline{45} x_8 \in dld(w).$$

3 Results

Given a *gac*-scheme $\sigma = (V, P)$, we say that a string $w \in V^*$ is *valid* with respect to σ, if w admits a unique decomposition as follows:

$$w = w_0\alpha_1 w_1 \alpha_2 \dots w_{k-1}\alpha_k w_k,$$

with $k \geq 0, \alpha_i \in P, 1 \leq i \leq k$, and $Sub(w_j) \cap P = \emptyset$, for all $0 \leq j \leq k$.

This definition appears also in (Vaszil, 2020).

Theorem 1. *Given a gac-scheme $\sigma = (V, P)$, one can algorithmically decide whether or not a given string $w \in V^*$ is valid in $\mathcal{O}(n + m)$, where $n = |w|$ and $m = \sum_{\alpha \in P} |\alpha|$.*

Proof. The algorithm is essentially the algorithm proposed by Aho and Corasick in (Aho & Corasick, 1975) for finding all occurrences of a finite set of strings in a given string. We actually use the algorithm for finding all occurrences of the pointers from P in w. The following conditions must be satisfied in order to validate w:

- At any position i in w, there is at most one pointer that can appear. In other words, there are no two pointers that appear at the same position in w.
- If a pointer α appears at the position i in w, and j is the next position in w, where a pointer appears, $j > i + |\alpha| - 1$.

Note that the complexity of the algorithm is the same to that of Aho-Corasick without the number of the output matches. □

As the Aho-Corasick is based on the construction of a special type of a finite automaton, we may also state:

Corollary 1. *Given a gac-scheme $\sigma = (V, P)$, the language of all valid strings over V with respect to σ is regular.*

3.1 Properties of the Non-iterated Versions

Theorem 2. *Let $\sigma = (V, P)$ be a gac-scheme.*
1. There is a finite transducer M such that $ld_\sigma(w) = g_M(w)$ for every $w \in V^*$.
2. There is no finite transducer M such that $hi_\sigma(w) = g_M(w)$ for every $w \in V^*$.
3. There is no finite transducer M such that $dlad_\sigma(w) = g_M(w)$ for every $w \in V^*$.

Proof. 1. Instead of giving the formal construction of the finite transducer M, we prefer to explain how it works on the input string w. Its computations have four phases:

- *Phase 1.* The transducer scans a prefix of w and writes each symbol read. In a nondeterministic way, it passes to Phase 2.
- *Phase 2.* Now, M checks whether a pointer follows in w and memorize it. If there are more pointers that appears in w at the same position (note that w is not necessarily valid), M chooses one nondeterministically. The pointer is printed out. Then it passes to Phase 3.
- *Phase 3.* Now M checks the presence of the second occurrence of the memorized pointer. If the pointer does not appear anymore, the computation is blocked. If the pointer appears, it checks also the presence of another pointer that does not overlap with the occurrences of the memorized symbol. If this is the case, the computation is blocked again. During this Phase, no symbol is written to output. If the checking process has been successful, the transducer passes to Phase 4, the last one.
- *Phase 4.* M writes the rest of the input string (after the second occurrence of the memorized pointer) to the output.

By these explanations, it immediately follows that $ld_\sigma(w) = g_M(w)$.

2. Assume by contradiction that there exists a finite transducer M such that $hi_\sigma(w) = g_M(w)$ for every $w \in V^*$. Let us consider the linear language $L = \{wcw^Rc \mid w \in \{a, b\}^+\}$ and the *gac*-scheme $\sigma = (V, \{c\})$. By our assumption, $hi_\sigma(L) = \bigcup_{w \in L} hi_\sigma(w) = \{wcw \mid w \in \{a, b\}^+\}$. This implies that $g_M(L) = \{wcw \mid w \in \{a, b\}^+\}$. But this is not possible because the language $\{wcw \mid w \in \{a, b\}^+\}$ is not context-free and the class of context-free languages is closed under finite transducer mappings.

3. A similar argument may be used for the third statement. We consider the context-free language $L = \{a^n cb^n cbcb^m ca^m \mid n, m \geq 1\}$ and the *gac*-scheme $\sigma = (\{a, b, c\}, \{c\})$. Thus, we get $dlad_\sigma(L) = \{a^n cb^m cb^{n+1} a^m \mid n, m \geq 1\}$, which is not context-free anymore. □

As an immediate consequence, it follows that the class of regular languages is closed under the operation ld, while the class of context-free languages is closed under none of the other two operations. Note that a similar result is also proved in (Freund et al., 2002).

One important question is: Which of the three operations preserves the validity property? The next statement answer this question.

Theorem 3. *Let $\sigma = (V, P)$ be a gac-scheme. If w is valid with respect to σ, then each of the sets $ld_\sigma(w)$, $hi_\sigma(w)$, and $dlad_\sigma(w)$ contains valid strings only.*

Proof. If w is a valid string with $w = x\alpha y\alpha z$, and $Sub(y) \cap P = \emptyset$, then the string u produced by ld, that is $u = x\alpha z$, is also valid. Indeed, if two pointers were overlapping in the prefix $x\alpha$ of u, they would overlap in the same prefix of w, which is not possible. The situation is similar if two pointers were overlapping in

the suffix αz of u, hence u is valid with respect to σ. It follows that ld preserves the validity property.

We now assume that $w = x\alpha y\alpha^R z$ and prove that $u = x\alpha y^R\alpha^R z$ is also valid. Assume the contrary; we distinguish three cases which are to be analyzed in the sequel. Let us assume that $\alpha = \gamma\delta$ and $y = y_1y_2$, such that δy_2^R is a pointer. By the definition of P, $y_2\delta$ must be also a pointer. But this is not possible as w is valid.

Now, let us assume that $\alpha = \gamma\delta$ and $y = y_1y_2$, such that $y_1^R\delta^R$ is a pointer. This implies that δy_1 is a pointer and this contradicts again the validity of w.

Finally, let $\alpha = \gamma\delta$ and $y = y_1y_2$, such that $\delta y^R\delta^R$ is a pointer. It follows that $\delta y\delta^R$ is also a pointer which is also a contradiction. Consequently, the hi operation preserves the validity property.

We now assume that $w = u\alpha v\beta x\alpha y\beta z$ and prove that $u\alpha y\beta x\alpha v\beta z$ is also valid. This immediately follows as soon as we notice that the underlined segments $\underline{u\alpha v\beta x}\underline{\alpha y\beta} z$ in w are interchanged by the $dlad$ operations. □

As we have seen above, during the process of obtaining its macronucleus, the ciliate removes some DNA sequences, and/or change some gene order, and/or reverts some inverted genes in the micronucleus. A natural question is whether or not the order of this evolutionary steps is important. More precisely, is it possible to obtain the same sequences if the order of applying the three operations is changed? This is solved by the following statement.

Theorem 4. *Let $\sigma = (V, P)$ be a gac-scheme, and x be a valid string in V^+. Then the followings hold:*
1. $ld_\sigma(hi_\sigma(x)) = hi_\sigma(ld_\sigma(x))$.
2. $ld_\sigma(dld_\sigma(x)) = dld_\sigma(ld_\sigma(x))$.
3. There are valid strings y such that $dld_\sigma(hi_\sigma(y)) \neq hi_\sigma(dld_\sigma(y))$.

Proof. 1. Let $z \in hi_\sigma(ld_\sigma(x))$; this means that there exists $y \in ld_\sigma(x)$ such that $z \in hi_\sigma(y)$. It follows that $y = x_0\alpha x_2$, provided that $x = x_0\alpha x_1\alpha x_2$, with $\alpha \in P$ and $Sub(x_1) \cap P = \emptyset$. We distinguish three cases depending on the substring of y that is modified by the hi operation:

Case (i). The modified substring is $x_0\alpha$. Two subcases must be considered: (i1) $x_0 = u\beta v\beta^R q$, $\beta \in P$, and (i2) $x_0 = u\alpha^R v$ (note that x is a valid string). In the subcase (i1), $z = u\beta v^R\beta^R q\alpha x_2$ holds. We infer that $z \in ld_\sigma(y')$, where $y' = u\beta v^R\beta^R q\alpha x_1\alpha x_2$, hence $y' \in hi_\sigma(x)$. In the subcase (i2), $z = u\alpha^R v^R\alpha x_2$ holds. We now infer that $z \in ld_\sigma(y')$, where $y' = u\alpha^R v^R\alpha x_1\alpha x_2$, hence $y' \in hi_\sigma(x)$.

Case (ii). The modified substring is αx_2. This may be treated analogously to Case (i).

Case (iii). The pointer α is a proper substring, that is neither a prefix nor a suffix, of the modified string. Formally, $x_0 = u\beta v, x_2 = q\beta^R t$, for some u, v, q, t. It follows that $z = u\beta q^R\alpha^R v^R\beta^R t$. We deduce that $z \in ld_\sigma(y')$, where $y' = u\beta q^R\alpha^R x_1^R\alpha^R v^R\beta^R t$, therefore $y' \in hi_\sigma(x)$. Thus, we have proved that $ld_\sigma(hi_\sigma(x)) \supseteq hi_\sigma(ld_\sigma(x))$ holds.

For the converse inclusion, let $z \in ld_\sigma(hi_\sigma(x))$. There exists $y \in hi_\sigma(x)$ such that $z \in ld_\sigma(y)$. We may consider that $y = x_1\beta x_2^R\beta^R x_3$, provided that $x = x_1\beta x_2\beta^R x_3$. It is easy to notice that the substrings of y that can be modified by the ld operation are: (i) $x_1\beta$, (ii) βx_2^R, (iii) $x_2^R\beta^R$, and (iv) $\beta^R x_3$. By the definition of ld which imposes that the loop does not contain any pointer, we immediately conclude that z can be obtained by applying first the ld operation to the substrings of x: $x_1\beta$, βx_2, $x_2\beta^R$, and $\beta^R x_3$, respectively, and then the hi operation guided by the pointers β and β^R.

2. Let $q \in ld_\sigma(dld_\sigma(x))$. We consider that $x = u\alpha v\beta w\alpha y\beta z$; there exists $t = u\alpha y\beta w\alpha v\beta z \in dld_\sigma(x)$ such that $q \in ld_\sigma(t)$. The possible site for the application of the ld operation to t could be one of the following substrings (i) $u\alpha$, (ii) $\alpha y\beta$, (iii) $\beta w\alpha$, (iv) the sub$\alpha v\beta$, and (v) βz. No matter which site is chosen, the two operations can be interchanged. It follows that $ld_\sigma(dld_\sigma(x)) \subseteq dld_\sigma(ld_\sigma(x))$.

Now, let $q \in dld_\sigma(ld_\sigma(x))$. We consider that $x = u\alpha v\alpha w$; there exists $t = u\alpha w \in ld_\sigma(x)$ such that $q \in dld_\sigma(t)$. Obviously, by the definition of ld, any application of dld to t can be applied to the same substring of x with just one exception, namely the substring that includes α, when the substring of x includes $\alpha v\alpha$. Consequently, $ld_\sigma(dld_\sigma(x)) = dld_\sigma(ld_\sigma(x))$ holds.

3. We take the string $y = a12a34a21a56a34a56a$, where the substrings of digits represent pointers. On the one hand, $t = a12a34a56a34a21a56a \in dld_\sigma(x)$. Actually, t is the only string in $dld_\sigma(x)$. Then, $hi_\sigma(t)$ is also a singleton that contains $q = a12a43a65a43a21a56a$. Therefore, $q \in hi_\sigma(dld_\sigma(y)$. On the other hand, the unique string in $hi_\sigma(y)$ is $r = a12a43a21a56a34a56a$, but $dld_\sigma(r) = \emptyset$. Now the proof is complete. □

Clearly, the first two statements of this theorem are valid for the simple variants. If we restrict to the simple variants, the equality $dld_\sigma(hi_\sigma(y)) = hi_\sigma(dld_\sigma(y))$ holds for any valid string y.

3.2 Properties of the Iterated Versions

In this section, all the strings considered are valid strings. Given a *gac*-scheme $\sigma = (V, P)$, and $\tau \in \{ld_\sigma, hi_\sigma, dlad_\sigma\}$ we define:

$$(i) \quad \tau^0(w) = \{w\},$$
$$(ii) \quad \tau^{i+1}(w) = \bigcup_{x \in \tau^i(w)} \tau(x),$$
$$(iii) \quad \tau^*(w) = \bigcup_{i \geq 0} \tau^i(w).$$

Given a gac-scheme $\sigma = (V, P)$, we say that a string w is an ld_σ-macronuclear string if $ld^*_\sigma(w) = \{w\}$. The set of all ld_σ-macronuclear strings evolved from a string x is denoted by $mld_\sigma(x)$ and is defined by

$$mld_\sigma(x) = \{w \mid w \in ld^*_\sigma(x) \text{ and } w \text{ is } ld_\sigma - macronuclear.\}$$

An ld_σ-macronuclear string may be viewed as a formal description of a protein-coding gene structure in a chromosome of a ciliate macronucleus. Before sexual reproduction the genomic copies of these genes in the micronucleus are separated by the presence of internally eliminated non-protein-coding DNA sequences which must be removed during sexual reproduction.

Theorem 5. *Let $\sigma = (V, P)$ be a gac-scheme. There exists a finite transducer M such that $mld_\sigma(w) = g_M(w)$ for every string w over V.*

Proof. We consider a finite transducer M that implements the following nondeterminstic algorithm.
Step 1. M finds the first occurrence of a pointer in P, say α. All symbols read up to the occurrence of α as well as α are left unchanged. The pointer α is memorized in the current state.
Step 2. Nondeterministically, M chooses one of the next two ways for continuing the computation:
Step 2.1. It "guesses" that the next pointer is also α. Then it deletes all the symbols of the input string from the current position of α up to the next one, as well as the found occurrence of α. If its guess was wrong, the computation is blocked. Note that M can do this as it saved α. Now, M goes back to *Step 2.*.
Step 2.2. M "guesses" that the next pointer is different than α. Then M leaves unchanged all the read symbols until the next occurrence of a pointer is met. If this pointer is α, then the computation is blocked. If it is not α, M writes the new pointer on the output tape, replace α in the memory by the new pointer and goes back to *Step 2.*

The finite transducer defined above actually applies all the possible iterations of ld_σ to w. □

As the set of all valid strings is regular, by the previous theorem it follows that

Corollary 2. *For any gac-scheme $\sigma = (V, P)$, the set of all ld_σ-macronuclear strings is regular.*

Given a gac-scheme $\sigma = (V, P)$, and a string $w \in V^*$, we define

$$(hi/dlad)_\sigma(w) = hi_\sigma(w) \cup dlad_\sigma(w),$$
$$gac_\sigma(w) = (hi/dlad)_\sigma(w) \cup ld_\sigma(w),$$

and iterate these operations as follows

$$(i) \quad \tau^0_\sigma(w) = \{w\},$$
$$(ii) \quad \tau^{i+1}_\sigma(w) = \bigcup_{x \in \tau^i_\sigma(w)} \tau_\sigma(x),$$
$$(iii) \quad \tau^*_\sigma(w) = \bigcup_{i \geq 0} \tau^i_\sigma(w),$$

where $\tau \in \{gac, (hi/dlad)\}$. As a consequence of Theorem 4, we have this result as a *normal form* of the gene assembly in ciliates.

Theorem 6. *Given a gac-scheme $\sigma = (V, P)$, and a string $w \in V^*$, $gac^*_\sigma(w) = ld^*_\sigma(hi/dlad)^*_\sigma(w)) = (hi/dlad)^*_\sigma(ld^*_\sigma(w))$ holds.*

Proof. Obviously, both inclusions

$$ld^*_\sigma(hi/dlad)^*_\sigma(w)) \subseteq gac^*_\sigma(w)$$

and

$$(hi/dlad)^*_\sigma(ld^*_\sigma(w)) \subseteq gac^*_\sigma(w)$$

hold.

We now prove the inclusion $gac^*_\sigma(w) \subseteq ld^*_\sigma(hi/dlad)^*_\sigma(w))$, while the other inclusion, namely $gac^*_\sigma(w) = (hi/dlad)^*_\sigma(ld^*_\sigma(w))$ may be proved analogously. The argument is the induction on the number of gac steps. More precisely, let $z \in gac^k_\sigma(w)$, for some $k \geq 0$. If $k = 0$, then $z = w$, hence $z \in ld^0_\sigma(hi/dlad)^0_\sigma(w))$. Assume that $gac^k_\sigma(w) \subseteq ld^*_\sigma(hi/dlad)^*_\sigma(w))$ holds, and take $z \in gac^{k+1}_\sigma(w)$. This implies that $z \in gac_\sigma(y)$ for some $y \in gac^k_\sigma(w)$. By the induction hypothesis, $y \in ld^*_\sigma(hi/dlad)^*_\sigma(w))$. We distinguish three cases depending on the operation applied to y in order to get z: ld, hi, and $dlad$.

If $z \in ld_\sigma(y)$, then $z \in ld_\sigma(ld^*_\sigma(hi/dlad)^*_\sigma(w))) \subseteq ld^*_\sigma(hi/dlad)^*_\sigma(w))$.

If $z \in hi_\sigma(y)$, then $z \in hi_\sigma(ld^*_\sigma(hi/dlad)^*_\sigma(w)))$. This means that $z \in hi_\sigma(ld^q_\sigma(hi/dlad)^r_\sigma(w)))$, for some $q, r \geq 0$, but not both being null. If $q = 0$, we are done. Otherwise, by Theorem 4, we can interchange q times the operations hi and ld from left to right and yield that $z \in ld^q_\sigma(hi/dlad)^{r+1}_\sigma(w))$.

As the operations ld and $dlad$ can also be interchaged by Theorem 4, the case $z \in dlad_\sigma(y)$ is treated analogously. □

This theorem says that instead of iteratively applying the three operations in an arbitrary order, one can iteratively apply first the ld operation, which decrease the length of the string, and then the other two operations.

4 Final Remarks

We have reconsidered the three formal operations on strings suggested by the intramolecular gene assembly in ciliates. Properties of both non-iterated and iterated variants have been investigated. We shortly discuss here a few other properties that appear to be interesting to us and were not considered in this paper.

Note that the operations hi and $dlad$ are reversible (but not the ld operation), that is for any gac-scheme $\sigma = (V, P)$ and $\tau \in \{hi_\sigma, dlad_\sigma\}$, $x \in \tau(w)$ if and and only if $w \in \tau(x)$. This leads to the following definition of an equivalence relation $\sim_\tau$:

$$x \sim_\tau y \text{ if } x \in \tau^*(y).$$

It immediately follows that the relation is indeed an equivalence relation, being reflexive, symmetric and transitive.

Note that the problem "Does $x \in ld^*_\sigma(y)$ holds, for any given strings x and y?" is decidable in linear time. Indeed, by scanning both strings x and y, when a mismatch is met, try to apply the ld operation to remove the mismatch. If this is possible, continue the scanning process, otherwise the answer is negative.

Proposition 1. *Let $\sigma = (V, P)$ be a gac-scheme and $\tau \in \{hi_\sigma, dlad_\sigma\}$.*
1. The problem of whether $x \sim_\tau y$, for two strings $x, y \in V^+$, is decidable.
2. If hi and dld are simple, then the problem of whether $x \sim_\tau y$, for two strings $x, y \in V^+$, is decidable in linear time.

The proof of the first statement is immediate as any of the equivalence classes defined above are finite. The second statement follows from the fact that a single simultaneous scan of x and y (as that explained above for ld) suffices for taking the decision.

Here a few open problems naturally arise: Is it a more time efficient algorithm for solving the first statement, than the brute force one? What is its complexity? Have both problems (for hi and $dlad$) the same complexity?

Given $\sigma = (V, P)$ a gac-scheme, $x, y \in V^*$, and $\tau \in \{ld_\sigma, (hi/dlad)_\sigma, gac_\sigma\}$ defined above, we define the partial mapping

$$d_\tau(x, y) = \begin{cases} min\{k \mid x \in \tau^k(y) \text{ or } y \in \tau^k(x)\}, \\ \quad \text{if such a } k \text{ exists}, \\ \text{not defined, otherwise.} \end{cases}$$

Clearly, d_τ is computable for any $\tau \in \{ld_\sigma, (hi/dlad)_\sigma, gac_\sigma\}$. While d_{ld_σ} is computable in linear time, the complexity of computing the other functions remains an open problem.

Acknowledgements

This work was performed through the Core Program within the National Research, Development and Innovation Plan 2022-2027, carried out with the support of MRID, project no. 23020101(SIA-PRO), contract no 7N/2022, and project no. 23020301(SAFE-MAPS), contract no 7N/2022.

References

Aho A.V. & Corasick, M.J. (1975). Efficient String Matching: An Aid to Bibliographic Search, *Communications of the ACM*, 18(6), 333–340.

Alhazov A, Petre I, & Rogojin V. (2008). Solutions to Computational Problems Through Gene Assembly. *Nat. Comput.* 7(3), 385–401.

Brijder, R., Daley, M., Harju, T., Jonoska, N., Petre, I., & Rozenberg, G. (2012). Computational Nature of Gene Assembly in Ciliates. In: Rozenberg, G., Bäck, T., Kok, J.N. (Eds) Handbook of Natural Computing. Springer, Berlin, Heidelberg.

Ehrenfeucht, A., Harju, T., Petre, I., Prescott, D.M., & Rozenberg, G. (2004). *Computation in Living Cells. Gene Assembly in Ciliates*, Springer-Verlag Berlin, Heidelberg, New York.

Freund, R., Martín-Vide, C., & Mitrana, V. (2002). On Some Operations on Strings Suggested by Gene Assembly in Ciliates, *New Gener. Comput.* 20(3), 279-294.

Harju, T., Petre, I., Rogojin, V., & Rozenberg, G. (2006). Simple Operations for Gene Assembly. In A. Carbone & N.A. Pierce (Eds.) *Proceedings of the 11^{th} Workshop of DNA Computing (DNA 11)* (pp. 96-111). Springer-Verlag Berlin Heidelberg.

Kari, L. & Rozenberg, G. (2008). The Many Facets of Natural Computing, *Communications of the ACM* 51(10), 72–83.

Landweber, L.F., & Kari, L. (2002). Universal Molecular Computation in Ciliates. In: L.F. Landweber & Winfree, E. (Eds.) *Evolution as Computation* (257–274). Springer, New York.

Onolt-Ishdorj, T., Petre I, & Rogojin V. (2007). Computational Power of Intramolecular Gene Assembly. *Int. J. Found. Comp. Sci.* 18(5), 1123–1136.

Păun, G., Rozenberg, G., & Salomaa, A. (1998). *DNA Computing. New Computing Paradigms*, Springer-Verlag Berlin, Heidelberg, New York.

Rozenberg, G., Bäck, T., & Kok, J.N. (2012). *Handbook of Natural Computing*, Springer-Verlag Berlin, Heidelberg.

Vaszil, G. (2020). Personal communication.

Design of system for collecting and analyzing of human gait parameters

Juraj Ďud'ák, Gabriel Gašpar
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava
Jana Bottu 2781/25, Trnava, Slovakia
University of Zilina
Reasearch Centre
Univerzitna 8215/1, Zilina, Slovakia
{juraj.dudak, gabriel.gaspar}@stuba.sk

Jakub Perička
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava
Jana Bottu 2781/25, Trnava, Slovakia
jakub.pericka@stuba.sk

Eva Vangelová
University of Zilina
Reasearch Centre
Univerzitna 8215/1, Zilina, Slovakia
eva.vangelova@uniza.sk

Daniel Gurín
Slovak Medical University in Bratislava
Faculty of Health Sciences
Sladkovicova 21, 974 05 Banska Bystrica, Slovakia
daniel.gurin@szu.sk

Abstract. *Human gait parameters reveal much information about a person's possible health problems. These parameters can be used for the early diagnosis of neurogenerative diseases and to monitor the healing process during rehabilitation after an injury. The methods of monitoring gait parameters can be divided into two basic approaches. The first approach is to monitor the movement parameters of the lower extremities, tracking with the help of a camera system or with the help of accelerometers placed on the patient's body. The second approach is to monitor the loading of individual leg parts during walking. This paper will present the design of a wearable device for monitoring gait parameters, where several Inertial Measurement Units (each containing an accelerometer, gyroscope, and magnetometer) will be used to process and evaluate the data. The presented design will consist of a microcontroller module, 1 to 3 IMU modules (accelerometer+gyroscope), and pressure sensors placed in the shoe's sole.*

Keywords. gait analysis, sensory system, microcontroller, accelerometer

1 Introduction

Walking is a complex activity that requires a high level of coordination between the central nervous system and the body. Currently, the direction of objective data acquisition for the kinematic analysis of human movement in the physiological or pathological plane is focused on commonly used devices. However, implementing sensors in wearable smart devices does not guarantee their use in all situations. Therefore, it appears appropriate to implement sensors for kinematic analysis or monitoring of vital signs for active rehabilitation into adjuvants. Active rehabilitation or active forms of rehabilitation are those in which the patient actively realizes the chosen treatment and does not simply passively receive treatments. The data collected during this process will be processed by multiple independent methods. This paper presents the design of a technical device to acquire inertial data during the rehabilitation or convalescence process. The overall system design consists of a wearable device with a sensor base and communication interface, a software part for data acquisition and a model for data evaluation. The output of this model will be information about the gait parameters according to the specification of a rehabilitation expert. In developing the proposed system, the authors followed proven hardware design practices (Dudak et al., 2011). The aim of the present project is to detect selected gait parameters (such as leg off center or excessive knee joint rotation) from the acquired data. Both analytical and machine learning methods

will be used to obtain these data (Dudak et al., 2020).

1.1 Systems for inertial data collection and analysis

There are several articles regarding the design of rehabilitation aids, which implement sensor equipment or are connected in some way to the Internet. Monica Tiboni et al. (Tiboni et al., 2022) produced an extensive analysis of 215 publications on the monitoring of the health status of different body parts, the sensing technologies, and the sensors used. The study was published in 2022 and analyses publications since 1975. However, relevant publications have existed only since 2006. Up to 57% of the analyzed publications deal with lower limb rehabilitation, with the ratio of publications about rehabilitation aid vs. assistive devices being approximately the same. According to this analysis, pressure sensors and Inertial Measurement Units (IMU) are used most widely for the lower limbs, to a lesser extent than force sensors, EEG, and camera systems. The results of this analysis overlap with the topic of this paper - the use of several types of sensors for monitoring selected parameters of human gait in the rehabilitation process. IMU sensors are used in multiple papers. A separate category of wearable devices is clothing with installed sensors. These sensors are most often IMU units, which can obtain information about the nature of the movement being performed or the orientation of the human body or limb. In (Wang et al., 2015), a garment application is presented in the form of a vest, with which the authors monitored a patient's posture.

Among the methods for gait data evaluation, neural networks are widely used. The advantage of using the machine learning algorithm is the ease of use for monitoring specific target parameters. As an example, we refer to (Kantoch, 2017), where the authors could discern a patient's activity level by monitoring heart rate and blood oxygenation. In (Prochazka et al., 2018, p. 1209), an application of machine learning algorithms is presented to determine the relationship between the level of physical exertion of the monitored subject and its heart rate. Machine learning algorithms are also suitable for recognizing human hand gestures. Geng Yang et al. created a wristband application (Yang et al., 2018) that detects which forearm muscles are used and can recognize the hand gesture (open/closed palm, thumbs up, raise index finger, and other gestures). The fundamental task in projects of this type is the selection of sensors suitable for sensing the monitored parameters. In human gait analysis, it is possible to focus on dynamic parameters such as stride length, stride rate, and others, for which the IMU sensors are most commonly used. In (Gujarathi and Bhole, 2019), the principle of using IMU sensors for gait monitoring is presented. The authors presented the basic design of the measuring device and algorithms for estimating stride length, standing time, and time of limb movement during walking. For the calculations, they used a basic mathematical model. Time domain analysis was also used by Wang et al. in (Wang et al., 2016), where they used 2 IMU sensors placed on the thigh and at the base of the shin. The use of machine learning algorithms in evaluating gait parameters is described by Nguyen et al. (Nguyen et al., 2020), where they used 7 IMU sensors (3 for each leg and one in the pelvic region) and used a convolutional neural network for evaluation. A similar methodology is also used in (Mundt, 2020), where the authors used an analytical motion model as the first method and, as the second step, created a neural network model. In (Sunarya, 2020), the authors Sunarya et al. present an extensive feature analysis of smart shoe sensor data, including pressure sensors, accelerometer, and gyroscope signals. They also investigated the optimal length of data segmentation based on gait cycle parameters. The results show that the highest accuracy (88.44% - 89.76%) was achieved by the data fusion of the accelerometer and gyroscope data and a support vector machine model (SVM).

The second interest group is the parameters of body weight distribution on the foot surface in contact with the walking surface. In this case, several types of sensors can be used, such as piezoelectric (Chandel et al., 2019), pneumatic (Kong et al., 2009), or resistive (Paredes-Madrid et al., 2017).

This paper deals with designing a wearable system for collecting sensor data from various sources. IMU sensors placed in the lower extremities will be used as the primary method. Resistive sensors for measuring pressure in different parts of the foot in contact with the walking surface will be used as a second data source.

Since the authors of the article have established cooperation with the Slovak Medical University, Faculty of Public Health, specifically with the Department of Physiotherapy, the presented data collection system emerged as an intersection of expertise from technical and health fields. The result of joint discussions was the design of a system that would be able to sense human gait parameters in two modes: a) during rehabilitation under the supervision of a specialist, and b) during home rehabilitation. In this paper, we present a redesign of the solution, where the primary focus is on use in the supervised rehabilitation mode. This paper presents an initial design of the system and will be optimized in the future with respect to its ease of use. The system will also include a mathematical model that will identify and evaluate parameters related to possible health defects.

2 The design of the wearable system for data acquisition

The proposed system consists of a controller unit which incorporates a microcontroller to which the sensors are connected, a communication sub-module, and software

for data collection and analysis. For easy attachment of the sensors to the lower extremity, a mechanical design is proposed for the placement of sensors in the form of wearable electronics.

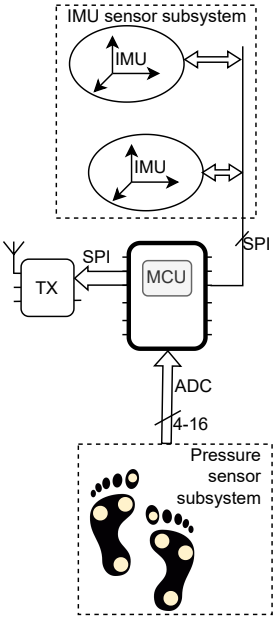


Figure 1: Block diagram of the proposed system

Figure 1 shows the main block diagram of the design. It is an application for collecting data from a single lower limb. This design consists of an STM32L432 microcontroller, NRF24L01 wireless data transmitter, and sensor modules:

- IMU sensor submodule. It contains 1 to 4 IMU sensors placed on the lower limb of the monitored person. In the initial design, we expect to place two sensors: one sensor will be attached to the calf and the other above the knee. The communication interface used is SPI.
- Pressure sensor submodule. It contains a pressure sensor module for one foot. The number of sensors in one module can range from 3 to 16. Since the physical principle of the sensor works with the change of electrical resistance depending on the pressure exerted on the sensor, the A/D converter interface of the MCU unit is used for the data measurement. Due to the magnitude of the measured electrical resistance change, each pressure sensor will be connected to a voltage divider circuit.

During the design of the module, the following requirements were defined:

1. IMU module
   (a) Sampling frequency min. 200Hz
   (b) Measurement range: 2-16g for accelerometer, 250-2000 $deg * s^{-1}$ for gyroscope
   (c) Resolution: 16 bit
   (d) Integrated low-pass filter
2. Pressure sensor module:
   (a) Number of sensors: min. 4 with proper placement of sensing areas
   (b) Sensor sensitivity suitable for the distribution of human body weight over the sensor area
3. Communication module
   (a) Communication speed: min. 1 Mb/s

3 Hardware design

IMU sensor module. For the IMU module, we have selected the ICM20948 sensor. This sensor contains a 3-axis accelerometer, 3-axis gyroscope, 3-axis compass, and a DMP (Digital Motion Processor) module that allows computations directly on the sensor module. It includes I2C and SPI communication interfaces with a maximum clock frequency of 7MHz for SPI and 400kHz for I2C. The accelerometer range is configurable from +/-2g to +/-16g. The gyroscope range is from +/-250deg*s-1 to 2000deg*s-1. According to the specification (TDK InvenSense, 2021), this sensor is suitable for wearable and IoT applications. The sensor allows for the maximum data acquisition frequency of 562Hz. The resolution for each axis is 16 bits. Additionally, the IMU includes a digital lowpass filter from 197Hz to 5.7Hz for the gyroscope and from 246Hz to 5.7Hz for the accelerometer.

Pressure sensor module. This module consists of multiple FSR sensors placed on a pad (Figure 2). A Force Sensing Resistor (FSR) is a two-terminal passive device that exhibits a dramatic decrease in its electrical resistance when stress is exerted over the Sensor Sensitive Area (SSA). The SSA is usually made from conductive particles randomly dispersed in a nonconductive polymer, resulting in a polymer composite; the SSA is later sandwiched between two metal electrodes to allow electrical measurements; the resulting device is an FSR (Paredes-Madrid et al., 2017). The parameters of the FSR sensor are:

- Dynamic sensing range spans from 0.2 N to 20 N
- Able to withstand millions of actuations
- No-Load resistance > 10 $M\Omega$
- Hysteresis: +5% (FSR datasheet)

MCU control module. The control module is implemented on the STM32L432 microcontroller, which belongs to the reduced power consumption family of STM32 microcontrollers. The created firmware works

Figure 2: Module containing FSR sensors

in two modes: configuration mode, where the connected sensors can be configured (range, digital low-pass filter, and sample rate), and the second mode is the measurement, during which the module operates autonomously, and the measured data are sent over the wireless interface. The principal algorithm is shown in Figure 3. The "IMU ready" and "RX data" blocks represent hardware interrupts from the IMU sensors and the communication interface, respectively. In the "measurement" mode, when the IMU sensors are active, a hardware interrupt is generated according to the sensor settings (5Hz to 562Hz). The meaning of this interrupt is a notification that the data are ready to read.

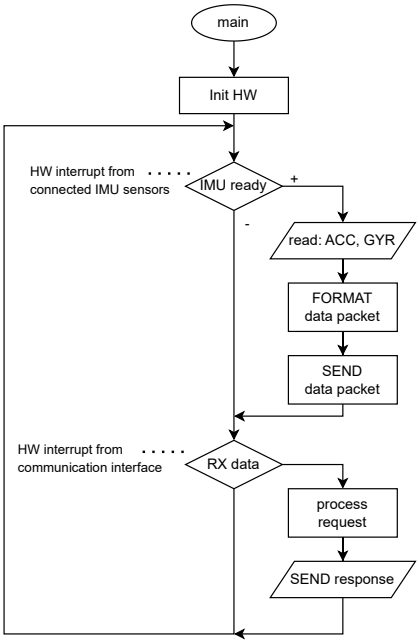


Figure 3: Principal algorithm for the communication with peripherals connected to the MCU

In the initial design phase, we selected the sampling rate of the IMU to be 280Hz. Data from both IMU sensors (Figure 6) and from 8 pressure sensors are available for each measurement. In terms of data intensity, there is a need to send data of at least 64B in one measurement. The structure of the data packet that is sent from the MCU module is shown in Figure 4.

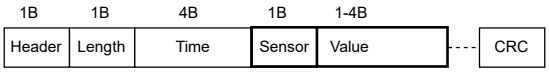


Figure 4: Format of the measurement data packet

The communication packet begins with a header, followed by the total packet length information. A mandatory part of the packet is the measurement timestamp. The time value is relative, and the timer is restarted at the beginning of the measurement process. The resolution of the time base is 0.1 ms. The next part is the data from the connected sensors. Each sensor has a label on which data can be correctly identified during data processing (Figure 4, Sensor section). This byte is followed by the value read from the sensor (value). This part contains data from all 3 axes for the accelerometer and gyroscope.

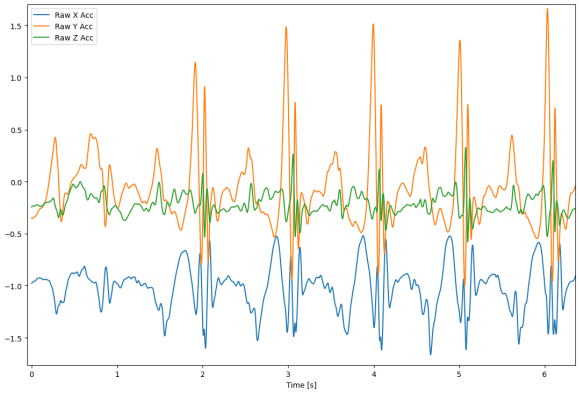


Figure 5: Result form Ambulatio software - data visualization

For data collection and evaluation, the software "Ambulatio" is being developed to configure the parameters of the IMU sensors, and also, in this software, basic visualization of the measured data is possible. In Figure 5 is a preview of the result of the "Ambulatio" software. The graph shows the data visualization from the IMU1 accelerometer in all 3 axes.

4 Conclusions

The embedded sensor system (Figure 1) forms the basis of a comprehensive data collection and analysis system in the recovery and rehabilitation process of the patient. The design presented in this paper forms the basis of the system, its hardware part, which is responsible for collecting and transmitting measured data. In the current version of the implementation, 2 IMU sensors are connected to a control module with a microcontroller.

Figure 6 shows the first prototype of this system. The IMU sensors are placed in a housing designed for this application. This housing design assumes that it will be attached to the lower limb of the monitored person - the shape of the housing is adapted to this. In Figure 6, these covers are marked with a yellow tag. In the current version of this system, a cover for the microcontroller is also included. In the next iteration of development, the MCU module is planned to merge with the IMU sensor module.

Figure 6: Ambulatio system prototype - IMU sensor placement

Using of system. In planning the functionality of the human gait monitoring system, it will be necessary to create a working prototype of a device that would monitor the parameters of both lower limbs. From a tactical standpoint, eliminate unnecessary cabling so that the device falls into the category of wearable electronics. Next, develop models to identify key parameters from the sensed data. To identify situations where persons with certain motor impairments are monitored. Tasks related to the last point will be addressed jointly with a rehabilitation expert. The first users of the proposed human gait monitoring system will be patients in the rehabilitation process. The patient will be fitted with 2 sensors on each lower extremity using straps (Figure 6). The aim is that these modules do not affect the patient's movement abilities in any way. 3D printing was used to create the mechanical design of the prototype hardware modules. In the next version, it is envisaged to minimize the dimensions and overall weight of the entire module. Currently, one module is approximately 200g. In the planned stage of the solution (rehabilitation under expert supervision), the system will be used for data collection and subsequently to confirm or modify the conclusions of the rehabilitation expert.

In the next step of system development, we will focus on the analysis of the collected data. The goal is to find suitable methods to analyze the measured data for later online analysis of qualitative gait parameters. We expect that both analytical and neural network models will be developed. Using these models, it will be easy to identify wrong walking habits or wrong habits for patient rehabilitation.

Acknowledgments

This paper was supported under the project of Operational Programme Integrated Infrastructure: Independent research and development of technological kits based on wearable electronics products, as tools for raising hygienic standards in a society exposed to the virus causing the COVID-19 disease, ITMS2014+ code 313011ASK8. The project was cofunded by the European Regional Development Fund.

References

Q. Wang, W. Chen, A. A. A. Timmermans, C. Karachristos, J. B. Martens, and P. Markopoulos, Smart Rehabilitation Garment for posture monitoring. In 2015 *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, pp. 5736-5739. doi: 10.1109/EMBC.2015.7319695.

E. Kantoch (2017). Human activity recognition for physical rehabilitation using wearable sensors fusion and artificial neural networks. In *Computing in Cardiology (CinC)*, Sep. 2017, pp. 1-4. doi: 10.22489/CinC.2017.296-332.

G. Yang et al. An IoT-Enabled Stroke Rehabilitation System Based on Smart Wearable Armband and Machine Learning. (2018) In: *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1-10, doi: 10.1109/JTEHM.2018.2822681.

A. Prochazka, H. Charvatova, S. Vaseghi, and O. Vysata. Machine Learning in Rehabilitation Assessment for Thermal and Heart Rate Data Processing. (2018) In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 6, pp. 1209-1214, doi: 10.1109/TNSRE.2018.2831444.

Gujarathi and K. Bhole (2019) GAIT ANALYSIS USING IMU SENSOR. In: *10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, pp. 1-5, doi: 10.1109/ICCCNT45670.2019.8944545.

X. Wang, M. Kyrarini, D. Ristic-Durrant, M. Spranger and A. Graser (2016), Monitoring of gait performance using dynamic time warping on IMU-sensor data. In: *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Benevento, Italy, pp. 1-6, doi: 10.1109/MeMeA.2016.7533745.

M. D. Nguyen et al. (2020), IMU-based Spectrogram Approach with Deep Convolutional Neural Networks for Gait Classification. In: *IEEE*

International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, pp. 1-6, doi: 10.1109/ICCE46568.2020.9042999.

M. Tiboni, A. Borboni, F. Verite, C. Bregoli, and C. Amici (2022), Sensors and Actuation Technologies in Exoskeletons: A Review, In: *Sensors, vol. 22*, no. 3, 2022, doi: 10.3390/s22030884.

Mundt M, Koeppe A, David S, Witter T, Bamer F, Potthast W and Markert B (2020), Estimation of Gait Mechanics Based on Simulated and Measured IMU Data Using an Artificial Neural Network. In: *Bioeng. Biotechnol*. 8:41. doi: 10.3389/fbioe.2020.00041

V. Chandel, S. Singhal, V. Sharma, N. Ahmed and A. Ghose (2019), PI-Sole: A Low-Cost Solution for Gait Monitoring Using Off-The-Shelf Piezoelectric Sensors and IMU, In: *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, pp. 3290-3296, doi: 10.1109/EMBC.2019.8857877.

K. Kong and M. Tomizuka (2009), A Gait Monitoring System Based on Air Pressure Sensors Embedded in a Shoe, In: *IEEE/ASME Transactions on Mechatronics*, vol. 14, no. 3, pp. 358-370, doi: 10.1109/TMECH.2008.2008803.

Chen, J.-L., Dai, Y.-N., Grimaldi, N. S., Lin, J.-J., Hu, B.-Y., Wu, Y.-F., Gao, S. (2022), Plantar Pressure-Based Insole Gait Monitoring Techniques for Diseases Monitoring and Analysis: A Review. In: *Adv. Mater. Technol.*, DOI: 10.1002/admt.202100566

TDK InvenSense: ISCM-20948 (2021), *Datasheet*. Document Number: DS-000189, InvenSense, a TDK Group Company. url: https://invensense.tdk.com/download-pdf/icm-20948-datasheet/

Paredes-Madrid, L.; Palacio, C.A.; Matute, A.; Parra Vargas, C.A. (2017), Underlying Physics of Conductive Polymer Composites and Force Sensing Resistors (FSRs) under Static Loading Conditions. In: *Sensors* 17, 2108. https://doi.org/10.3390/s17092108

FSR 400 Series Data Sheet. *Interlink Electronics*. Datasheet. url: https://www.interlinkelectronics.com/fsr-400

Nordic Semiconductors: nRF24L01+ Single Chip 2.4GHz Transceiver. *Datasheet*. url: https://datasheetspdf.com/pdf-file/906332/Nordic/nRF24L01+/1

Sunarya U, Sun Hariyani Y, Cho T, Roh J, Hyeong J, Sohn I, Kim S, Park C. (2020), Feature Analysis of Smart Shoe Sensors for Classification of Gait Patterns. In: *Sensors* (Basel). 20(21):6253. doi: 10.3390/s20216253.

Dudak J., Kebisek M., Gabriel. G, Fabo P. (2020), Implementation of machine learning algorithm in embedded devices, In: *19th International Conference on Mechatronics - Mechatronika (ME)*, DOI: 10.1109/ME49197.2020.9286705

Dudak J., Pavlikova S., Gaspar G., Kebisek M. (2011) Application of open source software on arm platform for data collection and processing. In: *Proceedings of 14th International Conference on Mechatronics, MECHATRONIKA 2011*, DOI: 10.1109/mechatron.2011.5961090

Deep Transfer Learning for Detecting Face Images with Moustache

Filip Babić, Petra Grd, Igor Tomičić, Ena Barčić
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42 000 Varaždin
fbabicucg@gmail.com, {petra.grd, itomicic, enbarcic}@foi.unizg.hr

Abstract. *Face detection is one of the most researched fields in deep learning, but some aspects have not yet been sufficiently researched. In this paper, we will focus on the application of deep transfer learning for detecting face images with moustache. For the purpose of this paper, we use a version of the MobileNetV2 convolutional neural network which was pretrained on the ImageNet dataset and the provided weights are used for transfer learning to moustache detection problem. In addition, we created a balanced dataset for moustache detection from FFHQ dataset. Using our approach we were able to achieve a testing F1-score of 89%.*

Keywords. face biometrics, neural networks, moustache detection, MobileNetV2, FFHQ dataset

1 Introduction

Deep transfer learning is a valuable technique in the field of computer vision that allows the transfer of knowledge learned from one task to another, and in recent years, it has been used to achieve state-of-the-art performance in various tasks, including object detection, image classification, and face recognition.

In this paper we will focus on the application of the deep transfer learning for detecting face images with moustache, where using deep transfer learning in such a context may have several advantages. Firstly, it can lead to improved accuracy as it leverages knowledge obtained from large datasets of face images, leading to improved performance in detecting face images with moustache. Secondly, it reduces training time since the model can be fine-tuned using a smaller dataset rather than being trained from scratch. Furthermore, deep transfer learning algorithms are designed to generalize well, resulting in improved generalization on new, untested data. They can also be engineered to be robust to variations in lighting, facial expressions, and camera viewpoint for example. Additionally, deep transfer learning algorithms require less computational resources compared to building a model from scratch. This makes them well-suited for deployment in resource-constrained environments, where efficiency is crucial. By leveraging pre-trained weights and knowledge from existing models, deep transfer learning minimizes the computational burden, enabling effective moustache detection on face images without excessive resource requirements.

Due to their cultural and historical relevance, moustaches in particular have been extensively studied as an important visual cue for gender classification and face recognition. Detecting face images with moustache can be challenging due to the variations in moustache shape, size, and color, and influenced by factors such as facial expressions, lighting, and camera viewpoint.

The further motivation for moustache detection can be found in various potential use-cases. For instance, it can enhance gender classification algorithms by detecting moustaches on face images, thus improving accuracy. Moustaches can also significantly impact face recognition performance, making moustache detection relevant in the field. In surveillance systems, identifying individuals based on their facial hair through moustache detection can be valuable in settings like security checkpoints, border control, and criminal investigations. Moustache detection also holds potential for personalized experiences, such as within augmented reality applications. Moreover, analyzing trends in facial hair and targeting specific demographics based on their facial hair preferences can be advantageous in marketing and advertising endeavors.

In this paper, we will present a state of the art research efforts within the field, and propose a solution by applying a deep transfer learning for detecting moustache on face images.

Section 2 will provide a related work overview; Section 3 will illustrate the network architecture of the proposed model; Section 4 will provide more detailed information on the used dataset, training, validation and testing protocol, and demonstrate and discuss achieved results. Section 5 will conclude and discuss the implications of the research and possible future work.

2 Related work

When talking about facial hair in the context of face recognition we are most commonly talking about two separate fields of research: facial hair detection and facial hair segmentation. Facial hair detection, as the name implies focuses on detecting facial hair in images of individuals or group images (Yang et al., 2018).

The authors in (Yang et al., 2018) present a DCNN for face detection leveraging facial attributes-based supervision called Faceness-Net. Some of the attributes they focus on are beard, no beard, goatee, 5 o'clock shadow, mustache, and sideburns. They use the Labeled Faces in the Wild (LFW) dataset (Huang et al., 2007) for training, and chose to combine the evaluation to one category 'hair+beard'. They use cropped images achieving a detection accuracy of 95.56% and uncropped images achieving a detection accuracy of 94.57%. They conclude that hair, eye, and nose parts perform much better than the mouth and beard.

In (Gudi, 2016) the authors focus on recognizing semantic features in faces using deep learning. They test the effects of various factors and hyper-parameters of deep neural networks for an optimal network configuration that can accurately recognize semantic facial features like emotions, age, gender, ethnicity, etc. In their research, they include 'add-on' features like glasses and facial hair (beard and moustache). The baseline architecture relies on a ReLu model and is trained and tested on the VV Dataset ("Vicarious Perception Technologies" 2023) and Facial Expression Recognition (FERC) Dataset (Goodfellow et al., 2015). The features of beards and moustaches are further classified into three categories consisting of non, light, or heavy. The authors achieve a final accuracy of 88.1% for the beards and 89.13% for the moustaches. They conclude that their network's precision for detecting glasses and heavy moustaches is very high, but the network does not learn very precise light beards and no beard classifiers.

In (Hand and Chellappa, 2016) the authors present a multi-task network for attribute classification. They present attribute relationships in three ways: using a multi-task deep convolutional neural network (MCNN) sharing the lowest layers amongst all attributes, sharing the higher layers for related attributes, and building an auxiliary network on top of the MCNN which utilizes the scores from all attributes to improve the final classification of each attribute. In addition, the authors have separated facial hair into five categories consisting of 5 o'clock shadow, moustache, no beard, sideburns, and goatee, using the CelebA (Liu et al., 2015) and LFWA (Huang et al., 2007) datasets for training and testing. They conclude that on the CelebA dataset using all three approaches the best results were achieved on the sideburns samples achieving 97.77% on the Independent CNN, 97.82% on the MCNN, and 97.85% on the MCNN-AUX, while the lowest results were achieved on the 5 o'clock Shadow categories achieving 93.94% on the Independent CNN, 94.41% on the MCNN, and 94.51% on the MCNN-AUX. On the LFWA dataset using all three approaches the best results were achieved on the moustache samples achieving 93.69% on the Independent CNN, 93.53% on the MCNN, and 93.43% on the MCNN-AUX, while the lowest results were achieved on the 5 o'clock Shadow categories achieving 77.39% on the Independent CNN, 77.70% on the MCNN, and 77.06% on the MCNN-AUX. On the LFWA dataset.

The authors in (Sghaier and Elfaki, 2021) present an efficient technique for human face occlusion detection and extraction. They extract three types of occlusions using artificial intelligence techniques consisting of a Viola-Jones algorithm based on Vision Cascade Detector and combined two performance methods: fuzzy C-means method and filters and morphological operations. This is done to boost the efficiency of face recognition systems. They divide occlusions into three categories: beard, moustache, and eyeglasses and use the Faces94 ("Face Recognition Data" n.d.) and CMU-PIE (Gross et al., 2008) datasets. The authors were able to achieve 100% accuracy with both the FCM method and Morphological operations on the Faces94 dataset, and a 99% accuracy with both on the CMU-PIE dataset.

Other approaches are more focused on facial hair segmentation such as the authors of (Le, Luu, Seshadri, et al., 2012) which present a novel system for beard and moustache detection and segmentation in challenging facial images. They use pre-processes images using the self-quotient algorithm and then use a sparse classifier to detect and segment regions containing facial hair from the MBGC, color FERET databases (Phillips et al., 1997). Facial hair is classified into two categories: beard and moustache. The authors are able to achieve 96.2% accuracy on beards, and 98.8% accuracy on moustaches on the MBGC dataset and 95.8% accuracy on beards, and 97.0% accuracy on moustaches on the color FERET dataset.

Some approaches combine both facial hair detection and facial hair segmentation like in (Ubaid et al., 2022) where the authors present an efficient state-of-the-art system for beard and hair detection and segmentation for changing color in challenging facial images based on a modified version of a Mask R-CNN model. They create their own dataset of 1 500 images which are equally divided into both hair and beard images. In the experiment phase, the authors were able to achieve a final accuracy of 91.2%.

The authors in (Le, Luu, Zhu, et al., 2017) present a robust, fully automatic, and semi-self-training system to detect and segment facial beard/moustache simultaneously in challenging facial images. The authors use the superpixel together with a combination of two classifiers, Random Ferns (rFerns) and Support Vector Machines (SVM) to obtain good classification performance as well as improve time efficiency.

They divide facial hair into three categories: facial hair, non-facial hair (male), and non-facial hair (female). In the end, they propose a facial hair detection algorithm and test it on the Multiple Biometric Grand Challenge (MBGC) ("Multiple Biometric Grand Challenge (MBGC) Presentations" 2010) and FERET (Phillips et al., 1997). They achieve the highest accuracy on the MBGC dataset with an accuracy of 88.5% for facial hair, 91% for non-facial hair (male), and 90.8% for non-facial hair (female), and the lowest accuracy on the FERET dataset with 85.1% for facial hair and 84.5% for non-facial hair (female), and the lowest accuracy for non-facial hair (male) was achieved on the PINELLAS dataset with 84.2%.

Among different deep learning models, MobileNetV2 architecture has emerged as a choice in different computer vision tasks such as face recognition, face mask detection and similar. Leveraging its streamlined structure, MobileNetV2 demonstrates exceptional prowess in accurately and rapidly discerning facial features, while optimizing the resources necessary for training and testing. The authors in (Almghraby and Elnady*, 2021) use mobilenetv2 and transfer learning for real-time mask detection. They used stochastic gradient descent (SGD) for optimization with 12 epochs in each experiment, a learning rate of 0.001, and a momentum of 0.85. They were able to achieve 99% training and 98% validation accuracy. Similarly the authors in (Kumar and Bansal, 2023) present their approach for mask detection on photos and video images using Caffe-MobileNetV2 transfer learning. They used the MobileNetV2 mask identification and added five different layers to the pre-trained MobileNetV2 architecture. They achieved an accuracy of 99.64% on photos.

The authors in (Huu et al., 2022) propose a model capable of distinguishing between masked and nonmasked faces using deep learning (DL)—MobileNetV2. The authors used Retina Face as their face detector model, next they used MaskTheFace to modify their dataset and finally they trained MobileNetV2. They were able to achieve an accuracy of up to 99.37%. In (Narduzzi et al., 2022) the authors focus on the adaptation of MobileNetV2 for the purpose of face detection on ultra-low power platforms. They used Tiny-Face as a starting point and computed a set of canonical bounding boxes derived by clustering them into 25 different sizes. Detection results were obtained after applying non-maximum suppression (NMS). MobileNetV2 was adapted to a fully convolutional architecture, and two breakpoints and three outputs were used. They concluded that MobileNetV2 can successfully be applied as an alternative to classical machine learning methods for embedded face detection.

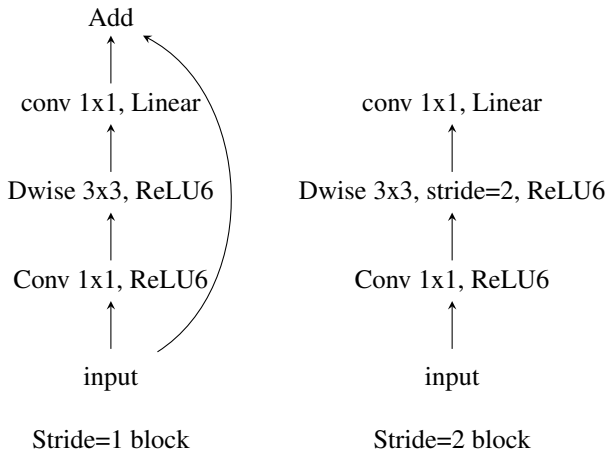


Figure 1: Convolutional blocks of MobileNetV2 (Sandler et al., 2018)

3 Network Architecture

The artificial neural network architecture used in the process of this research is a version of the MobileNetV2 convolutional neural network (Sandler et al., 2018). This particular architecture was chosen for its provided benefit in terms of low resource consumption during the training and testing processes. Lower resource consumption is the result of a trade-off in the manner of a reduced number of input parameters, leading to a higher efficiency score.

The basic building block of MobileNetV2 is "a bottleneck depth-separable convolution with residuals" (Sandler et al., 2018) (illustrated in Figure 1). The first layer of the model is a fully convolutional layer with 32 filters. 19 bottleneck residual layers follow it. The most prominently used activation function is ReLU6 (Howard et al., 2017), a modified rectified linear unit with the constrained maximum activation size of 6, primarily for the robustness it provides. A constant expansion rate is present in every layer, excluding the first layer. The utilised optimizer was AdamOptimizer (Table 1).

Table 1: CNN architecture hyperparameters

Hyperparameter	Value
Batch size	32
Activation function	ReLu6 and SoftMax
Loss function	binary cross entropy
Optimizer	AdamOptimizer
Learning rate	0.0001
Dropout	0.6

4 Experiments

4.1 Dataset

The dataset selected for this research is Flickr-Faces-HQ Dataset. This image dataset exhibits a substantial variation in terms of the faces it comprises. It encompasses individuals from different age groups, ethnicities, and backgrounds, thus offering a diverse representation. Moreover, the dataset also contains numerous images featuring faces adorned with accessories like glasses, sunglasses, hats, caps, and similar items. The presence of such accessories adds an additional level of complexity and diversity to the dataset, making it suitable for studying and developing algorithms that can effectively handle and recognize various facial attributes and accessories. The images are taken from the Flickr page and automatically aligned and cropped (Karras et al., 2018). The dataset contains more than 70 000 images with different resolutions. In this paper, a resolution of 128x128 pixels was chosen because it proved to be sufficient for the goal of the paper and manageable using available computer resources.

The selected dataset does not have labels for moustaches in images, so in order to train the neural network, the dataset has been manually divided into two parts. The first part consists of images without moustaches and the second part of 6 954 images with moustaches (Figure 2). For images without moustaches, a randomly selected subset of 7 060 images without moustaches was selected to create a balanced dataset.

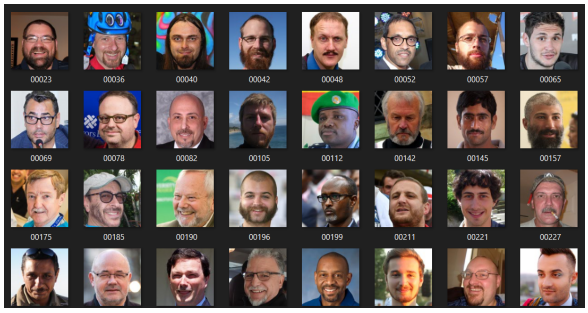


Figure 2: Examples of images with moustaches

The end results is the dataset with two subfolders corresponding to the classes necessary, one with moustache (49,62%) and one without moustache (50,38%).

4.2 Training and Validation

In order to train and later test the results of the network for moustache detection, the holdout method is used where 80% of images in each class is used for training and 20% for testing. The network was pretrained on ImageNet dataset and the provided weights are used for transfer learning to moustache detection. The lower part of the network is freezed and only the top of the network is trained with the training part of the dataset described in previous section.

To further increase the accuracy and reduce losses, artificial data augmentation is used to increases the diversity of the data available for model training without having to collect new data. Image cropping, rotating, shearing and horizontal flipping is used, which resulted with 17 500 images in the training set.

The network was trained on the machine whose main features are: processor - AMD Ryzen 7 3800X, graphics card - Nvidia GTX 1060 6 GB and RAM - 32 GB. CNNs were trained using a graphics card. Libraries on which neural networks are based and which are the basis of this paper are OpenCV (OpenCV, 2022), Keras (*Keras* 2022), TensorFlow (*TensorFlow* 2022) and scikit-learn (*scikit-learn* 2022).

In order to train the network, different hyperparameter values, shown in Table 1, have been utilized and the best results have been obtained by using learning rate of 0.0001, batch size of 32, dropout rate of 0.6 and training on 100 epochs. The training and validation results can be seen in Figure 3.

Figure 3: Training Loss and Accuracy

4.3 Testing and Results

As mentioned in the previous section, 20% of the dataset is used to test the model performance, while 80% of dataset has been used for training and validation. During the network validation, the accuracy was calculated, but accuracy is not always sufficient, especially if we are dealing with an unbalanced dataset. To this end, a comprehensive testing and creation of a confusion matrix has been done from which Precision, Recall, Accuracy and F1-score measures were calculated.

Information concerning actual and predicted classifications made by a classification system is contained in the confusion matrix. The data in the matrix is frequently used to evaluate the performance of such systems.

Precision is the proportion of positive cases that were correctly identified (Pedregosa et al., 2011):

$$Precision = \frac{TP}{TP + FP}. \quad (1)$$

Table 2: General Confusion matrix (Pedregosa et al., 2011)

	Predict Class 1	Predict Class 2
Actual Class 1	TP	FN
Actual Class 2	FP	TN

Recall is the proportion of actual positive cases which are correctly identified (Pedregosa et al., 2011):

$$Recall = \frac{TP}{TP + FN}. \quad (2)$$

Accuracy is the proportion of all cases that were correctly identified (Pedregosa et al., 2011):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3)$$

F1-score is a harmonic mean of Precision and Recall, calculated as (Pedregosa et al., 2011):

$$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall}. \quad (4)$$

Table 3 shows the testing results for the proposed approach and the examples of moustache detection can be seen in Figure 4.

Table 3: Test results

	Precision	Recall	F1-score
Moustache	0.87	0.90	0.89
No moustache	0.91	0.87	0.89
Overall Accuracy			0.89

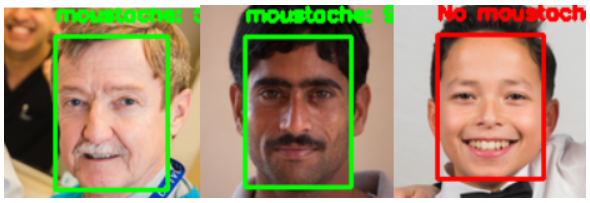


Figure 4: Examples of moustache detection

The results achieved within the context of this study cannot be directly compared with prevailing state-of-the-art achievements. This divergence emanates primarily from the datasets employed and the categories used for classification. Most datasets used in literature are private datasets or parts of the publicly available dataset are used without specification which parts or how many images. It is essential to underscore that a definitive, universally acknowledged benchmark dataset for facilitating direct comparisons of moustache detection outcomes, has not yet been established within the scientific community. Presently, a conspicuous gap exists in the domain of moustache detection evaluations, largely due to the scarcity of a standardized dataset that engenders a level playing field for comprehensive result assessment. An inherent predicament pertains to the prevalent datasets, often characterized by a notable class imbalance, wherein the number of non-moustache images significantly outweighs their moustache-bearing counterparts. This skewed distribution invariably renders conventional accuracy metrics insufficient for performance evaluation.

5 Conclusion

In this paper, we have provided an overview of the current state of the art in the field of moustache detection. Additionally, we have presented a solution that applies deep transfer learning techniques for detecting face images with moustaches. We have employed a version of the MobileNetV2 convolutional neural network and Flickr-Faces-HQ Dataset with a large variation of faces of different ages, ethnicities, backgrounds, and faces with accessories. We used the holdout method for training and testing our model. Specifically, we divided the dataset into two parts: 80% of the images from each class were utilized for training, while the remaining 20% were reserved for testing and evaluating the performance of the model. The network was pretrained on ImageNet dataset, and the weights provided were used in transfer learning for moustache detection.

For this research, a new balanced dataset specifically designed for moustache detection was created. This dataset was used to train and test the model in order to evaluate its performance. The evaluation of the model's performance was reported in Section 4, focusing on precision, recall, accuracy and F1-score as the performance metrics.

Deep transfer learning showed to be a promising technique for detecting moustaches on face images, where by leveraging the knowledge learned from large datasets of face images, it could improve the accuracy of moustache detection, even in challenging conditions. Future research in this area should focus on improving the accuracy and robustness of deep transfer learning algorithms for moustache detection, as well as exploring new applications of moustache detection in various fields.

Acknowledgments

This research was funded by the project "Development of CSTI platform for retrieval and analysis of structured and unstructured data". The project received funding from the European Regional Development Fund through OP Competitiveness and Cohesion 2014–2020 within the Call for Proposals "Development of the products and services arising from re-

search and development activities—Phase II", under grant number KK.01.2.1.02.0310.

References

Almghraby, M., & Elnady*, A. O. (2021). Face Mask Detection in Real-Time using MobileNetv2. *International Journal of Engineering and Advanced Technology*, *10*(6), 104–108. https://doi.org/10.35940/ijeat.F3050.0810621

Face Recognition Data. (n.d.). https://cmp.felk.cvut.cz/spacelib/faces/faces94.html.

Goodfellow, I. J., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., ... Bengio, Y. (2015). Challenges in representation learning: A report on three machine learning contests [Special Issue on "Deep Learning of Representations"]. *Neural Networks*, *64*, 59–63. https://doi.org/https://doi.org/10.1016/j.neunet.2014.09.005

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2008). Multi-pie. *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, 1–8. https://doi.org/10.1109/AFGR.2008.4813399

Gudi, A. (2016). Recognizing semantic features in faces using deep learning. https://doi.org/10.48550/arXiv.1512.00743

Hand, E. M., & Chellappa, R. (2016). Attributes for improved attributes: A multi-task network for attribute classification. https://doi.org/10.48550/arXiv.1604.07360

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. https://doi.org/10.48550/ARXIV.1704.04861

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (tech. rep. No. 07-49). University of Massachusetts, Amherst.

Huu, P. N., Quang, V. T., Le Bao, C. N., & Minh, Q. T. (2022). Proposed Detection Face Model by MobileNetV2 Using Asian Data Set. *Journal of Electrical and Computer Engineering*, *2022*, e9984275. https://doi.org/10.1155/2022/9984275

Karras, T., Laine, S., & Aila, T. (2018). A style-based generator architecture for generative adversarial networks. https://doi.org/10.48550/ARXIV.1812.04948

Keras. (2022). https://keras.io/

Kumar, B. A., & Bansal, M. (2023). Face Mask Detection on Photo and Real-Time Video Images Using Caffe-MobileNetV2 Transfer Learning. *Applied Sciences*, *13*(2), 935. https://doi.org/10.3390/app13020935

Le, T. H. N., Luu, K., Seshadri, K., & Savvides, M. (2012). Beard and mustache segmentation using sparse classifiers on self-quotient images. *2012 19th IEEE International Conference on Image Processing*, 165–168. https://doi.org/10.1109/ICIP.2012.6466821

Le, T. H. N., Luu, K., Zhu, C., & Savvides, M. (2017). Semi self-training beard/moustache detection and segmentation simultaneously. *Image and Vision Computing*, *58*, 214–223. https://doi.org/10.1016/j.imavis.2016.07.009

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. *Proceedings of International Conference on Computer Vision (ICCV)*.

Multiple Biometric Grand Challenge (MBGC) Presentations. (2010). *NIST*
Last Modified: 2021-06-02T18:40-04:00.

Narduzzi, S., Türetken, E., Thiran, J.-P., & Dunbar, L. A. (2022). Adaptation of MobileNetV2 for Face Detection on Ultra-Low Power Platform. *2022 9th Swiss Conference on Data Science (SDS)*, 1–6. https://doi.org/10.1109/SDS54800.2022.00008

OpenCV. (2022). *Opencv*. https://opencv.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, *12*(null), 2825–2830.

Phillips, P., Moon, H., Rauss, P., & Rizvi, S. (1997). The feret evaluation methodology for face-recognition algorithms. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 137–143. https://doi.org/10.1109/CVPR.1997.609311

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. https://doi.org/10.48550/ARXIV.1801.04381

Scikit-learn. (2022). https://scikit-learn.org/

Sghaier, S. M., & Elfaki, A. O. (2021). Efficient techniques for human face occlusions detection and extraction. *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*, 1–5. https://doi.org/10.1109/WiDSTaif52235.2021.9430214

Tensorflow. (2022). https://www.tensorflow.org/

Ubaid, M. T., Khalil, M., Khan, M. U. G., Saba, T., & Rehman, A. (2022). Beard and hair detection, segmentation and changing color using mask r-cnn. In A. Ullah, S. Anwar, Á. Rocha, & S. Gill (Eds.), *Proceedings of international conference on information technology and applications* (pp. 63–73). Springer Nature. https://doi.org/10.1007/978-981-16-7618-5_6

Vicarious Perception Technologies. (2023). *VicarVision*. https://vicarvision.nl/.

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2018). Faceness-net: Face detection through deep facial part responses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(8), 1845–1859. https://doi.org/10.1109/TPAMI.2017.2738644

Towards Sequence-to-Sequence Neural Model for Croatian Abstractive Summarization

Vlatka Davidović
University of Rijeka
Faculty of Informatics and Digital Technology
Radmile Matejčić 2, 51000 Rijeka
Business Department, Polytechnic of Rijeka
Trpimirova 2/V, 51000 Rijeka
vlatka.davidovic@uniri.hr

Sanda Martinčić Ipšić
University of Rijeka
Faculty of Informatics and Digital Technology
Center for Artificial Intelligence and Cybersecurity
Radmile Matejčić 2, 51000 Rijeka
smarti@uniri.hr

Abstract. *Abstractive text summarization is a natural language processing task of generating a summary from an input text while preserving the meaning of the text. Today, two prevalent deep learning architectures for automatic text summarization are sequence-to-sequence and transformer. The sequence-to-sequence architecture works with sequential data where the order of words in the text is important. The models have reached state-of-the-art performance in English, but the same task is challenging for low-resourced and morphologically rich languages. Moreover, there are few publicly available datasets prepared for training the summarization task. We focus here on Croatian, since the summarization dataset for Croatian is still missing. In this paper, we propose a solution to fill this gap. The first step is creating the summarization dataset in the Croatian language, using Google machine translation from English to Croatian. With the obtained Croatian version of the dataset, we perform initial training of the sequence-to-sequence model with an attention mechanism. The preliminary results of the Croatian abstractive summarization are presented using the evaluation metrics ROUGE and BERTScore.*

Keywords. abstractive summarization, sequence-to-sequence, Croatian

1 Introduction

The main goal of automatic text summarization is to generate a short summary of the input text without human intervention while preserving the main information and meaning (Allahyari et al. 2017). At the same time, the generated summary has to be fluent and similar to the abstract written by humans. The task can be defined as extractive or abstractive text summarization. While extractive text summarization can extract sentences from the original text, abstractive text summarization makes a summary by paraphrasing the main contents of the text with words that may not be in the original text (Widyassari et.al., 2022).

Datasets are an important part of any natural processing task, including automatic summarization, and they are used for training, validation, and testing of the models. Text summarization models map the long to a short text, so the dataset needs to be prepared as pairs of long text and its corresponding short summary. Automatic text summarization models based on deep learning architectures learn better on large datasets (Turc et al., 2019), so obtaining large and appropriate datasets is crucial for summarization tasks. Among published datasets, the most represented are in the English language. Furthermore, among English datasets, summarization research uses 55,65% of the publicly available dataset and 30,35% uses private datasets (Widyassari et.al., 2022).

The multilingual version of CNN/DailiyMail summarization dataset is a standardly used as the evaluation benchmark (Papers with Code, n.d.) available for five different languages (French, German, Spanish, Russian, Turkish). The Croatian translation of CNN/DailyMail is used here for the training of sequence-to-sequence abstractive summarization.

Existing Croatian neural models are trained on multilingual datasets. The CroSloEngualBERT model is trained in Croatian, Slovenian, and English languages (Ulčar & Šikonja, 2020) for NER, POS-tagging, and dependency parsing tasks. The model BERTić is trained in the Bosnian, Croatian, Montenegrin, and Serbian languages (Ljubešić & Lauc, 2021) for NER, POS-tagging, geo-location prediction, and commonsense causal reasoning tasks. The Context-aware Croatian Abusive Language (CoRAL) dataset (Shekhar et al., 2022) is used for detecting abusive language tasks. It contains user comments from one of the Croatian news portals (Shekhar et al., 2021). The Cro-CoV-cseBERT model automatically analyses the collected tweets in Croatian language, related to the communication about COVID-

19. Dataset is constructed of 10 000 manually annotated tweets (Babić et al., 2021).

The goal of this research is to obtain a dataset for training and testing the Croatian abstractive text summarization. For that purpose, a Croatian dataset is created using a machine-translated version of the well-known CNN/DailyMail English dataset. Then two different sequence-to-sequence models with attention mechanisms: long short-term memory (LSTM) and bidirectional long short-term memory (Bi-LSTM) are trained. The results of both models are evaluated with ROUGE 1, 2, L, and BERTSum evaluation metrics and compared with the same model trained in the original English language.

The paper is structured as follows: Section 2 overviews the sequence-to-sequence architecture and evaluation principles. The constructed dataset is elaborated on in Section 3, while Section 4 presents the results. The paper ends with a discussion.

2 Background architecture

2.1 Sequence-to-sequence

This work is based on a sequence-to-sequence with attention architecture (Sutskever et al., 2014). The architecture consists of encoder and decoder stacks of neural networks (Cho et al., 2014) that map input text to output text. Before the sequence is passed to the encoder, it is tokenized into words and converted into a word embedding. The output of the encoder is the context vector that goes to the decoder. The decoder predicts the next word based on information from the encoder and the previous predicted word. The attention mechanism additionally provides the decoder with information about relevant parts of the input sequence and the association between them (Fig.1).

Attention was originally applied to neural machine translation (NMT) tasks (Bahdanau et al., 2015), and then implemented in abstractive summarization (Rush et al., 2015; Nallapati et al., 2016). Attention enables focusing on the most relevant part of a source sentence where the information is concentrated, enabling the model to improve the prediction of the target word.

The neural networks in the encoder and decoder stack are layered on top of each other to provide better results (Suleiman & Awajan, 2020). They can utilize vanilla recurrent neural networks (RNN) (Nallapati et al., 2016), LSTM (Sutskever et al., 2014), gated recurrent unit (GRU) (Cho et al., 2014) or their bidirectional representatives (Kaichun et al., 2018; Prethee et al., 2022).

Analysis of the several approaches shows that recurrent neural networks with an attention mechanism and LSTM (Hochreiter & Schmidhuber, 1997; Cheng et al., 2016) are the most prevalent techniques for abstractive text summarization (Suleiman & Awajan, 2020).

While regular networks are trained in a forward direction, bidirectional networks (Schuster & Paliwal, 1997), are trained in both directions. Bi-LSTM (Graves et al., 2005) contains two LSTM layers: forward and backward LSTM, so they capture past and future information of the input word. The output of both layers is concatenated into functions such as average, sum, product or concatenation.

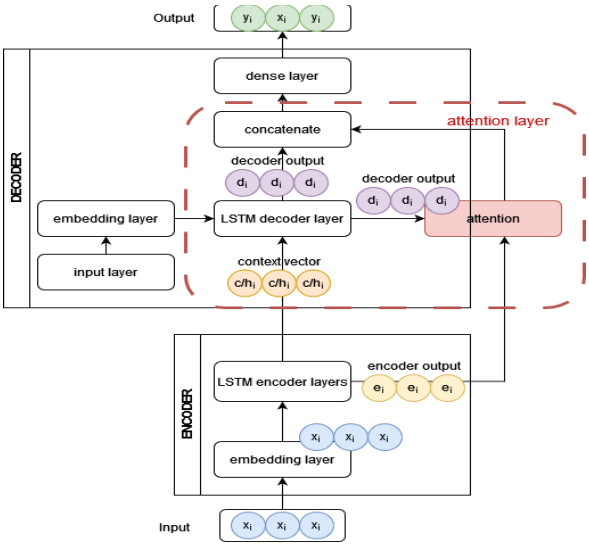


Figure 1. Encoder-decoder architecture with an attention mechanism (x_i=input vector, e_i=encoder hidden state, c/h_i= context vector, d_i=decoder output, y_i=output vector)

Transformers (Vaswani et al., 2017.) represent the other line of promising architectures for many text generation tasks. In this work, we limit the initial example to a sequence-to-sequence baseline, and we plan to train transformers in the following work.

2.2 Evaluation

The standard evaluation metric for summarization is Recall-Oriented Understudy for Gisting Evaluation (ROUGE) (Lin & Och, 2004). While ROUGE-N {N=1, 2, 3, …} calculates n-gram overlap between the machine-generated and reference summaries, ROUGE-L calculates the longest common subsequence (LCS) between the generated and reference summary so similar summaries have longer LCS. ROUGE and similar metrics correlate poorly with human evaluation (Novikova et al., 2017) because they do not consider the meaning and context of the evaluated text. BERTScore aims to overcome this drawback.

The BERTScore (Zhang et al., 2020) evaluation metric includes a contextual word embedding measure that calculates the similarity score between tokens in the original and the generated summaries. This metric is based on pre-trained BERT contextual embeddings (Devlin et al., 2019) and WordPiece tokenization of

input text (Wu et al., 2016). BERTScore uses multilingual WordPiece tokenization for non-English languages.

3 Dataset

Cable News Network (CNN)/DailyMail (Hermann et al., 2015; Nallapati et al., 2016) is an English-language dataset containing news articles, written by CNN and DailyMail journalists. The original version was created for the machine-reading and comprehension and question-answering tasks. One text in the dataset consists of id, article content, and highlights. *Id* is a string containing the hexadecimal formatted SHA1 hash of the URL where the story was retrieved from. *Article* is a string containing the text of the news article. *Highlights* is a string containing the summary of the article, written by the author of the article.

The dataset is divided into three separate parts: training, testing, and validation. The training subset contains 287,113; the test subset 11,490; and the validation subset 13,368 pairs of text and summary.

3.1 Croatian ML translated dataset

CNN/DailyMail dataset is commonly used for abstractive summarization tasks (Suleiman & Awaian, 2020). In this work, Google Machine Translation (GMT) is used to translate it into Croatian language (Google Cloud Translation, n.d.). The original division of the dataset into a training, testing and validation part is preserved and the resulting translations are prepared in JSON format, with UTF-8 encoding.

The analysis of the original and translated DailyMail/CNN dataset in English and Croatian languages is presented in Table 1.

Table 1. The average number of words in the text and summary in CNN/DailyMail in English and machine-translated Croatian datasets

CNN/DailyMail datasets	Avg. num. of words in the document	
	text	summary
English	659	49.9
Croatian MT	579.8	45.7

4 Experimental settings

In this work, we use a sequence-to-sequence architecture based on encoder-decoders with attention mechanisms. We trained two language models: one based on three layers of LSTM networks and another based on a Bi-LSTM network. Both models have an attention mechanism.

The LSTM and Bi-LSTM models were trained using the Croatian MT CNN/DailyMail dataset. Additionally, the LSTM model was also trained with the original English CNN/DailyMail dataset to compare the results

4.1 Preprocessing

Both Croatian MT and English datasets are initially preprocessed and tokenized. During preprocessing, non-alphabetic characters, multiple spaces, and single characters are removed from the training and validation text. Additional text cleaning is performed with the Python library SpaCy (Honnibal & Montani, 2017). There are available pipelines trained for specific languages. Pipeline *hr_core_news_lg* (Ljubešić et al., 2018) was trained on the Croatian language and *en_core_web_sm* (Weischedel et al., 2013; Choi & Palmer, 2012; Fellbaum, 2005) was trained on the English language.

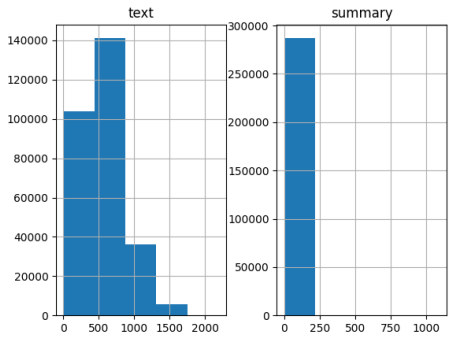


Figure 2. Distribution of words in DailyMail/CNN dataset. X-axis represents number of the words in the document. Y-axis represents number of the documents. Left: distribution in texts; Right: distribution in summary

During the preparation of the datasets for the training, we decided to set the maximum length of the texts to 800 words and summaries to 50 words. According to the word distribution (Fig.2) this seems to be an adequate threshold.

Table 2. Vocabulary size in texts and summaries

CNN/DailyMail datasets	Vocabulary size	
	text	summary
English	136,927	30,963
Croatian MT	305,163	64,445

Texts and summaries exceeding the maximum length are filtered out from the training dataset. Documents that have less than 800 words in text or 50 words in the summary are expanded by padding a zero up to the maximum length. After preparation, the English dataset contains 135,474 documents and the

Croatian dataset contains 171,107 documents that were used for training.

Two vocabularies are extracted per both datasets: one from the original texts and the other from the summaries. The vocabularies from the original text and the summary contain the most frequent words from the original text and the summary in both languages. The differences in the number of words in the vocabularies are displayed in Table 2. In contrast to the English, Croatian language has a rich morphology in which one word may have many forms (Seljan, 1999), and that explains the larger vocabulary of the Croatian dataset.

4.2 Language models

Two sequence-to-sequence models with attention mechanisms are compared. The encoder stack in the first LSTM model has three connected LSTM layers. The output of the last LSTM layer goes to the decoder side: the attention layer and the LSTM layer. The LSTM layer in the decoder gets a context vector from the encoder. The context vector is formed from the encoder hidden states and the attention weights. The outputs from the decoder's LSTM and attention layer go to the concatenation layer and from there to the last, dense layer. The dense layer has a softmax activation function that decides which token to generate next.

The encoder stack of the Bi-LSTM model has one Bi-LSTM layer. From this layer, outputs go to the decoder's concatenation layers and the attention layer. The concatenation layers merge inputs and pass them to the LSTM on the decoder side. The outputs from this layer and the attention layer are concatenated and passed to the dense layer to obtain the generated summary.

4.3 Training

Models were trained in the Tensorflow/Keras environment, using an AMD Ryzen Threadripper 3960X 24-Core Processor, CPU 2.2 GHz, and 256 GB of RAM. The hyperparameters for LSTM and Bi-LSTM models have the same values. The latent dimension was set to 300 and the embedding dimension to 200 tokens. The optimizer for models is RMSprop and the learning rate is set to 0.001. Details about the training are presented in Table 3.

For the Croatian dataset, the batch size is set to 250 and for the English dataset, the batch size is set to 400. A smaller batch size for the Croatian dataset was chosen because the model is trained with a large number of parameters and a larger batch size can exceed the memory limit.

Training is performed with maximum 50 epochs with an early stopping mechanism that is activated when the model does not improve results.

Table 3. Comparison of training data in different models and datasets (ES=early stopping)

Dataset and model	Num. of parameters	Batch size	Train. time	Num. of epochs
EN-LSTM	55,011,863	400	122h	33 (ES)
HR-LSTM	115,486,155	250	148h	26 (ES)
HR-Bi-LSTM	135,392,655	250	303h	50

5 Results

The models are tested during the inference phase with the testing subsets and evaluated using standard evaluation metrics ROUGE 1, 2, and L. Since the ROUGE metrics do not capture the semantics of the text, the BERTScore metric is used for additional evaluation. Table 4 shows the results of each model.

Table 4. Evaluation using ROUGE and BERTScore metrics

Dataset and model	Rouge 1 (F1)	Rouge 2 (F1)	Rouge L (F1)	BERTScore (F1)
EN-LSTM	20.29%	3.41%	15.10%	82.42%
HR-LSTM	16.91%	2.75%	12.58%	64.94%
HR - Bi-LSTM	18.71%	3.17%	13.22%	66.17%

The LSTM model trained on the Croatian MT dataset achieved the lowest results regardless of the used metrics. The Bi-LSTM model slightly improves the results over the LSTM baseline. Although the LSTM model on the English dataset shows better results compared to the models trained on the Croatian MT dataset, the overall results on ROUGE metrics are not impressive. In general, models achieve better results with BERTScore considering the content of texts. However, the BERTScore result on the Croatian MT dataset (64-66%) underperforms the English counterpart (82.42%). According to authors (Zhang et al., 2020), "for non-English language, the multilingual $BERT_{multi}$ is a suitable choice although BERTScore computed with this model has less stable performance on low-resource languages".

Despite modest summarization results, examples of the translated summary and the summary generated by LSTM and Bi-LSTM in the Croatian are given below:

Table 5. Examples of abstractive summarization with LSTM and Bi-LSTM models on the Croatian MT dataset

Translated summary	*pijani tinejdžer popeo se u nastambu s lavovima u zoološkom vrtu u zapadnoj indiji rahul kumar 17 potrčao je prema životinjama vičući ću ubiti lava srećom pao je u jarak prije nego što je stigao do lavova i spašen je*
Predicted summary LSTM	*slon je bio zoološkom vrtu za safari parku južnoj africi ali je pobjegao iz zoološkog vrta ali je preživio*
Predicted summary Bi-LSTM	*zoološkog vrta kini je životinje zoološkom vrtu central parku kina je upozorila da je životinja na putu za dobrotvorne svrhe ali je umro nakon što je pobjegao iz zoološkog vrta*
Translated summary	*reanne evans suočila se s kenom dohertyjem u kvalifikacijama za svjetsko prvenstvo doherty je osvojio svjetsko prvenstvo 1997 evans je izgubio prvi okvir 71-15 protiv dohertyja ali dudley native se izborio i poveo 4-3 ken doherty, međutim uspio je zatvoriti zadivljujuće natjecanje 10-8*
Predicted summary LSTM	*trener je rekao da je bio na putu za trening nakon što je izgubio 3 0 od njemačke finalu svjetskog prvenstva 2010 ali je uspio pobijediti na turniru na svjetskom prvenstvu 2014 ali je uspio pobijediti na turniru na igrama londonu*
Predicted summary Bi-LSTM	*evans je pobijedio 6 5 7 5 6 5 7 5 polufinalu na australian openu evans je počeo igrati za pobjedu od 7 6 7 6 7 6 7 6 7 6 7 6 1994 sada je pobijedila 6 0 7 6 7 5 polufinalu*

6 Discussion and Conclusion

The summarization task relies on a large amount of data that consists of pairs of text and summary. It represents a laborious challenge to obtain an adequate dataset for additional language such as Croatian. To acquire the dataset, we used Google to machine translate CNN/DailyMail into the Croatian language.

On Croatian MT CNN/DailyMail dataset, the baseline sequence-to-sequence models - LSTM and BiLSTM with attention mechanism are trained. The results are evaluated on ROUGE 1, 2, L, and BERTScore metrics and then compared with the evaluation performed on the original English dataset.

ROUGE is a simple metric that compares exact words between reference and generated summaries and does not give information about the semantic similarity of the words. For example, ROUGE cannot recognize that different words can represent the same term. ROUGE metrics show that sequence-to-sequence models in general do not infer good results, yet the English model achieves slightly better results than the Croatian. That is expected since the number of parameters is higher due to the larger vocabulary and for the training of Croatian summarization the larger quantities of texts are needed.

The results on the English dataset are much better than the results on the Croatian MT dataset. Regardless of the results, generated summaries are not of the desired content, but topics are well captured. The model correctly infers the topic but misses generating the correct information. Still, this initial attempt is indicative and motivational, that with a lot of improvements both in the selected architecture and in the dataset, we can progress towards the adequate solution for the abstractive text summarization in the Croatian language.

To conclude, the translated version of CNN/DailyMail can serve as a shortcut for training the initial summarization model for Croatian.

As the next step, the Croatian MT dataset will be extended with additional data and used to train transformer models. Transformers made huge progress in the field of abstractive summarization and regularly reached much higher evaluation scores than sequence-2-sequence models, however, they require a lot of computing resources and a huge amount of data to achieve better results.

Acknowledgment

This work has been fully supported by the University of Rijeka under project number uniri-drustv-18-20.

References

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K., Text Summarization Techniques: A Brief Survey, *International Journal of Advanced Computer Science and Applications (ijacsa)*, 8(10), 2017. http://dx.doi.org/10.14569/IJACSA.2017.081052

Babić, K., Petrović, M., Beliga, S., Martinčić-Ipšić, S., Matešić, M., Meštrović, A. (2021). Characterisation of COVID-19-Related Tweets in the Croatian Language: Framework Based on the Cro-CoV-cseBERT Model. In *Applied Sciences, 2021*. 11(21), 10442; https://doi.org/10.3390/app112110442

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align

and translate. In *3rd International Conference on Learning Representations (ICLR'15),* Bengio Yoshua and LeCun Yann (Eds.).

Cheng, J., Dong, L., & Lapata, M. (2016). Long short-term memory-networks for machine reading. In *2016 Conference on Empirical Methods in Natural Language Processing*, pp. 551–561, Las Vegas, NV, USA.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Empiricial Methods in Natural Language Processing (EMNLP 2014)*, October. arXiv preprint arXiv:1406.1078.

Choi, J.D., &Palmer, M. (2012). Guidelines for the CLEAR Style Constituent to Dependency Conversion, *Technical report 01-12: Institute of Cognitive Science*, University of Colorado Boulder

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, pp 4171–4186. arXiv preprint arXiv:1810.04805.

Fellbaum, C. (2005). WordNet and wordnets. In *Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition.* pp 665-670. Oxford: Elsevier

Google Cloud Translation. (n.d.). https://cloud.google.com/translate

Graves, A., Fernández, S., & Schmidhuber, J. (2005). Bidirectional LSTM networks for improved phoneme classification and recognition. In *Artificial Neural Networks: Formal Models and Their Applications–ICANN 2005, Part II 15* pp 799-804. Springer Berlin Heidelberg.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9(8),* pp 1735-1780, November 1997, doi:-10.1162/neco.1997.9.8.1735.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1), pp 411-420.

Lin, C. Y., & Och, F. J. (2004). Looking for a few good metrics: ROUGE and its evaluation. In *4th Ntcir Workshops*. 2004, pp. 1-8.

Ljubešić, N., Agić, Ž., Klubička, F., Batanović, V., & Erjavec, T. (2018). Training corpus hr500k 1.0., *Slovenian language resource repository CLARIN.SI*, ISSN 2820-4042, http://hdl.handle.net/11356/1183.

Ljubešić, N., & Lauc, D. (2021). BERTić--The Transformer Language Model for Bosnian, Croatian, Montenegrin and Serbian. In *8th Workshop on Balto-Slavic Natural Language Processing*, pp 37–42, Kiyv, Ukraine. arXiv preprint arXiv:2104.09243.

Nallapati, R., Zhou, B., dos Santos, C., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. In *CoNLL-16 - 20th SIGNLL Conference on Computational Natural Language Learning*, pp 280–290. https://doi.org/10.18653/v1/k16-1028

Novikova, J., Dušek, O., Curry, A. C., & Rieser, V. (2017). Why we need new evaluation metrics for NLG. In *2017 Conference on Empirical Methods in Natural Language Processing*. arXiv preprint arXiv:1707.06875.

Papers with Code. (n.d.). Document Summarization. https://paperswithcode.com/task/document-summarization

Rush, M., Chopra, S., & Weston, J., (2015). A neural attention model for abstractive sentence summarization. In *2015 Conference on Empirical Methods in Natural Language Processing*, pp. 379–389, Boston, MA, USA.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11), pp 2673-2681

Seljan, S. (1999). Lexical-Functional Grammar of the Croatian Language. In *22nd International Convention MIPRO '99* - CIS Computers in Intelligent Systems (CIS) Rijeka: Liniavera.

Shekhar, R., Karan, M., & Purver, M. (2022). CoRAL: a Context-aware Croatian Abusive Language Dataset. In *Findings of the ACL: AACL-IJCNLP*. arXiv preprint arXiv:2211.06053.

Shekhar, R., Pranjic, M., Pollak, S., Pelicon, A., & Purver, M. (2021). 24sata news comment dataset 1.0., *Slovenian language resource repository CLARIN.SI*, ISSN 2820-4042, http://hdl.handle.net/11356/1399.

Suleiman, D., & Awajan, A. (2020). Deep learning based abstractive text summarization: approaches, datasets, evaluation measures, and challenges. In *Mathematical problems in engineering, 2020*, pp 1-29. https://doi.org/10.1155/2020/9365340

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems, 27*. pp 3104-3112. arXiv preprint arXiv:1409.3215

Turc, I., Chang, M. W., Lee, K., & Toutanova, K. (2019). Well-read students learn better: On the importance of pre-training compact models. arXiv preprint arXiv:1908.08962.

Ulčar, M., & Robnik-Šikonja, M. (2020). FinEst BERT and CroSloEngual BERT: less is more in multilingual models. In *Text, Speech, and Dialogue: 23rd International Conference, TSD 2020*, Proceedings 23. pp 104-111. Springer International Publishing.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, pp 5998–6008.

Weischedel, R., Palmer, M., Marcus, M., Hovy,E., Pradhan, S., Ramshaw, L., …Houston, A. (2013). OntoNotes Release 5.0. *Philadelphia: Linguistic Data Consortium.* https://doi.org/10.35111/xmhb-2b84

Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., & Affandy, A. (2022). Review of automatic text summarization techniques & methods. In *Journal of King Saud University-Computer and Information Sciences*, 34(4), pp 1029-1046.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., …Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. In *International Conference on Learning Representations*. arXiv preprint arXiv:1904.09675.

Comparative Analysis of YOLOv5 and YOLOv6 Models Performance for Object Classification on Open Infrastructure: Insights and Recommendations

Marko Horvat, Ljudevit Jelečević, Gordan Gledec
University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Applied Computing
Unska 3, HR-10000 Zagreb, Croatia
{marko.horvat3, ljudevit.jelecevic, gordan.gledec}@fer.hr

Abstract. *In this paper, the performance of YOLOv5 and YOLOv6 models is evaluated on standardized object classification tasks using Kaggle open-access cloud processing infrastructure. The performance of these architectures is evaluated on a balanced subset of COCO dataset considering precision, recall, $mAP_{0.5}$, and $mAP_{0.5:0.95}$ metrics, learning and processing time. Our results indicate that the choice of the optimal model in the rapidly evolving field of computer vision and machine learning is heavily influenced by the dataset size, complexity, and restrictions imposed by the data processing platform. It is not always advisable to select the most recent algorithm for all applications. The obtained results, program code and the implemented architectures are available for non-commercial use at: https://github.com/mhorvat/yolov5_yolov6_comparison*

Keywords. computer vision, image classification, deep learning, algorithm evaluation, performance comparison, YOLO

1 Introduction

Identifying and categorizing objects in an image or video stream is a fundamental task in computer vision (Guo et al., 2022; Mohan et al., 2023). Recently, there has been significant research interest in advancing and refinement of deep neural network models for object classification tasks in different domains because of their superior performance compared to previous methods, as per example in medicine (Boaro et al., 2022). The You Only Look Once (YOLO) model is unique in that it performs object recognition in a single pass (Redmon et al., 2016). This makes it significantly faster than other recognition models such as R-CNN and its variants (Srivastava et al., 2021). In this regard, the YOLO model family of deep learning networks has emerged as the most popular and effective open-code solution for various real-time computer vision tasks including object detection and categorization (Jiang et al., 2022). It's well-suited for applications such as video surveillance, self-driving cars, and augmented reality (Sukkar, Kumar & Sindha, 2021; Zaghari et al., 2021; Zhang et al., 2023). The YOLOv5 model, released in 2020, achieved outstanding results on several benchmarks and has been widely accepted by the computer vision community (Jocher, 2020; Solawetz, 2020; Jiang et al., 2022). More recently, the YOLOv6 model was released in 2022 and has been proposed to further improve performance compared to its predecessor (Li et al., 2022). Given the advancements in both models, it is beneficial to conduct a thorough comparison between YOLOv5 and YOLOv6 to determine their respective strengths and weaknesses on a common dataset and in real-world applications.

In our previous study, the authors investigated the performance of different YOLOv5 models (Horvat, Jelečević & Gledec, 2022). Through a comparative analysis using an everyday image dataset, the authors provided recommendations for selecting the most appropriate YOLOv5 model based on the specific problem type, helping researchers make informed decisions. The authors continued this line of research with a comparative analysis of the already accepted YOLOv5 and the newer YOLOv6 architecture.

To evaluate the performance of the YOLOv5 and YOLOv6 models, experiments were performed on a subset of the COCO dataset (Lin et al., 2014). The performance of the models was evaluated based on several key metrics, including precision, recall, $mAP_{0.5}$, and $mAP_{0.5:0.95}$ metrics, as well as learning and processing time (Padilla et al., 2020). The analysis was facilitated by the open data storage and Graphical Processing Unit (GPU) hardware infrastructure provided by the online data science and machine learning platform Kaggle (Bojer & Meldgaard, 2021). In addition to the primary goal of comparing YOLOv5 and v6 models, this research also aimed to demonstrate

the practicality of utilizing a modern, freely available cloud computing infrastructure for conducting realistic deep learning computer vision tasks. It demonstrates that high-quality research can be carried out without the need for expensive, dedicated hardware resources.

The authors believe that evaluating YOLO models through openly accessible processing resources is crucial for researchers with access to limited resources. Advancements in computer vision methods may not be as beneficial for these practitioners, due to their inability to access high-end infrastructure and extensive image datasets for deep model training. Consequently, such users often resort to working with datasets containing only a few thousand images, which is not ideal for achieving optimal results. By resorting to open, cloud-based GPU infrastructure, these users can effectively overcome their computational and financial constraints for developing deep learning applications. This democratizes access to advanced machine learning techniques and promotes innovation and research.

The remainder of the paper is organized as follows: Section 2 describes our experimental setup, starting with an overview of the COCO and COCO minitrain in Section 2.1, followed by Section 2.2, which provides a comprehensive explanation of the dataset sampled specifically for training of the YOLOv5 and YOLOv6 models. The specific process of training these models is described in Section 2.3. Then, in Section 3, the results of our object classification experiments are presented. Section 3.1 discusses the processing time while Section 3.2 discusses the object recognition results for all classes. In addition, Section 3.3 focuses on the recognition results for just one class. Section 4 provides a detailed discussion of the obtained results and provides recommendations for choosing the best model for a given application. Finally, Section 5 concludes the paper, summarizes main findings, and provides outlook for further research.

2 Experimental setup

In this section the undertaken experimental approach is outlined which involves four primary steps: (i) the construction of the experimental dataset, (ii) the selection of YOLOv5 and v6 models for comparative analysis, (iii) the deployment of the selected models on a cloud-based platform, and finally, and (iv) the training of these models.

When deciding on an optimal dataset for testing, it is crucial to consider the standards set by the published literature. In Deep Learning (DL) and Artificial Neural Networks (ANNs) research, it is common to use purpose-built and established datasets to benchmark the performance of the models. One commonly employed dataset for DL models in the field of computer vision, including YOLO, is the COCO (Common Objects in Context) (Lin et al., 2014).

Experimental procedures in similar studies reported in the published literature usually require 256 or 300 epochs, which necessitates the use of one or more dedicated GPUs. A prominent example of such hardware is the Nvidia P100 and V100 GPUs, which have been specifically designed for machine learning (ML) applications and DL ANN training (Markidis et al., 2018).

It should be noted, however, that despite the availability of specialized hardware, the training process is anything but fast. Despite the simplicity of some models, the time required can still be considerable as investigated in this study.

2.1 COCO and COCO minitrain

The COCO dataset is a large-scale object detection, segmentation, and captioning dataset often used in computer vision research and applications (Lin et al., 2014). It is specifically designed for object recognition, segmentation, and captioning in the context of scene understanding. The dataset contains over 200,000 labeled images with over 1.5 million object instances and 80 object categories. The dataset is notable for its quality, diversity and complexity, with images featuring objects in a wide range of orientations, sizes, and contexts.

COCO minitrain is a carefully curated training set derived from the much larger COCO train2017 dataset, containing 25,000 labeled images (~20% of train2017), ~184,000 annotations and 80 object categories (Samet, Hicsonmez & Akbas, 2020). This smaller dataset, and the accompanying data processing software support, was developed to enable efficient hyperparameter tuning and reduce the computational cost associated with experiments in computer vision. Designed as a more compact alternative to the complete COCO dataset, the minitrain dataset allows researchers and engineering to optimize their models and perform initial tests in a time-saving manner without compromising the quality of their work. A sample of the COCO and COCO minitrain images are shown in Figure 1.

Typically, smaller datasets may be useful for rapid prototyping as training models on smaller datasets can be much faster, which allows for more iterations and fine-tuning of the model and its hyperparameters. Also, large datasets require significant computational resources (CPU/GPU power, memory, storage space) that may not always be available, especially in low-resource environments. More compact datasets are also used for model architecture debugging and performance checking before migrating to larger datasets or production environments.

The COCO minitrain is a valuable resource for the computer vision community because it enables faster experimentation and model development while providing a representative sample of the larger COCO dataset. This important feature supports reproducibility

and simplifies validation of the obtained research by other teams.

Figure 1. Examples of labeled and semantically segmented images in the COCO and COCO minitrain.

The statistical distribution of object classes in the minitrain dataset closely mirrors those of the COCO2017 dataset, ensuring that the performance and behavior of models trained on the minitrain dataset generalize well to the entire dataset. To ensure the quality of sampling, labels within the COCO2017 dataset are further divided into three size groups: S, M, and L, representing small, medium, and large object regions, respectively, as they appear in the images. The dimensions defining these size categories are defined by the specifications of the COCO metric (Lin et al., 2014).

To maintain a similar distribution, the minitrain images were randomly sampled from the full COCO2017 set while preserving the following three quantities as much as possible: 1) the proportion of object instances from each class, 2) the overall ratios of S, M, and L objects, and 3) the per-class ratios of S, M, and L objects.

In the original COCO minitrain version, sampling was done by iterative random sampling while minimizing the fitness function – the discrepancy between the largest and smallest ratio of the number of tags in both the COCO2017 set and the sample. To increase the likelihood of detecting a sample whose distribution is very similar to the label distribution in the COCO2017 dataset, the penalty function was modified (Jelečević, 2023). The Kullback-Leibler divergence was used along with Laplace smoothing as a more effective measure of similarity between the distributions (Zhuang et al., 2015). This approach helped to ensure that the selected training sample accurately represents the general class distribution of the COCO2017 dataset.

2.2 Dataset for YOLOv5 and YOLOv6 models training

The experimental dataset was generated using balanced sampling from the COCO train2017. The original dataset was sampled 200,000 times before generating a subset of 1,000 images with a total of 7,587 labels that was used for model training. The sampling pseudocode for generating the training subset is given here:

```
best_sample = none;
for i in range(0, 200,000):
 sample = rnd_select_sample(1000 images,
  from COCO_train2017_dataset);
 eval_score =
  eval_sample_representation(sample, based
  on label distribution);
 if (best_sample = none or eval_score >
   best_eval_score):
  best_sample = sample;
  best_eval_score = eval_score;
```

And the actual Python code run to obtain the experimental dataset uses COCO minitrain libraries:

```
!python sample_coco.py --coco_path
"/kaggle/datasets/coco" --save_file_name
"instances_train2017_minicoco_1k" --
save_format "json" --sample_image_count
1000 --run_count 200000 --
allow_empty_sample_classes
```

The statistical chi-square test was performed to verify the COCO train2017 dataset and the sampled subset had the same class distribution. The test resulted in a p-value $p=4.2 \cdot 10^{-17}$, well below the significance of $\alpha=0.05$, This led to the acceptance of the null hypothesis, indicating a correlation between the distribution of the COCO train2017 dataset and the sampled subset. Therefore, the chi-square test confirmed that the COCO train2017 dataset and the sampled subset indeed have the same distribution.

Also, the Pearson correlation between the population and the subset is $r=0.9938$, indicating a strong linear dependence between number of instances of all classes in the full COCO train2017 dataset and the sampled subset used for training.

Figure 2 shows the distribution of the 30 most frequent classes in the sampled dataset and in the original COCO train2017 dataset. The 30 most common object categories and their IDs are: Person (1), Car (3), Chair (62), Book (84), Dining table (67), Bottle (44), Cup (47), Umbrella (28), Bowl (51), Carrot (57), Horse (19), Kite (38), Handbag (31), Banana (52), and Potted plant (64). Some classes are repeated in different image sizes (S, M, or L).

2.3 YOLOv5 and YOLOv6 training process

Each YOLO model was trained on the entire training subset stored on the online platform Kaggle using the

virtualized Nvidia P100 GPU available in the programming environment[1].

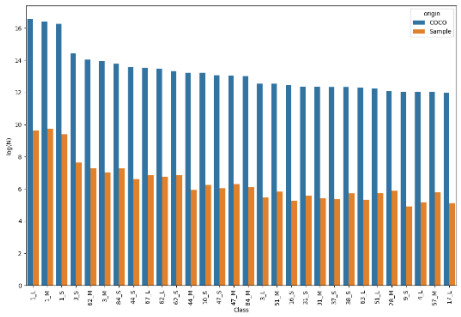


Figure 2. The distribution of the 30 most frequent classes in the COCO2017 and the experimental subsample.

Three pairs of mutually equivalent models were chosen for the training process: YOLOv5s and YOLOv6-N, YOLOv5m6 and YOLOv6-M, as well as YOLOv5x and YOLOv6-M6. Each pair consists of one model from the YOLOv5 and one from the YOLOv6 series.

Each training session lasted for 300 epochs, meaning that the Kaggle early stopping feature was not activated, which would halt the training if no progress was made in the last 100 epochs. Early stopping is a form of regularization method used to prevent overfitting in an iterative training. Thanks to the early stopping feature, at least some learning took place across all the models during the initial 200 epochs.

The scripts for training the YOLOv5 and YOLOv6 models were run with the following parameters:

```
!python train.py --data coco.yaml --epochs
300 --weights '' --cache --cfg
yolov5m6.yaml  --batch-size 16

!python tools/train.py --img 640 --batch 16
--epochs 300 --conf configs/yolov6n.py --
data data/coco.yaml
```

Tables 1 and 2 show the training durations, indicated in hours (h) and minutes (m), as well as the training speeds, denoted in the number of parameters per second (N/s), for the selected S, M, and L pairs of YOLOv5 and YOLOv6 models. The better results are highlighted in green.

Table 1. The duration of YOLO models training in hours and minutes

	YOLOv5		YOLOv6	
	Model	Duration	Model	Duration
S	YOLOv5s	05h 50m	YOLOv6-N	07h 26m
M	YOLOv5m6	09h 12m	YOLOv6-M	10h 21m
L	YOLOv5x	17h 20m	YOLOv6-M6	11h 18m

Table 2. The speed of YOLO models training in parameters per second

	YOLOv5		YOLOv6	
	Model	N/s	Model	N/s
S	YOLOv5s	19.05	YOLOv6-N	9.76
M	YOLOv5m6	59.88	YOLOv6-M	52.04
L	YOLOv5x	76.57	YOLOv6-M6	108.71

As expected, the more complex models require a longer training period. However, as can be seen in Table 2, complex models prove to be more efficient during training, effectively training more parameters in the same amount of time. Of all the models, YOLOv6-M6 stands out as the most efficient in terms of training.

3 Object classification results

As already explained, for the comparative analysis and benchmarking of YOLOv5 and YOLOv6 models the COCO val2017 dataset was used which includes a total of 5,000 images. It should be emphasized that the sampled training dataset with 1,000 images is not a subset of COCO val2017 thus ensuring a comprehensive testing process.

It should be noted that the experimental results were obtained without the use of Nvidia TensorRT, an SDK for high-performance inference runtime optimizer that delivers low latency and higher throughput for DL applications (Markidis et al., 2018; Shafi et al, 2021). However, the TensorRT SDK is commonly used in reference literature for YOLOv6 benchmarking (Li et al., 2022). This SDK is available as an optional feature on the Kaggle platform and, as reported in the published literature, can improve processing speed by three times (Markidis et al., 2018; Shafi et al, 2021).

Given the large number of image classes in the COCO2017 dataset, the relatively limited size of the evaluation dataset, and the significant proportion of the class “Person”, which accounts for one-third of all labels in the test subset, object classes detection results are presented separately in two parts: performance metrics calculated for all classes in the subset, and only to the “Person” class, in sections 3.2 and 3.3, respectively. This approach was chosen because of the unique characteristics and challenges presented by the size and class distribution of the evaluation subset.

3.1 Processing time

In the YOLO architecture, an input image goes through three different processing stages, each of which contributes significantly to the speed and accuracy of the output (Diwan, Anirudh & Tembhurne, 2023; Jiang et al., 2022).

1 Kaggle, https://www.kaggle.com/ .

Preprocessing is the first stage in which the input image is prepared for further processing. This includes resizing the image to the required dimensions, depending on the YOLO version, and normalizing the pixel values. During normalization, the pixel intensities are usually scaled to a range between 0 and 1 or -1 and 1 so that the model converges faster during training.

Inference is the core stage where the model makes predictions on the preprocessed image and the actual object recognition takes place. The preprocessed image is passed through the YOLO neural network, which outputs a three-dimensional tensor. This tensor contains the bounding box coordinates, class probabilities, and object scores for each grid cell in the image.

Non-Maximum Suppression (NMS) is the final stage that refines the results of the inference phase and helps to reduce the number of overlapping detections. The YOLO model can predict multiple bounding boxes for the same object, and NMS helps eliminate redundancies. To do this, first all bounding boxes whose objectness score is below a certain threshold are discarded. Then, for the remaining boxes, the box with the highest score is selected and all other boxes with a high overlap – measured by the Intersection over Union (IoU) metric – with the selected box are eliminated. This process is repeated until all boxes have either been selected or discarded. In Table 3 the duration of preprocessing of all input images in the evaluation dataset on the Kaggle's Nvidia P100 cloud infrastructure is shown in milliseconds [ms], separately for all three stages and in total. The best scores are highlighted in green.

Table 3. Duration required for processing all input images utilizing the Nvidia P100 cloud infrastructure

	Preprocessing [ms]	Inference [ms]	NMS [ms]	Total [ms]
YOLOv5s	0.20	2.80	1.90	4.90
YOLOv5m6	0.20	6.90	1.70	8.80
YOLOv5x	0.20	18.10	1.90	20.20
YOLOv6-N	0.12	1.57	18.40	20.09
YOLOv6-M	0.12	7.96	28.16	36.24
YOLOv6-M6	0.15	9.47	31.65	41.27

From Table 3, it can be seen that the preprocessing times are relatively consistent across all models, ranging from 0.12 ms to 0.2 ms. However, there are significant differences in the inference times and the Non-Maximum Suppression (NMS) times. The YOLOv5s model has the shortest total processing time of 4.9 ms, which is primarily due to the low inference time of 2.8 ms. On the other hand, the YOLOv6-M6 model has the longest total processing time of 41.27 ms, with the NMS stage taking up a significant portion of this time (31.65 ms).

Figure 3 shows the expected approximate number of frames per second (FPS) for each model. The YOLOv5s model outperforms all others with 204 FPS score, which is consistent with its shortest total processing time in Table 3. This is a very significant difference that results in the YOLOv5x processing video signals at twice as many frames per second as the YOLOv6-M6 model with the same complexity. However, the YOLOv6-M6 model still achieves 24 FPS despite its longer processing time which is marginally enough for real-time applications.

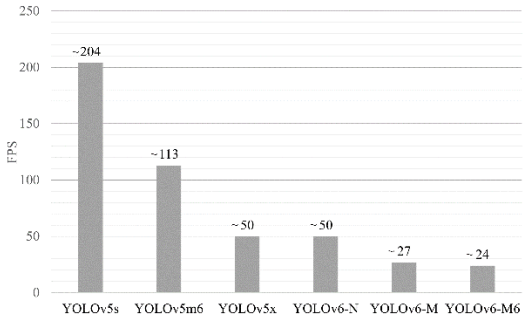


Figure 3. Approximate number of frames per second (FPS) for each model achieved on the Nvidia P100 cloud infrastructure

3.2 All class detection results

Table 4 shows precision, recall, $mAP_{0.5}$, and $mAP_{0.5:0.95}$ results for object detection of all COCO2017 classes in the evaluation dataset. The best results are again highlighted green.

Table 4. Object detection results for all COCO2017 classes in the evaluation dataset

	Precision	Recall	$mAP_{0.5}$	$mAP_{0.5:0.95}$
YOLOv5s	21.9	10.8	7.79	3.48
YOLOv5m6	24.8	12.7	9.28	4.52
YOLOv5x	22.1	12.5	9.10	4.59
YOLOv6-N	14.6	17.0	6.74	3.47
YOLOv6-M	15.9	20.0	8.51	4.81
YOLOv6-M6	15.2	20.0	8.14	4.59

From Table 4 it can be observed that the YOLOv5m6 model has the highest precision, indicating that it has the highest accuracy in detecting objects among all models. However, the YOLOv6-M and YOLOv6-M6 models have the highest recall, suggesting they are the most capable of identifying all relevant instances. In terms of $mAP_{0.5}$, the YOLOv5m6 and YOLOv5x models perform comparably well, with scores of 9.28 and 9.10 respectively. However, the YOLOv6-M model outperforms all others in the $mAP_{0.5:0.95}$ metric with a score of 4.81, indicating its superior performance across a range of IoU thresholds.

Figure 4 presents a comprehensive analysis of the relationships concerning the detection of all classes within the evaluation dataset. The top diagram in Figure 4 shows the relationship between $mAP_{50:95}$ and

the number of model parameters. This relationship is significant because it provides insight into how the complexity of the model, indicated by the number of parameters, influences its precision performance. The bottom diagram illustrates the correlation concerning $mAP_{50:95}$ and the duration of image processing. This correlation is crucial because it shows the balance between the precision of the model and the processing speed. Essentially, it provides an understanding of how fast the model can process images without compromising its precision.

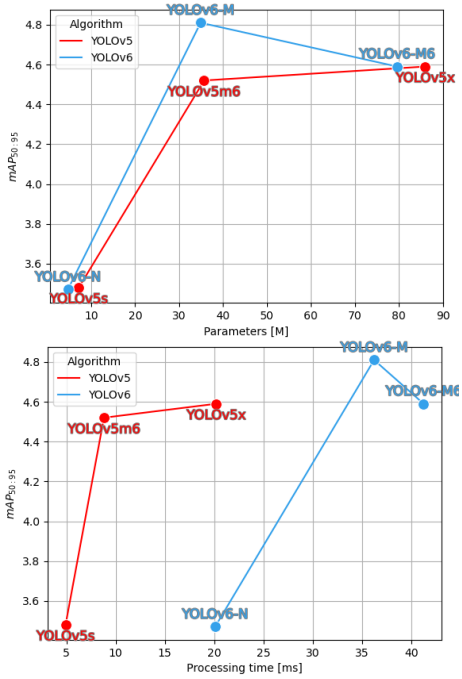


Figure 4. Relationships between $mAP_{50:95}$ and the number of model parameters (top), and $mAP_{50:95}$ and the duration of image processing (bottom) for detection of all classes

3.3 One class detection results

Table 5 presents object detection results only of "Person" class in the evaluation dataset.

Table 5. Object detection results only for class "Person" in the evaluation dataset

	Precision	Recall	$AP_{0.5}$	$AP_{0.5:0.95}$
YOLOv5s	39.6	45.4	39.8	16.9
YOLOv5m6	42.6	48.3	44.2	20.4
YOLOv5x	45.1	47.7	43.1	20.7
YOLOv6-N	54.6	41.0	41.9	19.7
YOLOv6-M	60.3	46.0	49.1	25.2
YOLOv6-M6	62.7	44.0	48.1	24.6

Figure 5 shows relationships between $mAP_{50:95}$ and the number of model parameters, and $mAP_{50:95}$ and the duration of image processing in the top and bottom diagrams, respectively, for object detection of only one class "Person" in the evaluation dataset.

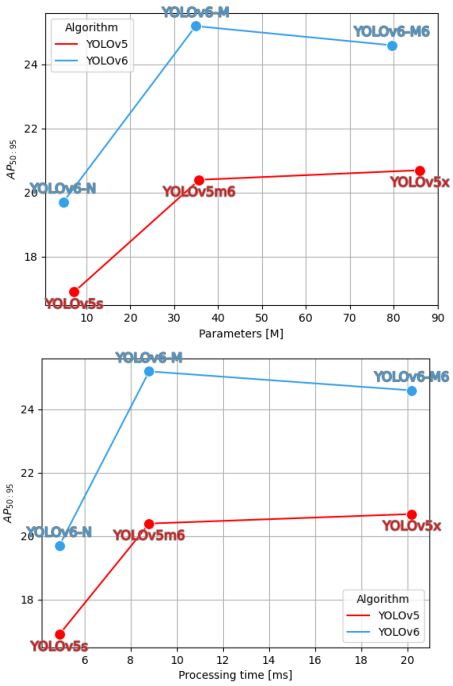


Figure 5. Relationships between $mAP_{50:95}$ and the number of model parameters (top), and $mAP_{50:95}$ and the duration of image processing (bottom) for detection of only one class "Person"

Based on the detection results for the "Person" class alone, the YOLOv6-M model outperforms others, irrespective of whether the evaluation is based on the constant Jaccard index metric ($AP_{0.5}$) or the metric that considers a range of index values ($AP_{0.5:0.95}$). This indicates that the YOLOv5 model is better at handling unbalanced labels, while the YOLOv6 performs better in object localization.

An important difference compared to the results for detection of all classes (in Section 3.2) is the significantly stronger correlation between the time required for image processing and the number of parameters in both the YOLOv5 and YOLOv6 models. This suggests that the computational efficiency of the YOLOv5 and YOLOv6 models is significantly positively affected by the number of parameters used, due to the complexity introduced by a larger parameter space. The increased time requirement could potentially impact the real-time capability of the models. Consequently, optimizing the number of parameters can significantly improve the processing time and overall performance of these models without sacrificing object detection accuracy.

4 Discussion

It is tempting to believe that a more recent deep learning architecture is always the preferable choice for

any research or software platform, but selecting the optimal data processing model is more delicate than that. To make an informed and correct decision, it is necessary to consider three crucial factors: 1) the YOLO model intended function, 2) available hardware, and 3) training time limits.

When using a pretrained model for image processing and processing time is not an issue, opting for the most complex model with the highest possible accuracy is a viable option. However, for real-time video signal processing involving localization and classification with multiple labels, the YOLOv5m6 model is recommended because it demonstrates a favorable speed-accuracy ratio. If the task is less demanding with fewer labels and the focus is on localization, the YOLOv6-M6 model is a more suitable alternative.

On the other hand, in scenarios where training time is limited, the YOLOv6 network is a better choice compared to the marginally slower YOLOv5, especially when working with multi-class models. YOLOv6, with its more advanced architecture, is designed to efficiently process complex object detection tasks, offering a balance between speed and accuracy.

Although this paper provides initial guidelines for choosing a suitable YOLOv5 or YOLOv6 model, it is advisable for researchers to adapt their selection to their specific requirements and use-cases. This can be achieved by consecutively testing different YOLO models on a subset of their data and evaluating the performance on the intended hardware after deployment. By conducting such iterative experimental assessments on a limited dataset, researchers can gather the necessary insights to make an informed decision about the most suitable YOLO model for their specific needs.

5 Conclusion

The YOLO framework deployed in cloud environment provides an out-of-the-box solution that offers great benefits to engineers and researchers who may lack extensive experience in computer vision algorithms and access to high-end processing infrastructure. However, to improve the efficiency of object recognition without delving into optimization of deep learning and machine learning algorithms, it would be recommended to improve the understanding of the scene. One possible method to improve classification performance is to neglect redundant image sections and focus exclusively on authentic regions of interest before integrating the image into the YOLO workflow. The use of knowledge graphs and ontologies, together with automated reasoning services such as expert systems is strongly recommended to obtain a truly semantically rich scene description. These methods facilitate a comprehensive understanding of the semantic regions detected in the image and their conceptual relationships (Horvat, Grbin & Gledec, 2013a; Horvat, Grbin & Gledec, 2013b).

In conclusion, the choice of a DL architecture such as YOLOv5 or YOLOv6 depends on several factors, including the intended function, available hardware, and time constraints for training. For image processing tasks where time is not an issue, more complex models with high accuracy can be used. For real-time video processing requiring classification and localization with multiple labels, YOLOv5m6 is recommended due to its optimal speed-to-accuracy ratio. For tasks with fewer labels where localization is the primary concern, YOLOv6-M6 is recommended. When training time is limited, YOLOv6 outperforms the slightly slower YOLOv5, especially for multiclass models, by providing a balance between speed and accuracy. Despite these guidelines, it is recommended that researchers tailor their selection to specific needs by testing different YOLO models on a subset of data and evaluating performance on the intended hardware after deployment. This iterative testing allows for informed decisions about the most appropriate YOLO model for their unique requirements.

References

Boaro, A., Kaczmarzyk, J. R., Kavouridis, V. K., Harary, M., Mammi, M., Dawood, H., Shea, A., Cho, E. Y., Juvekar, P., Noh, T., Rana, A., Ghosh, S., & Arnaout, O. (2022). Deep neural networks allow expert-level brain meningioma segmentation and present potential for improvement of clinical practice. *Scientific Reports*, *12(1)*, 1–3. https://doi.org/10.1038/s41598-022-19356-5

Bojer, C. S., & Meldgaard, J. P. (2021). Kaggle forecasting competitions: An overlooked learning opportunity. *International Journal of Forecasting*, *37*(2), 587-603. https://doi.org/10.1016/j.ijforecast.2020.07.007

Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*, *82*(6), 9243-9275. https://doi.org/10.1007/s11042-022-13644-y

Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., Zhang, S. H., Martin, R. R., Cheng, M. M., & Hu, S. M. (2022). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, *8(3)*, 331–368. https://doi.org/10.1007/s41095-022-0271-y

Horvat, M., Grbin, A., & Gledec, G. (2013). *WNtags: A web-based tool for image labeling and retrieval with lexical ontologies*. *Frontiers in artificial intelligence and applications*, *243*, 585–594. https://doi.org/10.3233/978-1-61499-105-2-585

Horvat, M., Grbin, A., & Gledec, G. (2013). Labeling and retrieval of emotionally-annotated images using WordNet. *International Journal of Knowledge-based and Intelligent Engineering Systems, 17*(2), 157–166. https://doi.org/10.3233/KES-130269

Horvat, M., Jelečević, Lj., & Gledec, G. (2022). A Comparative study of YOLOv5 models performance for image classification. In *Proceedings of the 33rd Central European Conference on Information and Intelligent System (CECIIS 2022).* (pp. 349–356).

Jelečević, Lj. (2023). Modified COCO minitrain. Retrieved from https://github.com/thelcrysis/coco-minitrain

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. *Procedia Computer Science*, *199*, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135

Jocher, G. (2020). YOLOv5. Retrieved from https://github.com/ultralytics/yolov5

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–755). Springer, Cham.

Markidis, S., Der Chien, S. W., Laure, E., Peng, I. B., & Vetter, J. S. (2018, May). Nvidia tensor core programmability, performance & precision. In *2018 IEEE international parallel and distributed processing symposium workshops (IPDPSW)* (pp. 522-531). IEEE. https://doi.org/10.1109/IPDPSW.2018.00091

Mohan, R., Elsken, T., Zela, A., Metzen, J. H., Staffler, B., Brox, T., Valada, A., & Hutter, F. (2023). Neural Architecture Search for Dense Prediction Tasks in Computer Vision. *International Journal of Computer Vision*, 1–24. https://doi.org/10.1007/s11263-023-01785-y

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237–242). IEEE.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).

Samet, N., Hicsonmez, S., & Akbas, E. (2020). Houghnet: Integrating near and long-range evidence for bottom-up object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16* (pp. 406-423). Springer International Publishing.

Shafi, O., Rai, C., Sen, R., & Ananthanarayanan, G. (2021, November). Demystifying tensorrt: Characterizing neural network inference engine on nvidia edge devices. In *2021 IEEE International Symposium on Workload Characterization (IISWC)* (pp. 226-237). IEEE.

Solawetz, J. (2020). Yolov5 new version-improvements and evaluation. *Roboflow. Seach date*. Retrieved from https://blog.roboflow.com/yolov5-improvements-and-evaluation/

Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. (2021). Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, *8*(1), 1-27.

Sukkar, M., Kumar, D., & Sindha, J. (2021, July). Real-Time Pedestrians Detection by YOLOv5. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 01–06). IEEE.

Zaghari, N., Fathy, M., Jameii, S. M., & Shahverdy, M. (2021). The improvement in obstacle detection in autonomous vehicles using YOLO non-maximum suppression fuzzy algorithm. *The Journal of Supercomputing*, *77*(11), 13421-13446.

Zhang, L., Xu, F., Liu, Y., Zhang, D., Gui, L., & Zuo, D. (2023). A posture detection method for augmented reality–aided assembly based on YOLO-6D. *The International Journal of Advanced Manufacturing Technology*, *125*(7-8), 3385-3399.

Zhuang, F., Cheng, X., Luo, P., Pan, S. J., & He, Q. (2015, June). Supervised representation learning: Transfer learning with deep autoencoders. In *Twenty-fourth international joint conference on artificial intelligence*

ComFac4LT – Computing Facilities for Language Technologies

Gaurish Thakkar, Marko Tadić
University of Zagreb
Faculty of Humanities and Social Sciences
Ul. Ivana Lučića 3, 10000, Zagreb
{gthakkar, marko.tadic}@ffzg.hr

German Rigau
University of the Basque Country
HiTZ Center - Ixa
M. Lardizabal 1, 20080, Donostia
german.rigau@ehu.eus

Abstract. *High-performance computing (HPC) is an essential service required in multiple research areas. This paper maps the existing HPC landscape and enumerates the various access types available to researchers working in Europe. Furthermore, the present study examines the perspectives of researchers in the field of language technology gathered through an online survey. Requesting resources from HPC service providers can be a time-consuming process; therefore, a dynamic access mode should be made available to reduce the request time. Collaboration with academia is a viable option for a SME user who wishes to utilise HPC services. Academic institutions in a nation without HPC resources should rely on international partnerships for HPC services. Users should have access to centralised information regarding HPCs to facilitate the location of services.*

Keywords. High-Performance Computing, Language Technologies, PRACE, EuroHPC-JU, Natural Language Processing

1 Introduction

Language Technology (LT) is a highly researched field with high socio-economic impacts. Textual analysis processes facilitate knowledge acquisition and strategic decision-making. The additional knowledge extracted from the text has been largely attributed to advances in the field of. All of these advancements have been made possible by the availability of data, improved data processing techniques (algorithms), and processing capabilities made available to researchers over time.

There has been a clear shift away from knowledge-based and human-engineered methods and towards data-driven methods, which has led to progress in the field of LT. One recent aspect associated with the paradigm shift in language processing is the use of large language models. Large-scale monolingual and/or multilingual text data is used to train language models. Pre-trained large language models, like BERT (Devlin et al., 2019), GPT (Brown et al., 2020), GPT-4 (OpenAI, 2023), and XLM-Roberta (Conneau et al., 2020), have offered a framework for using the knowledge acquired during the training process to be later applied to newer tasks.

As previously stated, one aspect associated with the boom of AI-based data-driven techniques for NLP is the ability to crunch data using efficient hardware in the form of Graphic Processing Units (GPUs). In neural language model training, the cost component is realised in the form of hardware and its operation. This directly results in organisations, with facilitated access to these hardware resources, having access to the research and development of LT (Ahmed and Wahed, 2020).

BLOOM (Scao et al., 2022) is a large language model with 176 billion parameters that can write text in 46 natural languages. The model was trained using the Jean Zay public supercomputer with 384 NVIDIA A100 80 GB GPUs (48 nodes) for 117 days. Building such models with numerous parameters that are learned during training necessitates an equally capable system with capable hardware.

This study reports on the results of an investigation of the available high-performance computing (HPC) facilities available to LT researchers. In addition to the different HPC infrastructures that are available, the paper looks at aspects like access protocols, calls, and eligibility. The compatibility of existing models is directly correlated with the available GPU. Hence, GPU hardware and associated details form another slice of data that is useful from an LT point of view. For research and innovation in the field of LT, a comprehensive understanding of the available infrastructure alternatives is essential.

The primary objective of this research is to evaluate the capacity of various high-performance computing (HPC) initiatives, including EuroHPC JU, PRACE, LUMI, and national consortia, in supporting contemporary LT. Specifically, the paper aims to:

- assess the hardware capabilities, especially the number of nodes and GPUs, of these HPC systems for both small and large-scale experiments.
- evaluate the accessibility protocols offered by these initiatives.

- examine the resource request and allocation procedures for various user roles.
- survey the various access modes provided by these HPCs, with a specific focus on accessibility for Small and Medium Enterprises (SMEs)..

2 Background

2.1 High-Performance Computing (HPC)

HPC, also called a supercomputer, provides the opportunity to solve complex problems in different applications[1]. Running applications in parallel to speed up performance is required for highly computational tasks such as pre-training a neural language model. In computing, floating point operations per second (FLOPS, flops, or flop/s) is a measure of computer performance, useful in fields of scientific computation that require floating-point calculations (Dolbeau, 2018). To put it into perspective, a laptop, or desktop with a 3 GHz processor can perform around 3 billion calculations per second. HPC solutions, on the other hand, can perform quadrillions of calculations per second, i.e., 1 Petaflops (10^{15}). The orders of magnitude in computer performance can be understood as follows:

- A 1 petaflops (PFLOPS) computer system is capable of performing one quadrillion (10^{15}) floating-point operations per second.
- A 1 exaflops (EFLOPS) computer system is capable of performing one quintillion (10^{18}) floating-point operations per second.

The following are the components[2] of an HPC solution:

- **Server**: responsible for computing
- **Network**: interconnection between the servers, responsible for high-speed transfers between servers and storage units.
- **Storage**: store for feeding data to servers, as well as persisting data received as output of the processing operation.

The collection of such servers (each server is a node) forms an HPC cluster. In addition to the above-mentioned components, there are accelerated nodes, i.e., computer nodes with GPUs or any other accelerator like a Xeon Phi[3]. At the time of collecting data for the paper, **Frontier**[4] HPC in the USA is rated at 1.685 exaFLOPS (Rpeak) and is the world's fastest supercomputer in operation. **Fugaku** HPC in Japan comes in second with 537 PFLOPS (Rpeak) followed by **LUMI** in Finland with 428 PFLOPS (Rpeak). The recent trend being followed by the infrastructure providers is to shift computing to the level of exascale (10^{18} floating point operations per second). Benchmarking (Luszczek et al., 2005; Jiang et al., 2019, 2021), performance monitoring (Truong et al., 2001), resource utilisation (Asch et al., 2018), user surveys and feedback (Wolter et al., 2006), and documentation review (Lathrop et al., 2019) are some methods used to investigate HPC facilities. Previous attempt at mapping the HPC landscape in Europe can be linked to Berberich et al. (2019); Eicker et al. (2020). However, visualising the HPC landscape with LT as the primary focus is an unexplored field.

2.2 HPC Initiatives in Europe

The European High-Performance Computing Joint Undertaking (**EuroHPC-JU**) (Skordas, 2019) is a joint project that brings together the resources of the European Union. It is involved in activities such as the procurement and installation of supercomputers throughout Europe. In addition, it is involved in developing sustainable HPC technologies for efficient and cleaner computing. Other objectives of EuroHPC-JU are to design and develop applications and algorithms for HPC services, as well ease access to potential HPC users like SMEs and HPC experts across Europe. To date, five supercomputers[5] are now fully operational: **LUMI** in Finland (which ranks number 3 in the world), **LEONARDO** in Italy (which ranks number 4 in the world), **Vega** in Slovenia, **MeluXina** in Luxembourg, **Discoverer** in Bulgaria, **Karolina** in the Czech Republic, and Supek in Croatia which was opened right after we completed this research, so, unfortunately, it wasn't included in this data collection. Two supercomputers are underway: **Deucalion** in Portugal, and **MareNostrum5** in Spain. The list of EuroHPC-JU public members can be found at `https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju_en`.

PRACE[6] (Partnership for Advanced Computing in Europe) (Hutton et al., 2019) is a not-for-profit international association that aims to facilitate access to a research infrastructure that enables high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. It has 25 member countries [7] whose representative organisations create a pan-European supercomputing infrastructure, providing access to computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level. The computer systems and their operations accessible through PRACE are provided by five PRACE members (BSC representing Spain, CINECA representing Italy, ETH Zurich/CSCS representing Switzerland, GCS representing Germany, and GENCI representing France). Figure 1 shows PRACE member countries.

[1] https://www.ff4eurohpc.eu/en/about/what-is-hpc/

[2] https://www.netapp.com/data-storage/high-performance-computing/what-is-hpc/

[3] https://en.wikipedia.org/wiki/Xeon_Phi

[4] https://en.wikipedia.org/wiki/Frontier_(supercomputer)

[5] https://digital-strategy.ec.europa.eu/en/policies/high-performance-computing-joint-undertaking

[6] https://prace-ri.eu/

[7] https://prace-ri.eu/about/members/

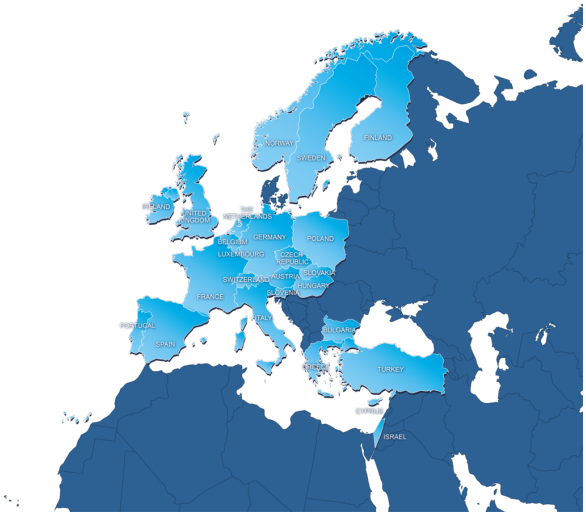


Figure 1: The list of PRACE members is displayed in the figure.

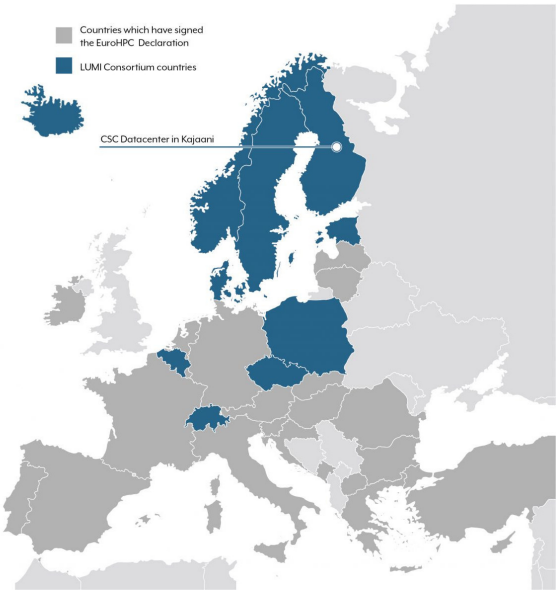


Figure 2: The list of LUMI consortium members is displayed in the figure.

LUMI consortium [8] consists of ten European countries and provides a high-quality, cost-efficient, and environmentally sustainable HPC ecosystem based on true European collaboration. The LUMI (Large Unified Modern Infrastructure) consortium countries are Finland, Belgium, the Czech Republic, Denmark, Estonia, Iceland, Norway, Poland, Sweden, and Switzerland. Half of the LUMI resources belong to the EuroHPC Joint Undertaking, and the other half of the resources belong to the participating countries, i.e., the LUMI consortium countries. Each consortium country has a share of the resources based on its contribution to the LUMI funding. The shares for each of the countries are allocated according to local considerations and policies, so LUMI is seen and handled as an extension of national resources. The LUMI shares belonging to the EuroHPC-JU are allocated by a peer-review process (comparable to that used for PRACE Tier-0 access). Figure 2 shows the LUMI consortium members.

Apart from the previously mentioned entities, individual countries in Europe provide and support HPC services to their respective researchers via national HPC centres or infrastructure managed via open research communities like universities.

There have been active attempts by EuroHPC-JU to increase the exposure of HPC to the existing member states. For example, the creation of new national competence centres (NCC) for HPC was taken up in the EuroCC[9] call. NCCs represent a focal point for HPC in the participating country, liaising with national initiatives in the area of HPC and facilitating access for national stakeholders to European HPC competences and opportunities in different industrial sectors and domains.

2.3 HPC Classifications

PRACE categorises European HPC facilities into three tiers: Tier-0 are European centres with petaflop machines; Tier-1 are national centres; and Tier-2 are regional centres. The resources under Tier-0 categorization are exclusively distributed via open-access calls and peer-review procedures. For Tier-1 HPC services, access is provided via DECI (Distributed European Computing Initiative), which is a programme under PRACE designed for research projects that require different resources from those currently available in the principal investigator's (PI) own country. At the same time, those projects should not require resources on the very largest (Tier-0) European supercomputers or very large CPU allocations. Another HPC category called Tier-3 exists to denote a university cluster. For example, Paderborn University's Noctua 1 [10] provides access to the members of Paderborn University.

2.4 HPC Calls

HPC resources are allocated either via open calls or by registering requests with the responsible authority via email or portals. In the case of national resources, i.e., Tier-1 HPC, which are linked to pan-European supercomputing infrastructure and ecosystems, the calls are divided into national and European calls. Researchers from universities, research institutions, and enterprises who match the eligibility conditions can utilise HPC

[8] https://www.lumi-supercomputer.eu/lumi-consortium/
[9] https://www.eurocc-access.eu/
[10] https://pc2.uni-paderborn.de/hpc-services/available-systems/noctua1/

services for free. An open-research agreement in which the results are made public serves as one of the most important prerequisites for free access to the resources. A third category of access to national resources exists in some cases where the work does not fall under the scope of the previous two types. In this case, the access is paid for, and the costs are calculated after analysis of the access request. Access to these systems is provided via open calls, where computing and data management resources are awarded through a peer review process[11]. Access to systems under the EuroHPC-JU is provided via open calls on the PRACE website.

2.5 Access Calls

2.5.1 PRACE

The following forms of access are available to PRACE systems:

- Preparatory Access is intended for short-term access to resources, for code-enabling and porting, and is required to prepare proposals for Project Access and to demonstrate the scalability of codes. The Call for Proposals for PRACE Preparatory Access is a continuously open call, with cut-off dates every 3 months. Preparatory Access[12] allows PRACE users to optimise, scale, and test codes on PRACE Tier-0 systems before applying to PRACE calls for Project Access.
  - **Benchmark Access** is designed for code scalability tests, the outcome of which is to be included in the proposal in a future EuroHPC Extreme Scale and Regular call. Users receive a limited number of node hours; the maximum allocation period is three months.
  - **Development Access** is intended for projects centred on the development and optimisation of code and algorithms. Users will typically be allocated a few node hours; the allocation period is one year and is renewable up to two times.
- **SHAPE Access**[13]: suitable for SMEs with the potential to use HPC. This access mode aims to help SMEs benefit from the expertise and knowledge developed within PRACE RI.
- **Distributed European Computing Initiative** (DECI): Suitable for smaller-scale projects that do not require Tier-0 systems. This access mode provides Tier-1 users access to supercomputing architectures from another European country for smaller-scale projects. Proposal submissions are accepted in response to annual calls.
- **Project Access** is intended for individual researchers and research groups and is suitable for established Tier-0 users. Access can be granted for 1-year production runs, as well as for 2-year or 3-year (multi-year access) production runs. Proposal submissions are accepted in response to biannual calls.
- The PRACE ICEI[14] is open to all European researchers and research organisations needing resource allocations, regardless of funding sources.

SYSTEMS AND CORE HOURS

System	Architecture	Site (Country)	Core Hours (node hours)	Minimum request (core hours)
HAWK*	HPE Apollo	GCS@HLRS (DE)	345.6 million (2.7 million)	100 million
Joliot-Curie KNL	BULL Sequana X1000	GENCI@CEA (FR)	37.5 million (0.6 million)	15 million
Joliot-Curie Rome	BULL Sequana XH2000	GENCI@CEA (FR)	195.3 million (1.5 million)	15 million
Joliot-Curie SKL	BULL Sequana X1000	GENCI@CEA (FR)	52.9 million (1.1 million)	15 million
JUWELS Booster*	BULL Sequana XH2000	GCS@JSC (DE)	85.2 million (1.78 million)	7 million Use of GPUs
JUWELS Cluster*	BULL Sequana X1000	GCS@JSC (DE)	35.04 million (0.73 million)	35 million
Marconi100	IBM Power 9 AC922 Whiterspoon	CINECA (IT)	165 million (1.87 million)	35 million Use of GPUs
MareNostrum 4*	Lenovo System	BSC (ES)	TBA	30 million
Piz Daint	Cray XC50 System	ETH Zurich/CSCS (CH)	510 million (7.5 million)	68 million Use of GPUs
SuperMUC-NG*	Lenovo ThinkSystem	GCS@LRZ (DE)	91 million	35 million

Figure 3: PRACE - Call for Proposals for Project Access. The provided figure illustrates the specifications of the systems included in the call, encompassing information such as name, location, and the range of available computing resources, denoted by minimum and maximum limits.

Figure 3 shows the PRACE call for proposals with the minimum number of core hours to be requested.

2.5.2 EuroHPC-JU Access Modes

- Extreme Scale Access (one-year or two-year projects)
- Regular Access (single-year projects)
- Benchmark Access are designed for code scalability tests, the outcome of which is to be included in the proposal in a future EuroHPC-JU Extreme Scale and Regular call.
- Development Access is designed for projects focusing on code and algorithm development and optimisation.
- Fast Track Access for Academia
- Fast Track Access for Industry Access

The calls are announced on the PRACE website[15]. Applicants interested in applying to any of

[11] https://prace-ri.eu/hpc-access/project-access/project-access-the-peer-review-process/

[12] https://prace-ri.eu/hpc-access/preparatory-access/preparatory-access-information-for-applicants/

[13] https://prace-ri.eu/hpc-access/shape-access/shape-access-information-for-applicants/

[14] https://prace-ri.eu/hpc-access/collaborative-calls/

[15] https://prace-ri.eu/hpc-access/eurohpc-access/

Access Mode	Extreme Scale	Regular	Benchmark	Development	Academic Fast Track	Industry Fast Track
Duration	1y renewable	1y renewable	2 to 3 months	1y renewable	< 6 months	1y renewable
Periodicity	Continuous calls, bi-yearly cut-offs	Continuous call, cut-offs every four months (3 cut-offs per year).	Continuous call, monthly cut-offs	Continuous call, monthly cut-offs	Continuous call, cut-offs ev. 2w/1m	Continuous call, cut-offs ev. 2w/1m
Share of resources	~70% Mostly pre-exascale	20 to 30% Mostly multi-petascale	Few % All systems	Few % All systems	~5% All systems	~5% All systems
Data storage needs	Large storage for medium to long term	Large storage for medium to long term	Limited	Data processing environment and platform		
Accessible to industry	Yes – Open R&D With specific evaluation criteria	Yes – Open R&D With specific track	Yes – Open R&D	Yes – Open R&D	No – use industry Fast Track instead	Exclusively Open R&D
External sc. Peer-review	Yes	Yes	No	No	No / Pre-identified	No / Pre-identified
Tech. assessment	Yes	Yes	Yes	Yes	Yes	Yes
Data Management Plan required	Yes	Yes	No	No	Yes	Yes
Application type	Full application	Full application	Technical application	Technical application	Light request + support documents	Full application
Prerequisite	Benchmark	Benchmark	None	None	Previous allocation or Benchmark	Benchmark
Submission period	> 2 months	> 2 months	N/A	N/A	N/A	N/A
Duration of evaluation process	3 months	2 months	≥1 week <2 weeks	≥1 week <2 weeks	≥2 weeks <1 month	≥2 weeks <1 month

Figure 4: EuroHPC-JU access modes. The diagram provides a comprehensive overview of the various elements pertaining to the access modes of EuroHPC-JU, including the duration of access, frequency of open calls, and prerequisites, among others.

the EuroHPC-JU calls need to apply via the PRACE peer-review platform[16]. The project scope and plan are also required for regular and extreme-scale access. More information about eligibility, access tracks, peer-review process, and scoring criteria can be found in the EuroHPC-JU access section on the PRACE website[17][18][19]. Figure 4 summarises the various EuroHPC-JU access modes. Figure 5 shows the EuroHPC-JU call for proposal with the minimum number of core hours to be requested.

2.6 HPC Users

EuroCC-JU classifies all the eligible users into the following categories:

- Academic users
- Industrial users
- Public Research Institutes

Researchers from academia, research institutes, public authorities, and industry established or located in a Member State or in a country associated with Horizon 2020/Horizon Europe are eligible to apply. Access to commercial companies and public organisations is provided solely for open R&D purposes.

[16]https://pracecalls.eu/

[17]https://prace-ri.eu/hpc-access/eurohpc-access/eurohpc-ju-regular-access-mode/regular-access-applicant-information/

[18]https://prace-ri.eu/hpc-access/eurohpc-access/eurohpc-extreme-scale-access/extreme-scale-applicant-information/

[19]https://prace-ri.eu/hpc-access/eurohpc-access/eurohpc-ju-benchmark-development-access-calls/benchmark-development-applicant-information/

System	Architecture	Site (Country)	Total Core Hours	Minimum request core hours
Vega CPU	BullSequana XH2000	IZUM Maribor (SI)	75 million	10 million
Vega GPU	BullSequana XH2000	IZUM Maribor (SI)	1.5 million	0.5 million
MeluXina CPU	BullSequana XH2000	LuxProvide (LU)	65.5 million	10 million
MeluXina GPU	BullSequana XH2000	LuxProvide (LU)	11.1 million	2 million
Karolina CPU	HPE Apollo 2000Gen10 Plus and HPE Apollo 6500	VSB-TUO, IT4Innovations, (CZ)	60 million	10 million
Karolina GPU	HPE Apollo 2000Gen10 Plus and HPE Apollo 6500	VSB-TUO, IT4Innovations, (CZ)	6 million	1 million
Discoverer CPU	BullSequana XH2000	Sofiatech, (BG)	104 million	10 million

Figure 5: EuroHPC-JU - Call for Proposals for Regular Access Mode. The provided figure illustrates the specifications of the systems included in the call, encompassing information such as name, location, and the range of available computing resources, denoted by minimum and maximum limits

3 Methodology

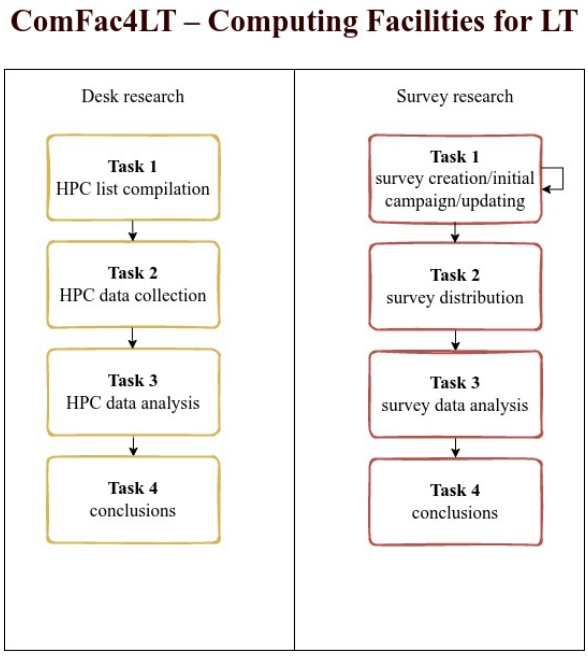


Figure 6: The methodology employed in this study encompasses two main approaches: desk-research and survey research.

The methodology used to study the various aspects related to the computing facilities for LT was performed using two distinct studies. The first part deals with desk research to study existing HPC facilities. This was accomplished by visiting each HPC's website and compiling information from the accompanying documentation, as well as contacting the appropriate authorities when necessary. This step allowed us

to concentrate on crucial aspects such as GPU specifications and access protocols. In the second part, a survey is conducted to study aspects related to HPC in practise. The survey captured the user's computational requirements as well as information about their computational facilities. The survey also gathered inquiries and comments from users regarding their existing HPC facilities. The survey method was chosen because it is critical to obtain up-to-date information on topics such as user hardware requirements and knowledge of EU-HPC initiatives.

3.1 Desk Research

A list of HPCs from the website Top500.org served as the seed list for the desk research. The Top500.org website publishes statistical lists of supercomputers twice a year. The website also includes metadata with the data releases, such as the HPC's location, ranking, and other hardware specifications. Two constraints were imposed to create the filtered list. First, only the European HPCs were retained. Second, HPCs belonging to the academic and research segments were retained. HPCs provided by vendor, private entities, others were not considered in the study as they did not relate to it directly. To this list, the supercomputers provided by EuroHPC-JU and PRACE (Partnership for Advanced Computing in Europe) were added. Given the nature of Tier-0 and Tier-1 HPCs being shared across EU member states and Horizon 2020-allied countries, we chose to focus on these supercomputers.

Desk research was conducted for each HPC in the fields listed below:

- Name
- Tier
- Performance (in petaflops)
- Location
- Hosting Institute
- HPC website link
- Types of access available to academic researchers, SMEs, and others
- Link to apply, register, or contact for the resources
- Manufacturer and specifications of GPU nodes, i.e., number of nodes, number of GPUs, and size of GPUs attached to each node
- Types of access provided as part of the institute: regular, benchmark, fast track, and others
- Additional notes or important points about the service

3.2 Online Survey

The survey, addressed to the LT researchers, sought to elicit the respondents' views to capture the real-world scenario. The survey had 11 questions in total. Two questions depended on previous answers. Table 1 shows an overview of the types of questions.

Table 1: Types of survey questions

Question types	Total
Closed	4
Open-ended	7
Total	11

None of the questions in the survey were mandatory.

- Part A. Respondent Profiling: The first part of the survey included questions for the demographic profiling of the respondents, with emphasis on
  - Country of the respondent
  - Active research field
  - Current role of the respondent
- Part B. HPC usage and requirements: assessed the current needs and usage patterns of computational facilities, i.e.,
  - HPC requirements for research and their specifics
  - Computational infrastructure
  - Hardware specifications in terms of processing time and memory
  - Awareness about Euro-HPC infrastructure
- Part C. Comments and suggestions: respondents' opinions, recommendations, and problems in relation to computational facilities

The survey was designed using Google Forms and underwent three iterations to capture verbose details. The survey was distributed by European Language Technology via a monthly newsletter in addition to mailing lists like META-NET-all, Corpora-List, MT-List, and In-Atala. The survey was open from March 7 to March 22. In total, 26 responses were collected. The responses collected as part of the survey, representing the views of the researcher in the field of LT, are analysed in the paper.

4 The HPC Landscape (desk research)

4.1 Analysis of HPCs

Given that performing language processing experiments necessitates the usage of GPUs, i.e., accelerated nodes, HPCs without GPUs were not considered. In total, the primary list contained 80 HPCs, but only 56 were analysed, as the ones that were filtered out were either offline or did not have GPUs.

4.1.1 Profile

In total, the final list accounts for 56 HPC from 20 EU member states, consisting of a mixture of various tiers. The distribution of tiers is depicted in Figure 7. The most common HPC level was Tier-1 (33) and then Tier-0 (15). This could be related to the predominance of EuroHPC-JU and PRACE supercomputers on the initial filtered list.

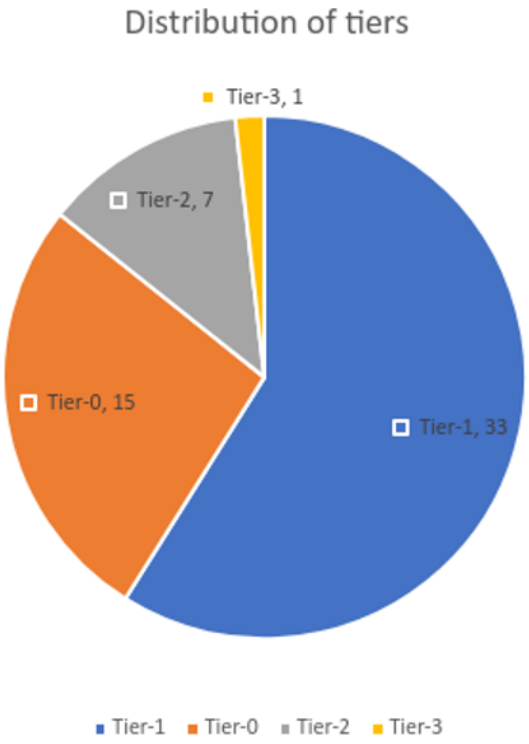


Figure 7: Distribution of tiers. This figure shows the distribution of tiers across different high-performance computing (HPC) systems.

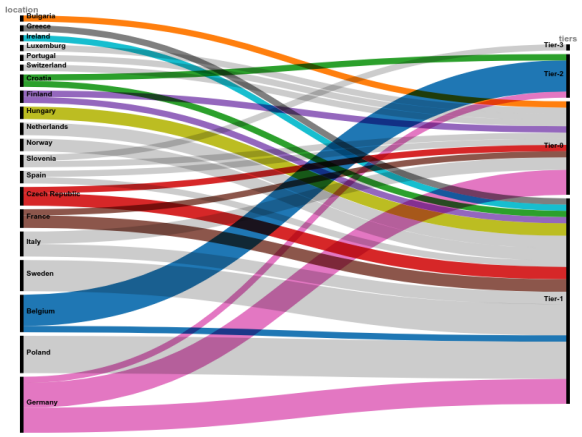


Figure 8: Country vs tiers. This diagram depicts the connections between the various tiers in various countries.

The relationship between countries and their tiers is shown in Figure 8. Germany had HPCs in all three tiers. France, Italy, Czech Republic, and Spain had Tier-0 and Tier-1 HPCs. The number of HPCs from each country that were analysed is shown in Figure 9. According to the HPC list, Germany had the most systems, followed by Poland, Belgium, Sweden, and other countries.

4.1.2 Hardware Performance

The countries are listed along with their overall HPC performance in Figure 10. With a performance of more than 25 petaflops, Finland is host to the third-fastest HPC LUMI in the world. Italy and France are ranked second and third. It was a clear finding that members of PRACE HPC hosting nations have higher cumulative performance than non-hosting nations. It is important to know that HPCs that are part of a consortium, such as LUMI, allocate their resources based on each country's contribution share.

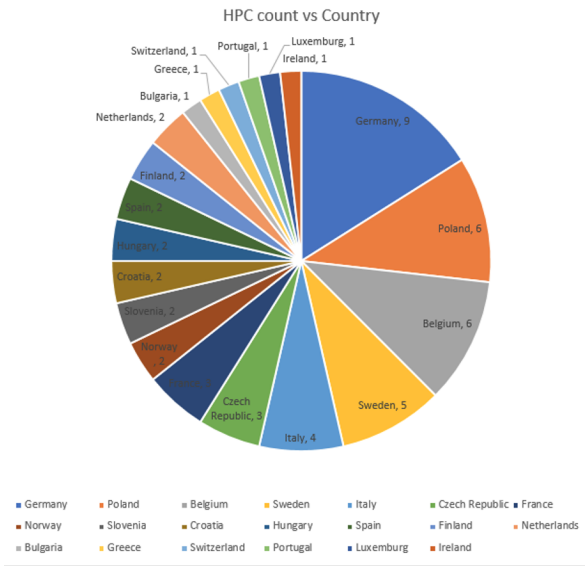


Figure 9: The figure depicts the countries, as well as the total number of HPCs that were included in the study.

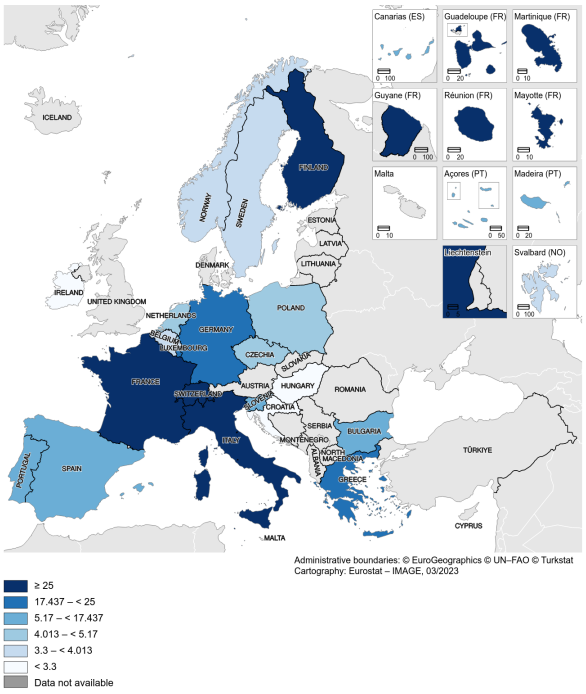


Figure 10: Countries and their cumulative performance. The figure presents focuses on the collective performance of various countries in terms of petaflops.

4.1.3 GPU Performance

This section analyses the GPUs used in HPC. In figure 12, we depict various GPU models. Nvidia V100 (28), A100 (23), and P100 were the most often installed graphics cards. There were six AMD Instinct cards in total. Nvidia CUDA with deep learning libraries such as PyTorch and TensorFlow enables easy access to GPU hardware, whereas ROCm is used to access AMD GPUs using such frameworks.

GPU cumulative VRAM is displayed per country in Figure 13. Finland, Italy, France, and Germany have the highest cumulative VRAM values. A comparison of the number of nodes vs. GPU cumulative VRAM is depicted in Figure 14.

4.1.4 HPCs Access

The HPC providers can also be grouped into the following categories:

- Open access to all researchers: An HPC provider grants open access to all the researchers linked to public research institutes like universities.
- Access to listed institutes: An HPC provider gives open access to universities or research institutes that have signed an agreement. For example, CSC's services[20] are free-of-charge for users affiliated with a Finnish higher education institution (universities, universities of applied sciences), or a state research institute. This is based on the agreement made with the Ministry of Education and Culture.
- Paid access or user contribution model: An HPC provider charges the users for the service based on the services being used.

In Figure 15, a breakdown of types of academic access is shown. The information shown pertains to non-PRACE access. In the vast majority of instances, academic users are granted free access.

Access to the industry can be classified into three categories. First, free access is granted if the research is publicly available and access is gained from the HPC service provider in the industry's respective country. Second, the HPC service provider does not offer access to industry users, but PRACE provides access. Third, access for commercial purposes is available to industrial users. The breakdown of access for industry is shown in Figure 16.

There is also strategic and discretionary access allocated for emergency-related work, such as research on pandemics such as COVID-19. Commercial access is typically an option for users that wish to utilise HPC services. Figure 17 depicts the breakdown of the different other types of access.

4.1.5 HPCs Access Calls

The calls to access computational resources for an HPC can be divided into national and international calls. International calls are handled through PRACE or a publicly funded project[21]. For the national-level calls, all the HPCs provide electronic means of submitting applications. The applications are usually accompanied by a project proposal listing requirements like a detailed plan of experiments, benchmarking scores, hardware needs, etc. The project proposal is then subjected to a technical feasibility assessment and scientific review.

The access call for HPCs can be classified as follows:

- **Experimental, benchmarking, or testing call**: The calls are usually open throughout the year and processed in a stepwise fashion. The call provides access to hardware with a few hours of computation

[20] https://research.csc.fi/free-of-charge-use-cases
[21] https://ni4os.eu/

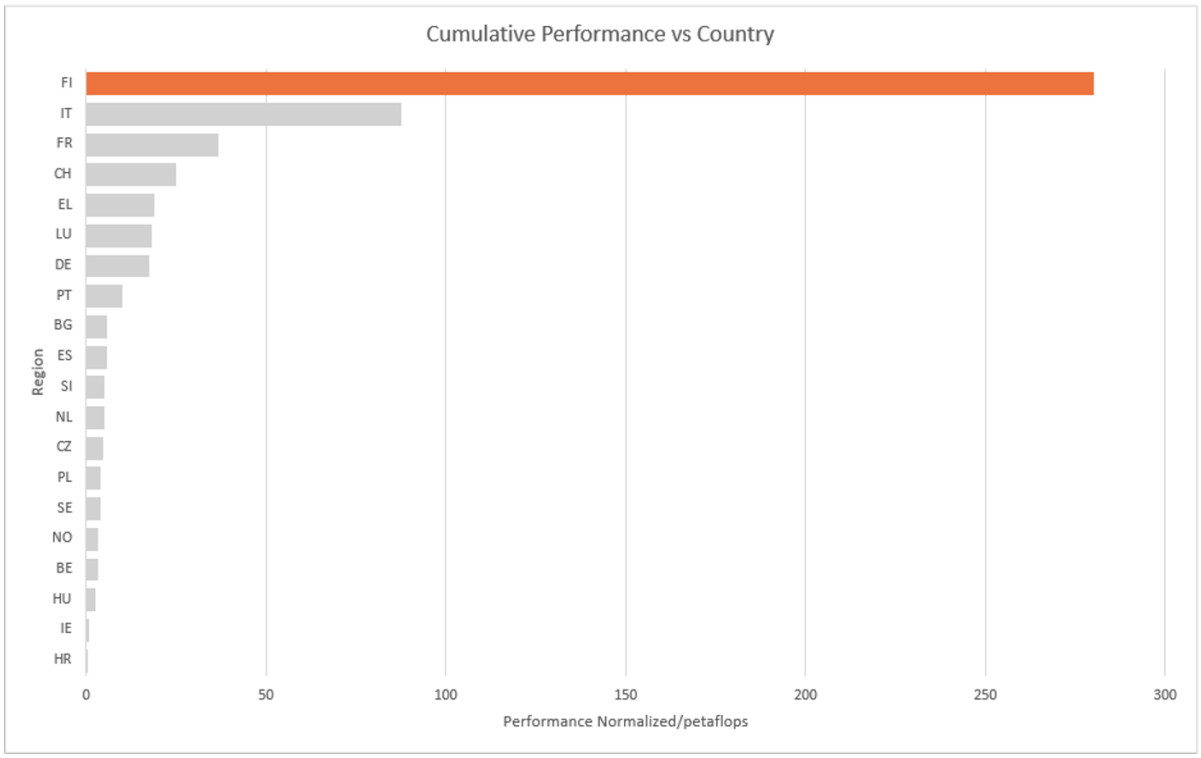


Figure 11: Countries and their cumulative performance

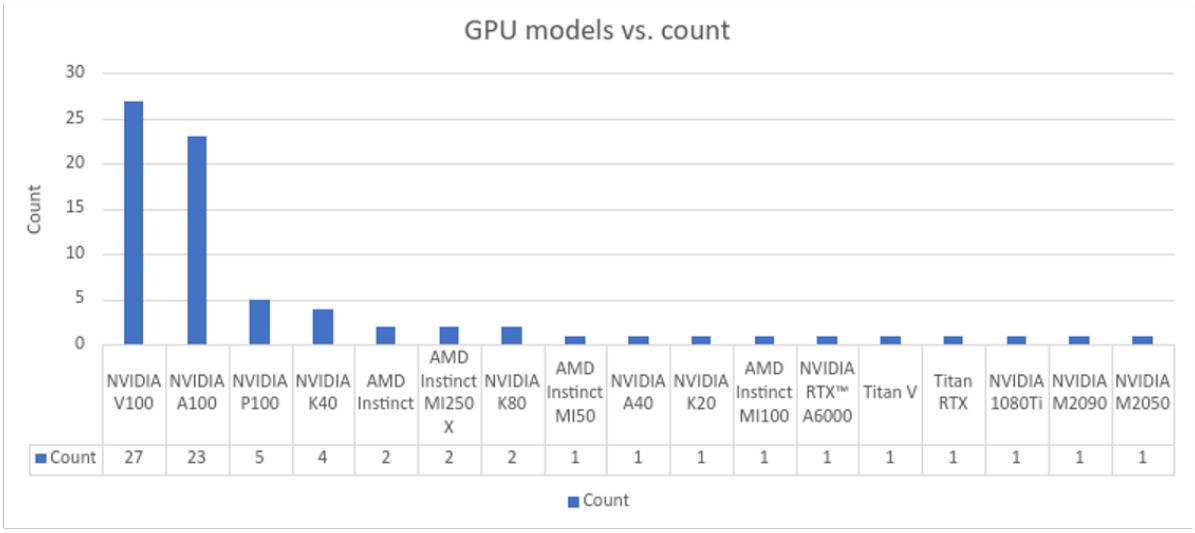


Figure 12: Model distribution of GPUs. The graph depicts the various GPU models and their counts, illustrating current GPU preference trends.

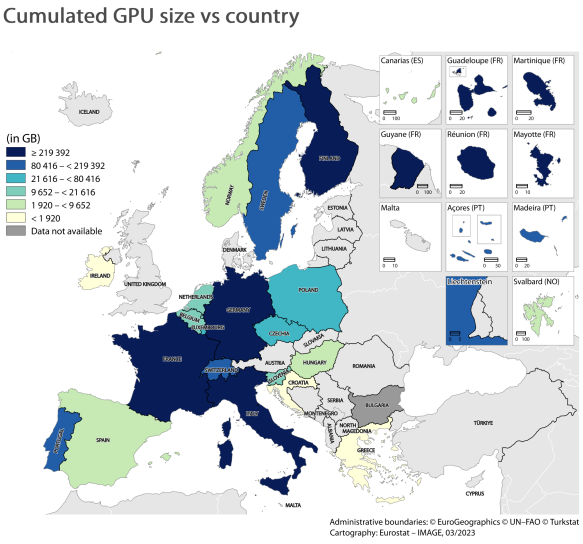


Figure 13: Countries and their cumulative GPU VRAM. Figure shows the availability of GPU memory in BG across various countries.

resources to test and compute the performance of the experiments. The benchmarking figures are then used in the application of regular access.

- **Fast track calls**: The calls are targeted towards users for projects that need fast access to HPC resources, which is limited in time and smaller in resources when compared with regular access call projects.
- **Regular calls**: The calls are opened for projects needing high-performance computational resources. The projects can last from 9 to 12 months, and calls are typically opened 2–4 times per year. Depending on the HPC service provider, requests can also be processed continuously. If the allocated resources are depleted, they can be extended.
- **Large-scale**: The call is similar to a regular access call but requires resources over a longer duration of time. Empirical estimates suggest that it could be more than 2% of total resources of the full HPC setup, computed over a year. The runtime of the projects ranges from 1 year to 3 years.
- **Director's Discretion/Discretionary Access**: A portion of the computational resources are reserved and made available upon project approval. An application can be submitted at any time. The computational resources are allocated irregularly based on evaluation by the management.
- **Extreme-scale**: The call is for the sectors to justify the need for and capacity to use extremely large allocations in terms of compute time, data storage, and support resources.

Each call has a processing period, which is the time it takes to look at the proposal and come to a decision. After this time, the applicants are given the resources they asked for.

4.1.6 Dynamic Access

The eDARI portal is used to request resource hours at French national computing centres. The portal allows two types of access to resources.

- Regular access
- Dynamic access

Depending on the number of requested hours, the requested access will be either Dynamic Access or Regular Access. If the number of hours requested is <= 50,000 GPU hours (and/or 500,000 CPU hours), it will be Dynamic Access (AD). If the amount is larger than these values, the request will be considered Regular Access (AR). The Dynamic Access skips the need for additional supplementary details. Requests for resources for Dynamic Access files may be made throughout the year and are renewable. Two project calls for Regular Access are launched each year.

This mode of access is discussed in a distinct section, since its general accessibility is so conducive to research. This access mode is a great choice due to the streamlined procedure and minimal documentation, particularly for CPU and GPU hours with moderate demands.

4.2 Comparison Study

In order to clarify the points made in the previous section, let us examine two countries in greater depth. Consider these two nations: Croatia, France.

At the time the study was conducted, two HPCs were publicly listed for Croatia, one for Tier 1 (SRCE) and the other for Tier 2 (BURA). There was one Tier 0 (Joliot-Curie IRENE) and two Tier 1s for France (Jean Zay, Adastra). The cumulative hardware performance of the French HPC exceeded 110 petaflops, while the cumulative performance of the Croatian HPC was close to 0.42 petaflops. The total GPU VRAM accessible in France remained at 273152 GB (number of GPU = 4616), whereas its Croatian counterpart recorded 480 GB (number of GPU = 24). In addition, France's LT research benefits from having dynamic access to a vast quantity of resources. It is critical to highlight that access to the Croatian HPC service (SRCE) is easier via the application portal. This disparity in the availability of HPC resources to researchers must be addressed if we are to realise the aim of language equality.

4.3 Summary

Figure 19 shows a summary of all the current HPC services that academic researchers and small and medium-sized enterprises (SME) can use. Users should use their local resources if they are available and appropriate. If the requirements are not very high, in most local HPC centres (Tier-2 and Tier-3), there is no need to write a

[22] https://www.edari.fr/schema/acces/ressource

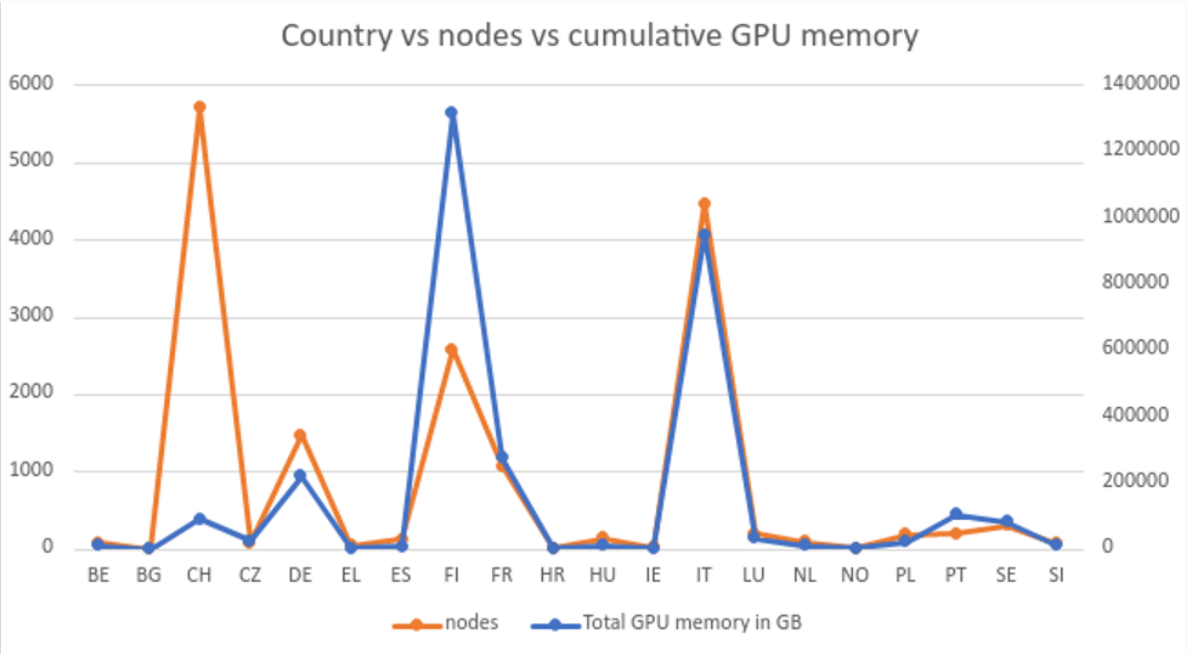


Figure 14: Countries with their total node count and GPU VRAM. The provided figure illustrates the quantity of nodes and GPU memory allocated to each country, effectively representing the sizes of their high-performance computing (HPC) nodes.

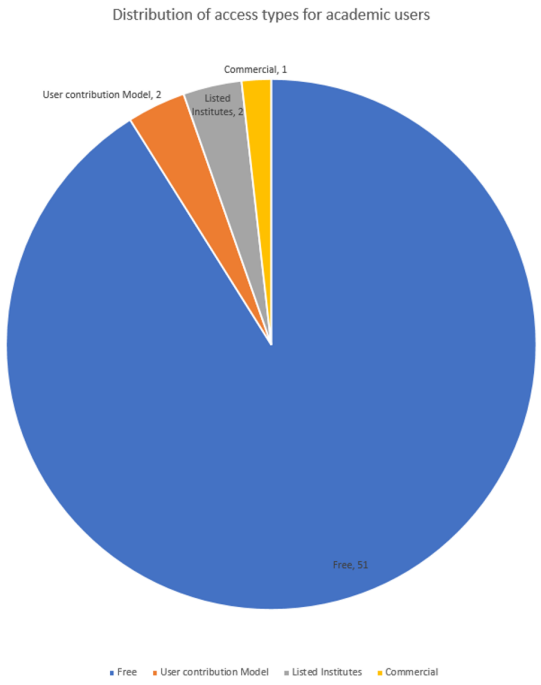


Figure 15: Distribution of different types of academic access. The majority of academic users' access is free of charge.

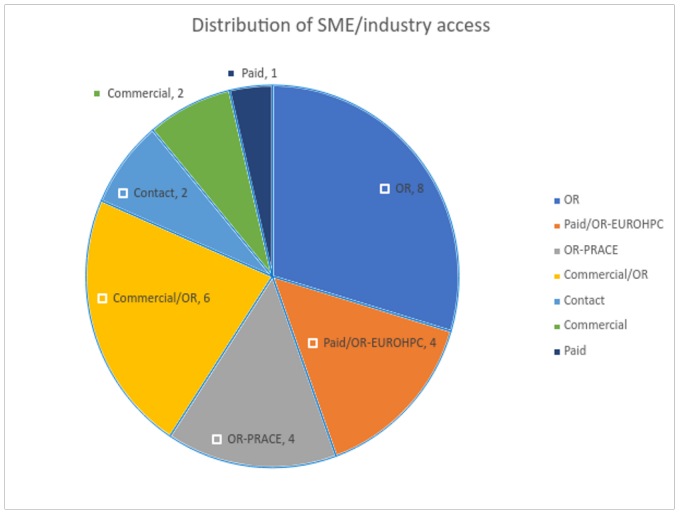


Figure 16: Distribution of different types of access for SME/industry users. The majority of SME/Industry users' access is paid. The free access is provided when SME users are involved in open research.

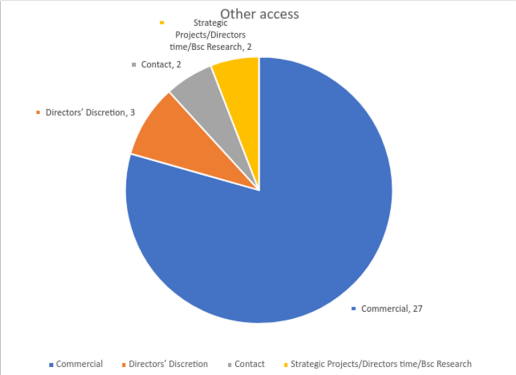


Figure 17: Distribution of other types of access. Access to the High-Performance Computing (HPC) service is exclusively granted by the governing authorities, specifically to users who have submitted a specialised application.

detailed application, as all the researchers are provided with fair-share quota access. If you require more resources than your centre can provide, and you don't have a local HPC centre, or you identify special needs (e.g., larger memory, more Cores/CPU, GPUs), you may contact another HPC centre or apply for compute time at a higher level (e.g., Tier-2/Tier-1). Only very experienced users with well-scaling codes and high demands on compute time should apply for large-scale projects on the Tier-1/Tier-0 level. In any case, you can contact your local HPC support with your queries. If the research is open, which means that the results will be available to the public after publication, a researcher from industry can work with academics or apply to the resources on their own through PRACE. In the case where research is private, the option of commercial access to resources provided by private vendors and other HPC providers is available.

5 Analysis

5.1 Survey Responses

5.1.1 Respondents' Profile

The majority of the answers came from European countries, except for a few. States covered via the survey include: Croatia, Denmark, Finland, France, Germany, Ireland, Italy, Luxembourg, Malta, Portugal, Spain, the UK, Pakistan, the USA, Ukraine, Russia.

Table 2 shows the breakdown of answers.

The majority of respondents had LT as the active area of research, with 25 actively associated with NLP and 1 marking themselves as not active NLP researchers. Table 3 shows the respondents associated with the area of NLP/LT.

Table 2: The geographical distribution of the survey participants

Country	Count	Country	Count
Croatia	2	Denmark	2
Finland	3	France	5
Germany	1	Ireland	1
Italy	1	Luxembourg	1
Malta	1	Pakistan	1
Portugal	1	Russia	1
Spain	2	UK	1
Ukraine	1	USA	1

Table 3: Participant's active area of research. "Is NLP/LT your active area of research".

Answers	Total
Yes	25
No	1

Most of the individuals who responded are either academic researchers or students. One respondent identified himself as a public sector researcher. No responses from the researchers working in industry were received. The breakdown of respondent associations is shown in Table 4.

Table 4: Respondents' role. "What is your current role ?".

Roles	Total
Researcher-Academia	17
Student	8
Researcher/developer - Public sector	1

5.1.2 Respondents' LT Infrastructure and Requirements

The majority of the survey respondents reported using HPCs for their experiments. The responses can be classified into three types.

- HPC users
- Cloud service users like Google Colab
- Local hardware or personal computer users

Respondents who didn't use an HPC said they did the experiments on a single GPU or a setup with more than one CPU. When asked how many GPU hours were needed, the answers ranged from "it depends on the experiment" to a precise number that suggested a certain number of hours per day, week, or month. Regarding the multi-GPU requirement, most respondents wished to use more GPUs, especially for the task of machine translation (text and speech). 50% replied they do not have the multi-GPU requirement, while 50% reported they do wish to use more GPUs. Another question was posed regarding memory requirements, and a variety of responses were provided.

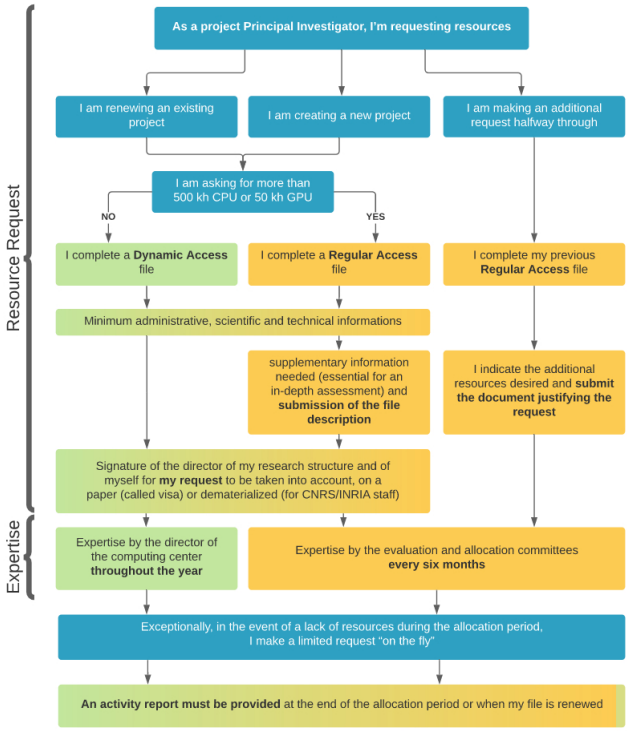


Figure 18: Reference card for access via edari.fr [22]. The provided diagram serves as a reference for individuals who are seeking to apply for High-Performance Computing (HPC) services.

5.1.3 EuroHPC-JU Usage

On the question of using the resources from EuroHPC-JU, the majority of answers suggested not using EuroHPC-JU. 57% reported never hearing about it, while 34.6% responded, negating the use of the service.

Table 5: EuroHPC-JU usage. "Have you used resources from EuroHPC-JU?".

Answer	Total
Yes	2
No	9
Never heard of it	15

5.1.4 Respondents' Comments, Suggestions, and Recommendations

The open-ended question capturing the comments and suggestions with respect to the computational facilities used by users presented multiple aspects, which can be described in the following points.

- one of the respondents said, "Euro HPC applications are extremely heavyweight and not fit for our field". Aspects like "time to solution" and terminology, like simulation, which relates to the field of physics, used during the application process introduce non-conformity.
- users' access to HPC is temporary and linked to a project.
- unavailability due to the number of GPUs and more users.
- opacity with respect to job scheduling

5.2 Summary

Although only 26 responses were received, the study can be considered a preliminary step in mapping the current HPC facilities available to the LT researchers. The country respondents, such as those from Finland, France, Italy, and Germany, use the HPC resources collected and analysed during desk research. The majority of respondents had GPU requirements of fewer than 100 hours per month on average. Concerning, however, is the lack of knowledge of the European-level HPC services available to users. The conditions set by the responders do not correspond to the minimum amount of hours required to be requested in the PRACE calls.

6 Conclusions

HPC services enable the solution of computational tasks at a rate exponentially greater than that of a desktop computer. These services have existed and have proven to be crucial in advancing the state of the art in LT. There are numerous projects that provide HPC nowadays, like EuroHPC-JU, PRACE, LUMI, national

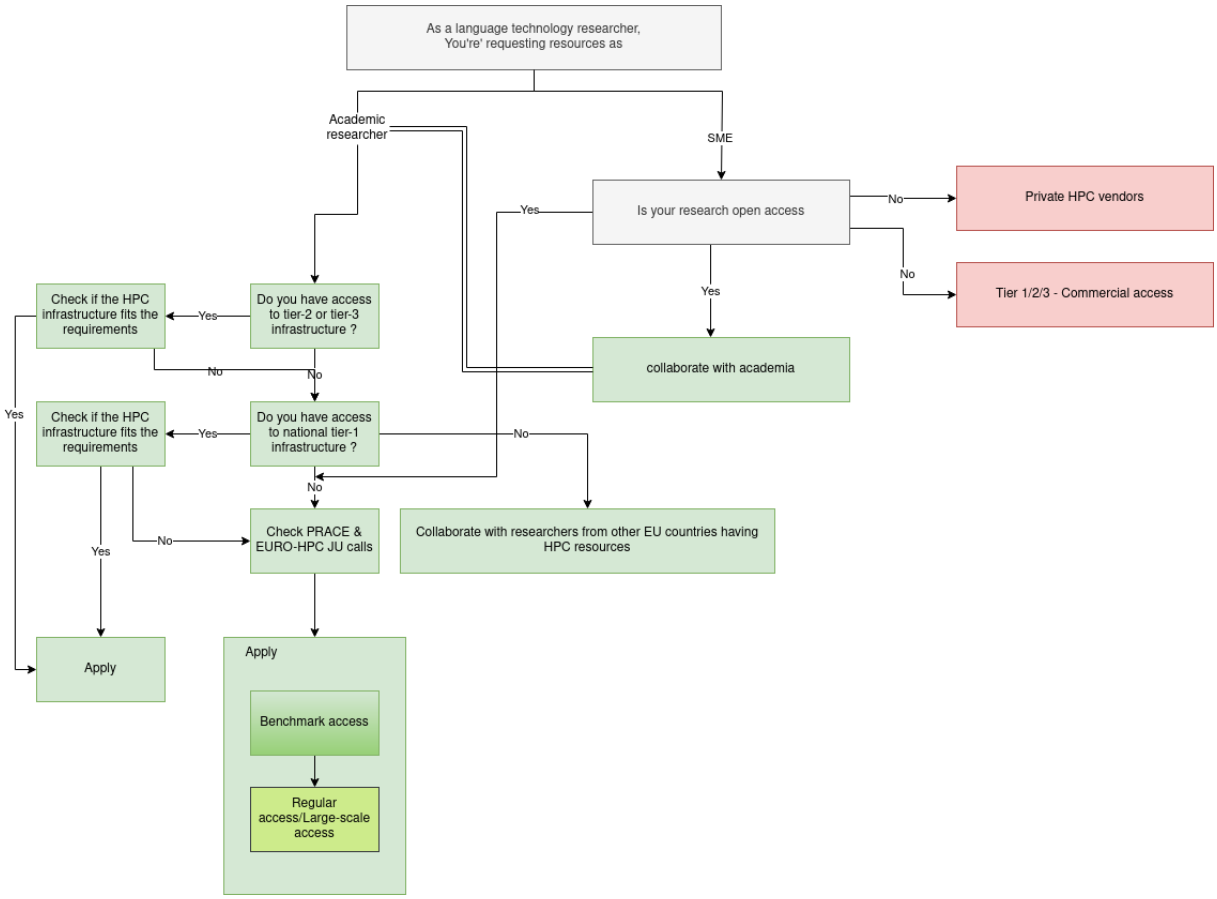


Figure 19: Overall summary. The provided diagram compiles all the aspects studied in the desk-research and serves as a reference for individuals who are seeking to apply for High-Performance Computing (HPC) services.

consortia, etc. Yet, for the subject of LT, unified research on many facets of HPCs was required.

In this paper, we give an overview of HPC services for LT research. We focused on elements such as available hardware, access types, and the requirements associated with each access type. In addition, we provide a simple reference card to be followed while seeking HPC services.

Before going into detail with all the conclusions of our analysis, we emphasise two points that, in our opinion, will be particularly critical to ensuring digital language equality in Europe:

- HPCs are important for LT research and development. Thus, competence with HPCs is a fundamental requirement for language equality[23].
- Availability of HPC for smaller and larger requirements is crucial. Users can access the European, national, and regional HPC services. Thus, efforts should be focused on facilitating easier and quicker access for these users.

Next, we present a summary of the key insights and recommendations regarding HPC services and access to them in the context of digital language equality in Europe.

[23]More about the European Language Equality project you can find at https://www.european-language-equality.eu.

- **Access to HPC resources for light-weight requirements**. As seen previously, the PRACE and EuroHPC-JU calls demand very high minimum node hours in the request. Although EuroHPC-JU offers academic fast-access, these calls are difficult to get. As previously indicated, dynamic access is an excellent solution for providing easy and quick access to requirements that are not demanding in terms of node hours. Hence, we suggest an access mode similar to dynamic access to speed up the process of resource request and allocation.
- **Collaboration within the EU community and SMEs**. An alternate way of accessing HPC resources from EU countries is through collaboration. The LT community should provide the required tools for collaboration, particularly with nations lacking HPC resources. This would be advantageous not only for academic researchers, but also for industry users.
- **Centralise access to HPC related information**. Websites like https://atlas-cric-dev.cern.ch/core/rcsite/list/ and https://gauss-allianz.de/en/hpc-ecosystem give centralised information on HPCs accessible in the country, including Tier classification and hardware specifications. Our desk study helps move in this direction, but we recommend a centralised website that would allow

users to locate and filter HPC prospects based on requirements and particular criteria.

- **Hardware absence == No LT exploration**. Last but not least, a relatively uniform image can be obtained from the survey of LT users about HPC usage. We can fairly assume that the LT researchers will stick to tasks that fit their current hardware availability rather than anything else. For instance, if a GPU is capable of fine-tuning a model, a researcher is more likely to pursue fine-tuning than machine translation or language modelling. Even in the case where the minimum hardware is available, users fiddle with hyperparameters like batch-size to finish training. This does increase the overall time required as compared to using an HPC service. Another issue related to this point is the capacity to execute inference[24] on LLMs such as BLOOM. As the model needs 352 GB in bf16 (bfloat16) weights (176*2), the most efficient set-up is 8×80 GB A100 GPUs. Also, 2×8×40 GB A100s or 2×8×48 GB A6000 can be used. The inability to employ these LLM models without access to numerous GPUs does provide a challenge. Mosaic ML[25] makes it easy to train a billion parameter models in hours instead of days, with no lock-in to a single vendor and coordination across multiple clouds. With the ability to scale across multiple providers, the OOM can be prevented. We recommend such an infrastructure be realised in the context of EuroHPC-JU and PRACE systems to cater to the dynamic needs of various NLP tasks from different strata of LT users.

Finally, HPC services are available at multiple levels (regional, national, and European). At the same time, addressing the challenges related to the availability and accessibility of HPC is of the utmost priority. Our hope is that this document provides enough overview and insights into existing HPC resources available to academia and industry. We also hope our recommendation in the final section will provide enough pointers for the stakeholders to plan and implement future steps effectively. Private providers such as Azure and Amazon are significant players in the LT market because they offer commercial access to a huge number of GPUs. This type of access is feasible with sufficient funds. This element was not addressed in this study; it's expected to be the subject of future research.

7 Limitations

Following are some of the study's immediate limitations:

- The list of HPCs is not exhaustive. For the desk research, a curated list of HPCs was compiled from the websites top500.org, PRACE, and EuroHPC-JU. Hence, the majority of the analysed systems were either Tier-0 or Tier-1, with fewer Tier-2 and even fewer Tier-3 systems. Thus, our observations and reasoning may be influenced by the HPCs analysed. Covering the EuroHPC-JU and PRACE systems does provide a European-level perspective, but country-specific observations cannot be confirmed with the same degree of certainty. During the time of data compilation, a number of new HPC systems[26] became operational and were not included in the analysis.
- The survey was intended to reach a larger audience, but due to time constraints, the sample size was insufficient, making it difficult to generalise findings to a larger group.
- In comparison to the number of LT researchers in the EU, the size of the survey's sample is significantly smaller. There were no answers from industry researchers, who may have had different computational needs or perspectives on HPC usage and access.

[24]https://huggingface.co/blog/bloom-inference-pytorch-scripts
[25]https://www.mosaicml.com/platform

Acknowledgments

The work presented in this paper has received funding from the European Union under the grant agreements No. LC-01641480 – 101018166 (ELE) and No. LC-01884166 – 101075356 (ELE 2).

References

Ahmed, N. M. and Wahed, M. (2020). The de-democratization of ai: Deep learning and the compute divide in artificial intelligence research. *ArXiv*, abs/2010.15581.

Asch, M., Moore, T., Badia, R., Beck, M., Beckman, P., Bidot, T., Bodin, F., Cappello, F., Choudhary, A., De Supinski, B., et al. (2018). Big data and extreme-scale computing: Pathways to convergence-toward a shaping strategy for a future software and data ecosystem for scientific inquiry. *The International Journal of High Performance Computing Applications*, 32(4):435–479.

Berberich, F., Liebmann, J., Nominé, J.-P., Pineda, O., Segers, P., and Teodor, V. (2019). European hpc landscape. In *2019 15th International Conference on eScience (eScience)*, pages 471–478. IEEE.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language

[26]https://www.tportal.hr/tehno/clanak/u-zagrebu-predstavljeno-najjace-super-racunalo-u-hrvatskoj-pokrenuta-nova-generacija-nacionalne-e-infrastrukture-hr-zoo-foto-20230328

models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dolbeau, R. (2018). Theoretical peak flops per instruction set: a tutorial. *The Journal of Supercomputing*, 74(3):1341–1377.

Eicker, N. et al. (2020). Performance monitoring and analysis of eurohpc supercomputers. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (SC '20)*.

Hutton, R. et al. (2019). Prace: A european hpc ecosystem. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (SC '19)*.

Jiang, Z., Gao, W., Tang, F., Wang, L., Xiong, X., Luo, C., Lan, C., Li, H., and Zhan, J. (2021). Hpc ai500 v2. 0: The methodology, tools, and metrics for benchmarking hpc ai systems. In *2021 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 47–58. IEEE.

Jiang, Z., Gao, W., Wang, L., Xiong, X., Zhang, Y., Wen, X., Luo, C., Ye, H., Lu, X., Zhang, Y., et al. (2019). Hpc ai500: a benchmark suite for hpc ai systems. In *Benchmarking, Measuring, and Optimizing: First BenchCouncil International Symposium, Bench 2018, Seattle, WA, USA, December 10-13, 2018, Revised Selected Papers 1*, pages 10–22. Springer.

Lathrop, S., Mendes, C., Enos, J., Bode, B., Bauer, G., Sisneros, R., and Kramer, W. (2019). Best practices for management and operation of large hpc installations. *Concurrency and Computation: Practice and Experience*, 31(16):e5069.

Luszczek, P., Dongarra, J. J., Koester, D., Rabenseifner, R., Lucas, B., Kepner, J., McCalpin, J., Bailey, D., and Takahashi, D. (2005). Introduction to the hpc challenge benchmark suite. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

OpenAI (2023). GPT-4 technical report. *CoRR*, abs/2303.08774.

Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilic, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S., Webson, A., Ammanamanchi, P. S., Wang, T., Sagot, B., Muennighoff, N., del Moral, A. V., Ruwase, O., Bawden, R., Bekman, S., McMillan-Major, A., Beltagy, I., Nguyen, H., Saulnier, L., Tan, S., Suarez, P. O., Sanh, V., Laurençon, H., Jernite, Y., Launay, J., Mitchell, M., Raffel, C., Gokaslan, A., Simhi, A., Soroa, A., Aji, A. F., Alfassy, A., Rogers, A., Nitzav, A. K., Xu, C., Mou, C., Emezue, C., Klamm, C., Leong, C., van Strien, D., Adelani, D. I., and et al. (2022). BLOOM: A 176b-parameter open-access multilingual language model. *CoRR*, abs/2211.05100.

Skordas, T. (2019). Toward a european exascale ecosystem: the eurohpc joint undertaking. *Communications of the ACM*, 62(4):70–70.

Truong, H.-L., Fahringer, T., Madsen, G., Malony, A. D., Moritsch, H., and Shende, S. (2001). On using scalea for performance analysis of distributed and parallel programs. In *Proceedings of the 2001 ACM/IEEE conference on Supercomputing*, pages 34–34.

Wolter, N., McCracken, M. O., Snavely, A., Hochstein, L., Nakamura, T., and Basili, V. (2006). What's working in hpc: Investigating hpc user behavior and productivity. *CTWatch Quarterly*, 2(4A):9–17.

Utilization of Open-Source in Upgrading Technological Devices for Industry 4.0 readiness

Gabriel Gaspar, Juraj Dudak
University of Zilina
Research Centre
Univerzitna 8215/1, 010 26 Zilina, Slovakia
Slovak University of Technology in Bratislava
Faculty of Materials Science and Technology in Trnava
Jana Bottu 25, 917 24 Trnava, Slovakia
{gabriel.gaspar, juraj.dudak}@uniza.sk
{gabriel.gaspar, juraj.dudak}@stuba.sk

Ivan Sladek, Ingrid Zuziakova
University of Zilina
Research Centre
Univerzitna 8215/1, 010 26 Zilina, Slovakia
{ivan.sladek, ingrid.zuziakova}@uniza.sk

Tomas Meravy
Slovak University of Technology in Bratislava
Faculty of Materials Science and Technology in Trnava
Jana Bottu 25, 917 24 Trnava, Slovakia
tomas.meravy@stuba.sk

Abstract*. The concept of Industry 4.0 currently represents the first choice according to which industrial equipment is designed and implemented. There are visible trends in integrating open-source solutions into the industry. At the same time, the advancement in science and technology development has decreased prices for sensing, processing, and sharing technologies. Retrofitting existing equipment brings additional opportunities for incorporation into factory automation. In this contribution, the focus was put on upgrading an older climate chamber to meet current requirements utilizing the Node-RED solution. The solution fully replaces the legacy control system while providing options for low-cost additional hardware and software extensions.*

Keywords. open-source, Industry 4.0, Node-RED, sensor, PLC.

1 Introduction

The concept of Industry 4.0 currently represents the first choice according to which industrial equipment is designed and implemented. Smart devices and automation are the trend of the times and will continue to expand unstoppably.

This paper aims to propose the concept of using open-source solutions to upgrade existing technological equipment. Many such devices have become obsolete during their many years of usage, and it is impossible to upgrade their automation level to a higher level easily. Such an upgrade involves a complete replacement of the control system and possibly further technological enhancements in the areas of sensors, actuators, and drives. Many advanced technologies are available and already in use. For example, the authors of this paper dealt with using 1-wire networks for industrial applications. Later they designed a universal module for connecting various sensors (Dudak et al., 2018).

However, several factors nevertheless make such equipment upgrades advantageous:

- can still carry out its function,
- simple repairs,
- spare parts availability,
- economic sense of replacement with new equipment.

The upgrade of the control system of an older climate chamber used by a local company in the Slovak Republic was chosen as a practical example for this proposal. It is a device with analog control, which does not allow for programmed heating and cooling. The device also does not allow remote management, control, or reporting of operating states.

Climate chambers, also known as environmental test chambers or climatic chambers, are specialized enclosures used to simulate and control various environmental conditions such as temperature, humidity, light, and sometimes other factors like

pressure or vibration. The primary purpose of climate chambers is to replicate real-world conditions in a controlled setting, allowing researchers and engineers to assess how materials, components, or products perform under specific environmental conditions. By subjecting items to different temperature and humidity levels, for example, manufacturers can ensure their products can withstand extreme climates or identify weaknesses that might lead to failure.

The proposed design aims to simplify operators' use and validate open-source technologies applicable to the company's products. The focus was put on using easily accessible and cost-effective solutions so that the results would be widely applicable in the small and medium-sized enterprises (SME) sector, for which significant investments are not always possible.

2 Related work

There are now visible trends in integrating open-source solutions into the industry, which was proved in authors' previous research on identifying essential attributes and needs for small and medium-sized enterprises (Barton et al., 2022). One of the reasons for this is the drop in price and the widespread availability of suitable hardware components. As the authors note, industrial and consumer internet of things (IoT) solutions are becoming economically and technologically viable, and in some instances, they are implementable in highly automated lines. At the same time, the general advancement in science and technology development has decreased prices for sensing, processing, and sharing technologies. Thereby, IoT technologies are now widely adopted in many commercial and industrial products (Martikkala et al., 2021). The involvement of open-source solutions in industrial applications is further discussed by the authors (Rathee & Chobe, 2022), who argue that industries that have been around for more than a century, like automotive and agriculture, and the industries formed much more recently, like aerospace, have all embraced open-source technologies and there is hardly any industry that has not dabbled in open-source software. Another example of open-source utilization is implementing a supervisory control and data acquisition (SCADA) system for a hybrid power system based on the Node-RED solution (Omidi et al., 2023). The authors declare their study to be a pioneer in developing open-source SCADA systems that are low-cost and low-power in their design. In most cases, SCADA systems are made up of hardware and software components that collaborate to gather and analyze data from sensors, control devices, and other sources before giving operators a graphical user interface (GUI) for managing and overseeing these operations. Due to their many benefits, open-source solutions have grown in favor recently, even among SCADA systems. Software with publicly available source code, allowing anybody to examine, alter, and distribute the code, is called "open source". In contrast the proprietary software, whose owner is a particular entity, does not provide source code to the general public. Typically, a community of individuals or organizations that work together and contribute to the codebase develops open-source software.

These characteristics can be utilized and, on their basis, can be concluded:

1. **Cost-Effective:** Compared to proprietary solutions, open-source SCADA systems are frequently more affordable, which is one of its main advantages. Traditional SCADA systems may require pricey licensing, hardware, and continuing maintenance fees, making them costly to adopt and operate. Contrarily, open-source SCADA systems are frequently cost-free to operate, require no license costs, and can function on affordable hardware, making them more affordable for enterprises on a tight budget.
2. **Customization and Flexibility:** Open-source SCADA systems allow program modification to satisfy particular needs. Companies can alter the source code to add or remove features, combine the software with other systems, or customize it to suit their requirements. In proprietary SCADA systems, where companies may have little control over the software's functionality and may need to rely on the vendor for modification, this level of customization is not often accessible.
3. **Transparency and Security:** Since the source code is made accessible for public inspection by the community, open-source SCADA systems provide transparency. It makes it possible for impartial security audits, which can assist in finding and fixing flaws and guarantee the security and dependability of the software. Because the source code is available, security upgrades and patches are not dependent on a single vendor, which might give more control over security precautions.
4. **Interoperability and Integration:** Open protocols and standards are frequently used to design open-source SCADA systems, which can help with interoperability and device integration. Enterprises can link the SCADA system with their current infrastructure, such as programmable logical controllers (PLC), remote terminal units (RTU), human-machine interfaces (HMI), and other industrial devices, without being constrained by proprietary protocols or interfaces.
5. **Community Support:** Open-source SCADA systems frequently have a vibrant user and development base that offers support via forums, mailing lists, and other platforms. It can give businesses access to a plethora of information and know-how, as well as plugins, modules, and documentation made by the community.

Open-source SCADA systems have several benefits: affordability, adaptability, customization, transparency, security, interoperability, and community support. Before selecting an open-source SCADA system, organizations should carefully assess their unique needs and available resources. It is because these systems may have drawbacks like a potential lack of commercial support, varying levels of user-friendliness, and potential differences in features from proprietary systems.

Retrofitting existing machines into Industry 4.0-ready equipment also brings additional opportunities for incorporation into factory automation. Those are, in particular, M2M communication options, with many applications one can adopt in any domain (Chakravarthi, V.S., 2021). Another example is that deploying such new technologies continuously increases demands for real-time wireless large data collection and acquisition (Big data), emphasizing simplicity and speed (Vagas et al., 2019).

3 Original equipment status and upgrade options analysis

The climatic chamber was custom-built for low temperature curing of potting compounds up to 100°C about 20 years ago. The work procedure is as follows:

1. The potting compound is prepared according to the datasheet. It is used for potting of electronic products, specifically specialized sensors.
2. Products are then placed in the climate chamber, operating parameters are set, and the process of curing the compound is started.
3. After a defined time, the products are unloaded, and the curing quality and the functional test are checked.
4. If the product is finalized according to the manufacturer's requirements, it is further packed and ready for shipment from the factory.

The climate chamber body chassis is made of stainless-steel plates for easy maintenance and cleaning. Heating was provided using heating elements connected in parallel, which were located on three sides of the chamber (excluding the door). A fan homogenizes the temperature inside the chamber. The schematic layout of the original chamber design is shown in Figure 1.

Such a device is a typical representative of the time of its creation, focused on simplicity of control solutions without the need for integration into other systems or other communication features. Due to the nature of the use and the high financial cost of such extensions at the time, the equipment was used until last year in its original state. The high entry costs of proprietary solutions have been reduced over time due to the rapid development of open-source solutions.

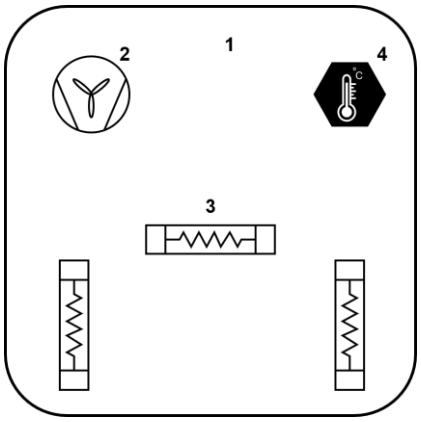


Figure 1. Chamber block schematics
1. Chamber chassis, 2. Fan, 3. Heating elements, 4. Thermometer

Open-source solutions have also been marginalized in the past due to their low uptake and lack of support. This concern is no longer valid today, and many suitable solutions are available as free software. An example is the Linux OS, several distributions of which are broadly used as both server and desktop solutions. In addition, the deployment of Linux-based systems is often seen in commercial manufacturers' solutions (SIMATIC Industrial OS - The Linux-based operating system for SIMATIC IPC - ID: 109766374 - Industry Support Siemens, 2019). A significant advantage is that hardware manufacturers also support open-source solutions, often providing their devices with pre-installed or easy-to-install open-source software. An example is the Raspberry Pi single-board computer (Teach, learn, and make with the Raspberry Pi Foundation, 2023), which is often the first choice for deployment in prototypes and final solutions. It is due to its acquisition price, the number of peripherals, and the broad client support from the manufacturer and the community. A suitable local example is the Unipi family of PLC controllers (Unipi, 2023), widely applicable in several areas ranging from home and building automation to industrial solutions. It is also preferable to use an open-source solution for programming the devices.

Since there was PLC Unipy Axon S155 on hand, it was decided to use this hardware with available heating elements, sensors, and other parts to implement the solution. The following chapter describes its set-up, hardware wiring, and programming.

4 Upgrade implementation

The proposal for upgrading the climate chamber was already implemented during several phases in 2022. The design also considered the students' final theses (Fabusova, J., 2022), (Hlavatovic, J., 2022). In this solution was used PLC Unipi Axon S155 as a controller for handling inputs, outputs, and

communication with the user shown in a proposed block schematics in Figure 2.

The digital output of the PLC was used to control the switching of the heater. In the same way, for the fan switching, the digital output of the PLC utilizing pulse width modulation (PWM) was used. The PWM was used as a typical digital output technique to regulate the energy supplied to various types of electrical equipment. In this case, to control the fan speed.

A digital PLC input was used to operate the emergency stop button. A 1-wire digital communication interface was used to operate a 1-wire bus on which 4 DS18B20 digital thermometers were connected. Sensors installed at multiple locations inside the chamber measure the temperature distribution. Based on these measurements, the heater and the fan circuits will engage. An Ethernet interface was used for communication with the user, of which two are available on the PLC.

The diagram in Figure 3 shows the basic chamber control algorithm. The operating parameters are initially set, including curing time, curing temperature, fan run time, and dead band. Consequently, the closed-door check is performed, followed by the emergency safety button check. If the checks are successful, the temperature in the chamber is measured. It is followed by a check of the elapsed curing time, and if the time has elapsed, the process ends, and the chamber performs a controlled shutdown. If the time has not yet passed, another temperature comparison occurs. Here are two options again; if the temperature is lower than the set curing temperature, the heating element is turned on with the fan to swirl the air in the chamber to homogenize the chamber temperature.

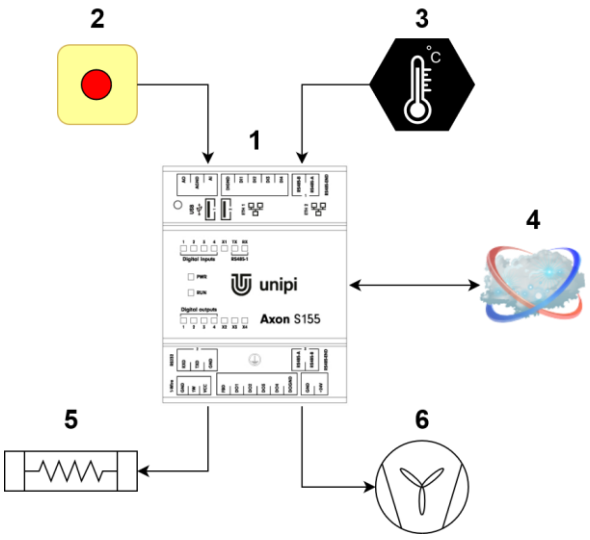


Figure 2. Block diagram of chamber components
1. PLC Axon S155, 2. An emergency stop button, 3. Thermometers, 4. Network connection, 5. Heating elements, 6. Fan.

However, suppose the temperature is higher than the set curing temperature. In that case, the heating element and the fan will turn off after the set cooldown time of the fan since the heater will not stop heating immediately after disconnecting. This process continues until the set curing time has elapsed.

The Node-RED solution (Node-RED, 2023) was chosen as an open-source control and data visualization solution. Node-RED is a tool for programming event-driven applications and provides interconnection between hardware, devices, APIs, and various online services. With the help of the extensive selection of nodes in the palette, Node- RED's browser-based flow editor makes it simple to link flows together. The event-driven, non-blocking paradigm of Node.js is fully utilized in the runtime's construction. Because of this, solutions can operate both in the cloud and at the network's edge on inexpensive hardware such as Raspberry Pi's already mentioned. It is simple to expand the selection of palette nodes to include new functions because the package repository contains more than 225,000 modules. Node-RED flows are saved using JSON format, which is simple to integrate and export for sharing with others.

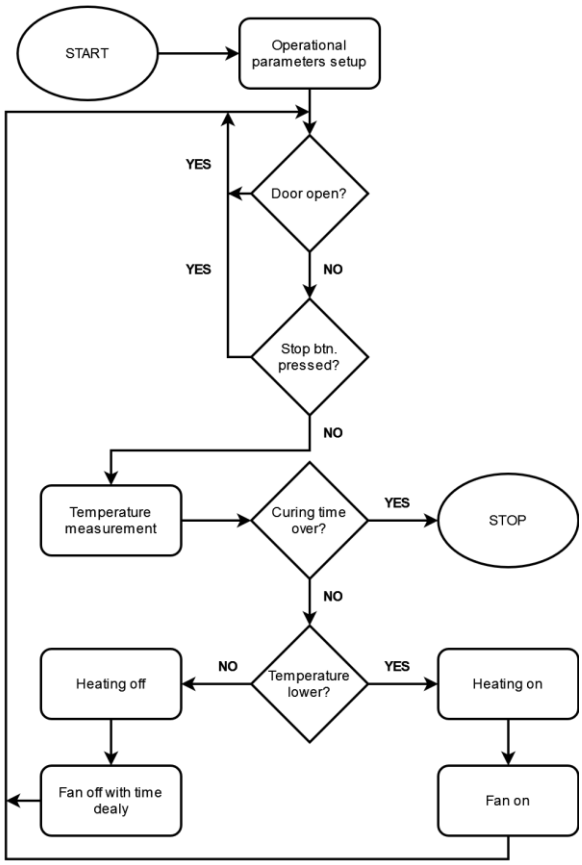


Figure 3. Chamber control flow diagram

Heating control utilizing electric heating elements according to the flow diagram is implemented in the Node-RED environment with the nodes shown in

Figure 4. It represents a typical two-position control implementation.

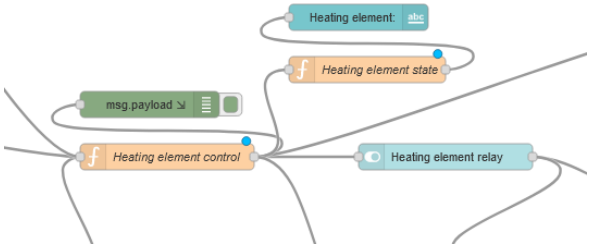


Figure 4. Heating control-related nodes

The fan speed control is implemented using PWM on the digital output of the PLC. The PWM settings are not directly available as a node and must be accessed via the MODBUS PLC registers. The method of setting the PWM parameters can be seen in the right part of Figure 5. through the nodes Set value, Modbus Flex Write and Modbus Response.

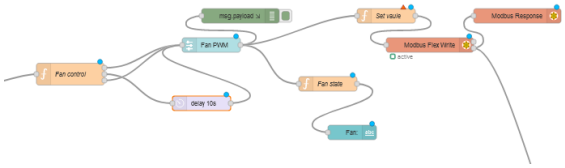


Figure 5. Fan control-related nodes

The MODBUS frame sent to the PLC to set the PWM parameters via node *Set value* has the following format,

```
msg.payload = {
    'quantity': 1,
    'value': msg.payload,
    'count': 1,
    'fc': 06,
    'address': 18,
    'unitid': 1
}
return msg;
```

where *value* represents set PWM duty, *fc* preset single register, *address* MODBUS register and *unitid* connecting route.

Figure 6. shows the implementation of the operating parameters settings, which include curing time, curing temperature, fan run time, and dead band. All settings are user configurable in the UI dashboard through input boxes.

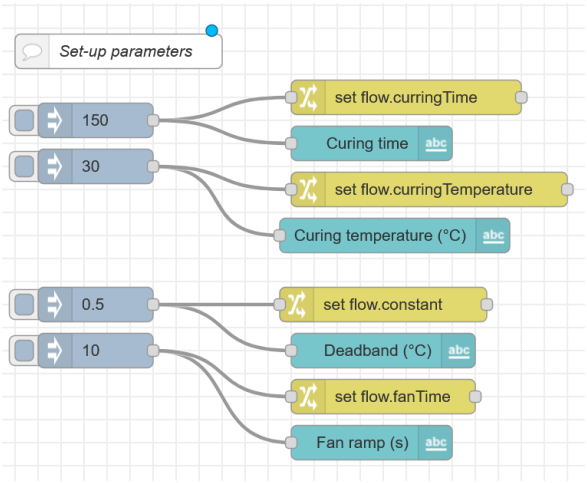


Figure 6. Set-up related nodes

Figure 7 shows the measurement and processing of temperature values. First, values are read from connected digital thermometers, then converted to °C values. Consequently, they are split and formatted for individual temperature values.

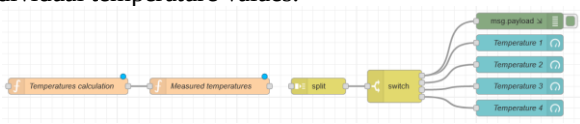


Figure 7. Temperature-related nodes

A simplistic visualization of the climate chamber control process is depicted in Figure 8. In the first column are used defined operating parameters, and in the second column are shown other operating parameters. The chamber graphical visualization is shown in the middle with possible visual alarms on a specific location according to error causing part. On the very right are displayed measured temperature values from connected digital thermometers.

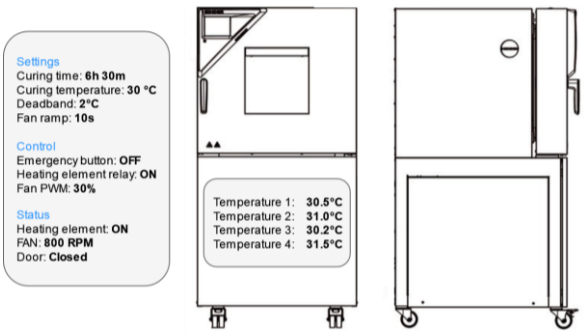


Figure 8. Dashboard example

5 Conclusions

Nowadays, open-source solutions are indisputably part of the toolbox for practical automation tasks. These

tasks include building automation, sensor systems, industrial applications, and many others. The area of upgrading existing equipment is proving to be an interesting one. Such equipment does not have the technical level to meet the requirements placed on current equipment. Despite their reliability and the possibility of even longer use, other solutions often replace them.

The proposed solution uses open-source tools that allow fully-fledged upgrade of existing devices for Industry 4.0 requirements with commonly available components such as PLCs, sensors, and others. Node-RED, a tool for programming event-driven applications, provides interconnection between hardware, devices, APIs, and various online services. After programming the basic functionality, the solution fully replaces the legacy control system while providing options for low-cost additional hardware and software extensions.

Despite the limited scope of this work, its outputs can form the basis for deploying similar solutions in complex tasks.

Acknowledgments

This paper was supported under the project of Operational Programme Integrated Infrastructure: Independent research and development of technological kits based on wearable electronics products, as tools for raising hygienic standards in a society exposed to the virus causing the COVID-19 disease, ITMS2014+ code 313011ASK8. The project was co-funded by European Regional Development Fund.

References

SIMATIC Industrial OS – The Linux-based operating system for SIMATIC IPC - ID: 109766374 - Industry Support Siemens (2019). Retrieved from https://support.industry.siemens.com/cs/us/en/view/109766374

Unipi (2023). Retrieved from https://www.unipi.technology/

Teach, learn, and make with the Raspberry Pi Foundation (2023). Retrieved from https://www.raspberrypi.org/

Node-RED (2023). Retrieved from https://nodered.org/

Dudak, J., Tanuska, P., Gaspar, G., Fabo, P. (2018). ARM-based universal 1-wire module solution", Journal of Sensors, 2018. https://doi.org/10.1155/2018/5268247

Barton, M., Budjac, R., Tanuska, P., Gaspar G., Schreiber, P. (2022). Identification Overview of Industry 4.0 Essential Attributes and Resource-Limited Embedded Artificial-Intelligence-of-Things Devices for Small and Medium-Sized Enterprises. *Applied Sciences* 12(11), 5672. https://doi.org/10.3390/app12115672

Martikkala et al. (2021). Trends for Low-Cost and Open-Source IoT Solutions Development for Industry 4.0. *Procedia Manufacturing*. 55, 298-305.

Rathee, S., & Chobe, A. (2022). Open Source Technology in Industry. *In: Getting Started with Open Source Technologies.* Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-8127-7_7

Omidi, S.A., Baig, M.J.A., Iqbal, M.T. (2023). Design and Implementation of Node-Red Based Open-Source SCADA Architecture for a Hybrid Power System. *Energies, 16(5),* 2092. https://doi.org/10.3390/en16052092

Chakravarthi, V.S. (2021). M2M Communication and Technologies. *In: Internet of Things and M2M Communication Technologies.* Springer, Cham. https://doi.org/10.1007/978-3-030-79272-5_10

Vagas, M., Galajdova, A., Simsik, D., Onofrejova, D. (2019). Wireless data acquisition from automated workplaces based on RFID technology. *In: IFAC-PapersOnLine*, 52(27), 299-304. https://doi.org/10.1016/j.ifacol.2019.12.677.

Fabusova, J. (2022). Využitie nástroja Node-RED pri meraní teploty technologických procesov. Master's thesis.

Hlavatovic, J. (2022). Využitie nástroja Node-RED pri meraní teploty technologických procesov. Master's thesis.

Modelling the Fermentation Process in Winemaking using Temperature and Specific Gravity

Ivana Kovačević, Mihaela Orić, Ivana Hartmann Tolić, Emmanuel Karlo Nyarko
Josip Juraj Strossmayer University of Osijek
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Kneza Trpimira 2b, 31000 Osijek, Croatia
ivana.kovacevic@ferit.hr, mihaela.oric@student.ferit.hr, ivana.hartmann@ferit.hr, karlo.nyarko@ferit.hr

Abstract. *This article examines the fermentation process in winemaking. The authors analyse fermentation data from a local winery in Croatia and propose two preliminary models to determine the optimality of ongoing fermentation processes using measurements of specific gravity and temperature. The data include measurements of temperature and specific gravity taken during successful fermentations of three different white wines. The models developed are a multivariate linear regression model and an artificial neural network model. The performance of each model is evaluated using metrics such as R-squared, Mean Squared Error, Root Mean Squared Error, and Maximum Absolute Error. The results show that both models accurately represent the fermentation process, with the artificial neural network model slightly outperforming the multivariate linear regression model.*

Keywords. Fermentation, Temperature control, Specific gravity, Modelling

1 Introduction

Traditional fermented foods are usually produced by spontaneous fermentation, which is dependent on climatic conditions. The winemaking process, like most of the other food production processes, has several distinct stages of production, namely: harvesting, crushing and pressing, fermentation, aging and maturation, clarification and stabilization, bottling and, finally, distribution and consumption. These stages in wine production involve a careful combination of art and science, with each step contributing to the final quality, characteristics, and enjoyment of the wine. Fermentation is a key step in winemaking, where yeast converts sugars in the must into alcohol, carbon dioxide, and other compounds. Temperature and specific gravity are vital parameters for controlling and monitoring fermentation. Temperature influences yeast metabolism, while specific gravity provides insights into sugar concentration and potential alcohol production. Maintaining an optimal fermentation temperature is crucial for yeast activity and the overall success of the fermentation process. Temperature influences yeast growth rate, sugar utilization, and the production of aroma and flavour compounds. High temperatures can lead to excessive yeast metabolism, resulting in the production of undesirable off-flavours and off-odours. Conversely, low temperatures can slow fermentation or cause yeast dormancy. Winemakers employ various techniques, such as cooling jackets, temperature-controlled tanks, or refrigeration systems, to regulate and maintain the desired fermentation temperature range.

Specific gravity is a valuable parameter for assessing sugar content and potential alcohol levels in the must. It is typically measured using hydrometers or refractometers. Specific gravity readings are taken before and during fermentation, providing insights into the progress and completeness of fermentation. Initially, the specific gravity is high due to the sugar content in the must. As yeast consumes sugar and produces alcohol, the specific gravity gradually decreases. Monitoring specific gravity allows winemakers to estimate the alcohol potential of the resulting wine and determine when fermentation has finished.

The authors of this article are involved in a project to develop an expert system to help winemakers in the production of wine. As mentioned earlier, one of the steps in wine production is fermentation. In this paper, the authors analyse the data recorded by a local winery in Croatia during the successful fermentation of three different white wines. Based on this data, the authors propose two preliminary models that could be used to determine whether the ongoing fermentation process is optimal or not.

2 Methodology in related research

In recent years, machine learning (ML) techniques have gained significant popularity and have been applied in various fields to detect and classify patterns in complex data structures. ML techniques are applied in various fields such as computer vision, natural language processing, healthcare, finance, recommender systems and many others. The choice of ML method to be used depends on the type of data, availability of labelled data, and the problem or task at hand.

Wine is a fascinating and sophisticated product with special characteristics that distinguish it from other products. Accordingly, the analysis, modelling and control of each stage in wine production has been researched exhaustively in an effort to simplify the control of each stage.

In (Tardaguila et al., 2021) an overview of smart applications and digital technologies that are currently applied in viticulture is provided. The test methods used to assess wine quality are demanding and varied. The advice of a wine expert is valuable, but also expensive and subjective. For this reason, various machine learning-based methods for predicting wine quality have been proposed (Chiu et al., 2021, Lee et al., 2021).

In (Román et al., 2011), the authors propose ANNs for effectively recognizing nonlinear patterns in wine fermentation processes and accurately predicting fermentation problems. They explored different sub-cases by varying the predictor variables and the time of fermentation. The predicted variables were total sugar, alcohol, glycerol, density, organic acids, and nitrogen compounds, and the time of fermentation was examined at three different durations. The results highlight the potential of ANNs as a valuable tool for monitoring and optimizing wine production processes. In (Urtubia et al., 2021) two cases were studied for the prediction of problematic wine fermentations. In the first case, the SVM algorithm was used to analyse different groups of chemical variables, including amino acids, saturated fatty acids, unsaturated fatty acids, organic acids, and fermentation control variables. In the second case, individual chemical variables, namely density, YAN (yeast assimilable nitrogen), Brix (sugar content), and acidity, were studied as predictors of problematic wine fermentation using SVM. The integration of artificial intelligence (AI) and machine learning (ML) techniques in fermentation bioprocesses is proposed in (Florea et al., 2022). They combine the fields of computer engineering and food engineering to demonstrate the importance of digitalization in modern industrial processes. The research involves the development and implementation of a software application that models a fully configurable 3-layer feed-forward multilayer perceptron neural network. This neural network, trained with experimental data, is able to predict the evolution of quantities characteristic of the alcoholic fermentation process in white wines. The study of the volatile organic compounds (VOCs) profile of pomelo wine and the development of an index to predict the degree of fermentation led to the improvement of the quality of pomelo wine in industrial production (Wei et al., 2022). Gas chromatography-mass spectrometry was used to analyse the changes in VOC content during the different fermentation periods. Principal component analysis, cluster analysis and partial least squares (PLS) regression were used to evaluate the variations of VOC content in the different fermentation phases. The PLS models identified the α-phellandrene/geraniol ratio as a potential index for determining the degree of fermentation of pomelo wine. Multivariate analysis techniques were also used to create an index to predict the degree of fermentation of pomelo wine under controlled laboratory conditions. The proposed index provides winemakers and wine chemists with valuable insights into the VOC composition and profile of pomelo wine throughout the fermentation process.

Modelling cold fermentations with non-conventional Saccharomyces species for establishing an experimental-modelling pipeline is proposed in (Henriques et al., 2018). The experimental pipeline includes conducting small-scale micro-vinifications and wine fermentations under controlled conditions. During these fermentations, the authors monitored growth rate and critical extracellular metabolites such as glucose, fructose, ethanol, glycerol, and acetic acid. They compared several candidate models that account for different biomass growth, transport, and inhibition mechanisms described in the literature. Developing a minimal model with the fewest parameters necessary for structural identifiability ensures that the model can be fit uniquely to the data while improving practical identifiability. For the most successful models, a combined modelling strategy is used to increase their robustness and minimize predictive uncertainty. The results are presented quantitatively and they use an ensemble of models to make robust predictions for processing conditions, such as initial inoculation and temperature, in order to optimize alcohol and glycerol production.

In (Nelson et al., 2022), the authors combine in *situ* measurements of density based on differential pressure measurements with a wine kinetic model and parameter estimation routine to predict the progression of a 1500-L wine fermentation.

3 Wine Fermentation Data

The data analysed are fermentation data recorded by a local winery in Croatia over a period of eleven years: 2010-2020. The winery uses measurements of temperature and specific gravity to implement control strategies during fermentation. Temperature control involves monitoring and adjusting the fermentation temperature to create optimal conditions for yeast

activity. This can be achieved by cooling or heating, depending on the desired wine style and yeast strain used. The specific gravity measurement is used to help decide when to add nutrients, adjust sugar levels or decide when to stop fermentation. The data provided are recorded series of fermentations where each data sample included two measurements of the wine must, namely the specific gravity (γ) in the unit Oechsle [°Oe] and the temperature T [°C]. The Oechsle unit [°Oe] is a unit of measurement commonly used in winemaking to estimate the potential alcohol content of grapes or grape must. It is based on the specific gravity of the grape juice or must and measures the soluble solids, mainly sugars, in the juice, which directly correlate with the potential alcohol that can be produced during fermentation. The Oechsle scale usually ranges from 50°Oe to 200°Oe, with higher values indicating a higher sugar concentration and potential alcohol content. It is important to note that the Oechsle unit is mainly used in German winemaking, especially for classifying German wines based on maturity and sugar content. Other scales and units may be used in other winegrowing regions, such as Brix in the United States or Baumé in France. All data provided refer to successful fermentations (series) of three different white wines, i.e. white grape varieties: Graševina, Chardonnay and Pinot Gris. An overview of the distribution of the measured data can be found in Fig. 1 and Table 1.

Table 1. Total number of data samples per grape variety

	Graševina	Chardonnay	Pinot Gris
Total no. of fermentation processes	20	22	22
Total no. of data samples	397	393	329

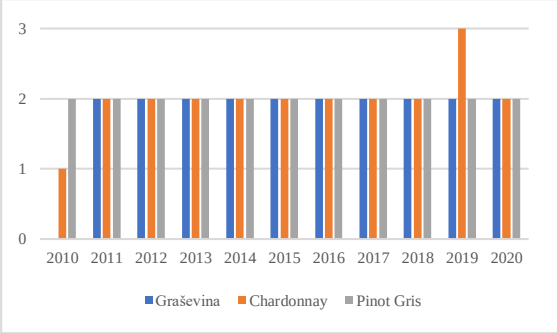


Figure 1. Number of fermentation processes recorded per year and per grape variety

A total of 64 different fermentation processes were recorded, comprising 1119 data samples. The minimum and maximum readings of the data samples are shown in Table 2.

It can be seen in Table 2 that the measured values for specific gravity ranged from 26 °Oe to 104 °Oe, while the temperature values ranged from 12.4 °C to 24 °C. It can also be seen that the interval between the measurements of the data samples is between 1 and 5 days. Of the 1119 data samples, 1073 measurements were taken every 24 hours, 35 after 48 hours, 10 after 72 hours and 1 after 5 days (Table 3).

Table 2. Minimum and maximum values of the data measured during fermentations.

Parameter Measured	Min. value	Max. value
Specific Gravity, γ [°Oe]	26	104
Temperature T [°C]	12.5	24
Interval between measurements [days]	1	5
No. of data samples per fermentation process	7	34

Table 3. Distribution of measurement interval

Interval between measurements [days]	No. of data samples
1	1073
2	35
3	10
5	1

The distribution of the temperature recorded is given in Fig. 2.

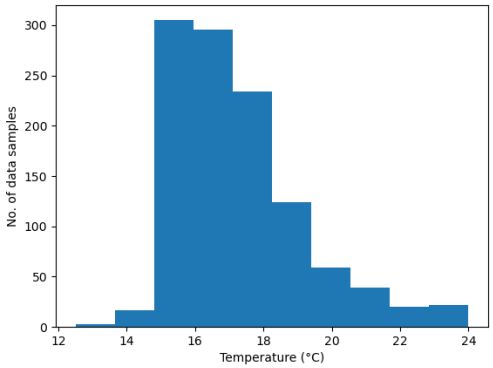


Figure 2. Histogram of temperature measured

Fig.2 shows that most of the measured data samples have a temperature between 15 °C and 18 °C, indicating that the winery controls the fermentation process mainly within this temperature range. An example of a fermentation process is shown in Fig. 3. It can be seen that during the fermentation process, the specific gravity value decreases with time, indicating a

decrease in the sugar content, and an increase in the alcohol content.

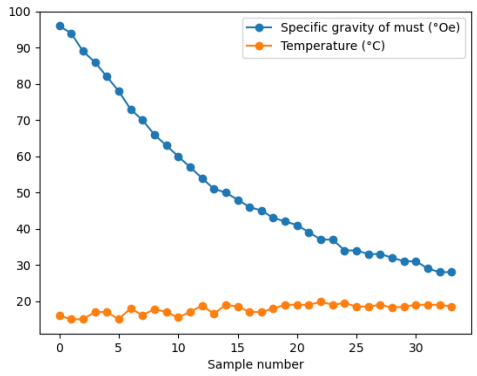


Figure 3. Example of a fermentation process showing the specific gravity and corresponding temperature.

Based on the recorded data, the authors attempt to model the fermentation process. For this purpose, the data is first divided into two subsets: the training set and the test set. The training set is used to determine the parameters of the model, while the fit of the model is determined using the test set. It should be noted that the data was split without taking the grape variety into account. However, it is imperative that a complete fermentation process or series is included in either the training set or the test set. 22% of the complete fermentation processes were randomly selected as the test set and the rest were used as the training set. The division of the measured data into the training set and the test set is given in Table 4.

Table 4. Division of the measured data into train set and test set

	Total	Train set	Test set
Total no. of fermentation processes	64	50	14
Total no. of data samples	1119	868	251

4. Model evaluation

All models considered in this paper are evaluated using the following evaluation metrics: R-squared (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Maximum Absolute Error (MAE).

R^2 represents the proportion of the variance in the dependent variable that is explained by the independent variables. It ranges from 0 to 1, with higher values indicating better fit. It is defined by eq. 1:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - y_M)^2} \quad (1)$$

where:

y_i – represents the measured or actual value,

f_i – represents the predicted value,

y_M – represents the mean of the measured value.

MSE (eq. 2) measures the mean squared difference between predicted and actual values. It provides a general measure of the accuracy of the model, with lower values indicating better performance.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - f_i)^2 \quad (2)$$

RMSE (eq. 3) is the square root of MSE and provides a more interpretable metric in the same units as the dependent variable.

$$RMSE = \sqrt{MSE} \quad (3)$$

MAE (eq. 4) is the maximum absolute difference between predicted and actual values. It gives an indication of the largest individual prediction error of the model.

$$MAE = max(|y_i - f_i|) \quad (4)$$

5 Fermentation Modelling

Based on the specific gravity measured at time *k*, and the temperature at time *k*+1, the model should provide an estimate of the specific gravity at time *k*+1 for a given fermentation process. Such a model would help a wine producer to control the fermentation process by adjusting the temperature based on the current specific gravity value and the desired specific gravity value after one day. The model to be determined, *F*, is defined by the following equation:

$$\gamma(k+1) = F(\gamma(k), T(k+1), k) \quad (5)$$

where:

k – sample number,

γ(k) – specific gravity measured at time *k*,

T(*k*+1) – temperature at time *k* +1.

The model to be determined, *F*, thus has 3 input variables (features) and one output variable.

As this is a preliminary study, two models are analysed and compared: a multivariate linear regression model and an artificial neural network model. The reason for selecting these models at this stage was because of their key differences: Multivariate linear regression is simpler and suitable for linear relationships, while artificial neural networks are more complex and versatile and can capture both linear and non-linear patterns in the data. All modelling was done in Python using the machine learning library scikit-learn (https://scikit-learn.org/stable/index.html). Before training the models, all data were preprocessed by performing data standardization with normal distribution, i.e. all input parameters were transformed to have a mean of 0 and a standard deviation of 1.

5.1 Multivariate linear regression model

Based on the training data, the specific gravity at time *k*+1 for a given fermentation process is given by the model:

$$\gamma(k+1) = 50.138 + 19.880\gamma(k+1) - 0.520T(k+1) + 1.504k \quad (6)$$

The performance of the linear regression model on the test dataset is shown in Table 5 and Fig. 4. Analysing Table 5, it can be concluded that the multivariate linear regression model has a high R^2 value of 0.9854 and a low RMSE value of 2.3902.

Table 5. Performance metrics of the multivariate linear regression model on the test dataset

R^2	MSE	RMSE	MAE
0.9854	5.7129	2.3902	7.2288

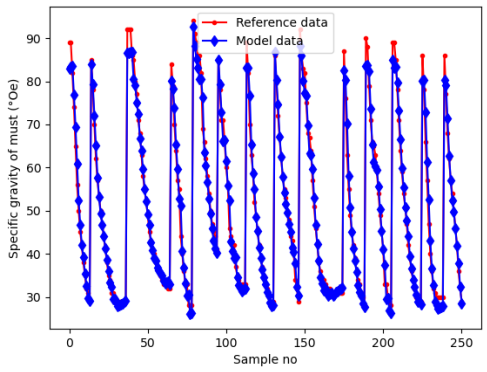


Figure 4. Comparison of the output values of the multivariate linear regression model with the actual values

Looking at Fig. 4, it can be concluded that the multivariate linear regression model models the fermentation process relatively well. However, it should be noted that it does not estimate the initial fermentation values very well.

5.2 Artificial Neural Network Model

A simple feed-forward neural network model with a single hidden layer was used. The structure of the neural network model was determined using the Grid Search Cross-Validation "GridSearchCV" function in the scikit-learn library. It was assumed that the neural network model has only one hidden layer. The number of neurons in the hidden layer was varied from 2 to 15, and three activation functions of this layer were also considered: logistic, tanh and relu. The linear activation function was used in all cases for the neuron in the output layer. The best artificial neural network model had three neurons in the hidden layer with the logistic-sigmoid activation function. The performance of the selected artificial neural network model on the test dataset is shown in Table 6 and Fig. 5.

Table 6. Performance metrics of the artificial neural network model on the test dataset

R^2	MSE	RMSE	MAE
0.9885	4.502	2.122	6.6514

Table 6 shows that the artificial neural network model has a high R^2 value of 0.9885 and a low RMSE value of 2.122.

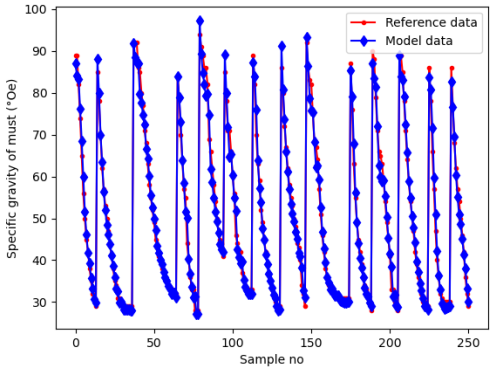


Figure 5. Comparison of the output values of the artificial neural network model with the actual values

Looking at Fig. 5, it can be concluded that the artificial neural network model represents the fermentation process quite well.

6 Discussion

The comparison of the two models shows that the artificial neural network slightly outperforms the multivariate linear regression model. The R^2 value increases by 0,3 % and the values of MAE and RMSE decrease by 8 % and 11 % respectively. The slightly better performance of the artificial network model could be due to the fact that the relationship between the data in the proposed model has some non-linearity that cannot be captured by the multivariate linear regression model.

7. Conclusions and Future Work

The aim of this work was to develop models for assessing the fermentation process of white wines based on specific gravity and temperature measurements. The models demonstrate good performance, with the artificial neural network model having a slight advantage over the multivariate linear regression model in terms of accuracy. The multivariate linear regression model showed high accuracy in estimating the specific gravity at the next time step, as indicated by the high R^2 value of 0.9854 and low RMSE value of 2.3902. However, the model

had difficulty in accurately estimating the values for the initial fermentation. In contrast, the artificial neural network model with a single hidden layer and three neurons provided a good representation of the fermentation process. These results suggest that the models have potential for predicting fermentation outcomes in winemaking and provide winemakers with a valuable tool for optimizing the process and ensuring high-quality wine production. Further research could explore the generalizability of the models to other wine varieties and the investigation of additional factors that influence the fermentation process.

Acknowledgments

We thank Krauthaker Vineyards and Winery for providing the fermentation data used in this research. This work results from implementing research activities on the project *Development of an expert system for food production and processing management (ESIA-Expert System for Intelligent Agriculture KK.01.1.1.07.0036)* funded by the European Regional Development Fund.

References

Chiu, T. H.-Y., Wu, C.-W., & Chen C.-H., (2021). A Hybrid Wine Classification Model for Quality Prediction. In Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, Proceedings, Part IV. Springer-Verlag, Berlin, Heidelberg, 430–438. https://doi.org/10.1007/978-3-030-68799-1_31

Florea, A., Sipos, A., & Stoisor, M. C. (2022). Applying AI Tools for Modeling, Predicting and Managing the White Wine Fermentation Process. Fermentation, Vol. 8, Page 137, 8(4), 137. https://doi.org/10.3390/FERMENTATION8040137

Henriques, D., Alonso-del-Real, J., Querol, A., & Balsa-Canto, E. (2018). Saccharomyces cerevisiae and S. kudriavzevii synthetic wine fermentation performance dissected by predictive modeling. Frontiers in Microbiology, 9(FEB), 317309. https://doi.org/10.3389/FMICB.2018.00088/BIBTEX

Lee, C. K., Law, K. M., & Ip, A. W. (2021). A Rule-Based Quality Analytics System for the Global Wine Industry. Journal of Global Information Management (JGIM), 29(3), 256-273. http://doi.org/10.4018/JGIM.20210501.oa1

Nelson, J., Boulton, R., & Knoesen, A. (2022) Automated Density Measurement With Real-Time Predictive Modeling of Wine Fermentations," IEEE Transactions on Instrumentation and Measurement, 71(1-7), 3509607. https://doi.org/10.1109/TIM.2022.3162289.

Román, R. C., Hernández, O. G., & Urtubia, U. A. (2011). Prediction of problematic wine fermentations using artificial neural networks. Bioprocess and Biosystems Engineering, 34(9), 1057–1065. https://doi.org/10.1007/S00449-011-0557-4/METRICS

Tardaguila J., Stoll M., Gutiérrez S., Proffitt T., & Diago M. P. (2021). Smart applications and digital technologies in viticulture: A review, Smart Agricultural Technology, 1,100005. https://doi.org/10.1016/j.atech.2021.100005.

Urtubia, A., León, R., & Vargas, M. (2021). Identification of chemical markers to detect abnormal wine fermentation using support vector machines. Computers & Chemical Engineering, 145, 107158. https://doi.org/10.1016/J.COMPCHEMENG.2020.107158

Wei, Q., Liu, G., Zhang, C., Sun, J., & Zhang, Y. (2022). Identification of characteristic volatile compounds and prediction of fermentation degree of pomelo wine using partial least squares regression. LWT, 154, 112830. https://doi.org/10.1016/J.LWT.2021.112830

Mapping RGB-to-NIR with Pix2Pix Image-to-Image Translation for Fire Detection Applications

Amila Akagić, Emir Buza
University of Sarajevo
Faculty of Electrical Engineering
Zmaja od Bosne bb, 71000 Sarajevo, BiH
{aakagic, ebuza}@etf.unsa.ba

Marko Horvat
University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Applied Computing
Unska 3, 10000 Zagreb, Croatia
marko.horvat3@fer.hr

Abstract. *In this research paper, we present a novel application of image-to-image translation techniques for fire detection applications. The focus of the study is on translating RGB images into Near InfraRed (NIR) images, which can serve as a preprocessing step for fusion or other methods that benefit from such information. The Pix2Pix model is employed to generate NIR images from existing RGB images, thereby adding an extra data source. The experimental results show the ability of the model to successfully learn the translation process, capturing the desired characteristics in the generated NIR images. The histograms and structural similarity index confirm the fidelity of the translations, with an average index value of 82.42%. The model effectively detects the position and shape of fire in the images, although some details on the edges may appear less prominent. Data augmentation could be employed to enhance the ability of the model to produce high-quality NIR images in various scenarios.*

Keywords. Computer Vision, Image Processing, Deep Learning, Image-to-Image Translation, NIR

1 Introduction

Computer vision-based wildfire prediction and detection systems have gained significant attention and popularity in recent years (Buza and Akagic, 2022; Akagic and Buza, 2022). The advancements in technology, especially in the fields of machine learning and deep learning, have enabled the analysis of large volumes of digital images with greater accuracy and efficiency (Šabanović et al., 2023; Kapo et al., 2023). These systems hold great promise for detecting and predicting wildfires. In this process, unmanned aerial vehicles (UAVs) equipped with computer vision-based technology play a crucial role in monitoring and fighting wildfires.

In recent trends of computer vision and image processing, there is compelling evidence supporting the idea that the inclusion of Near InfraRed (NIR) information is beneficial for pattern recognition, as it offers significant advantages in capturing a comprehensive representation of a scene across various applications (Ghiass et al., 2014; Mishra et al., 2022). The inclusion of NIR information is motivated by its ability to provide valuable insights that enhance accuracy and utility in real-world scenarios. Unlike prevailing methods that predominantly analyze 2D images within the visible light spectrum (Brdjanin et al., 2020; Dardagan et al., 2021), which are vulnerable to variations in environmental illumination that can degrade their performance, NIR bands exhibit resilience to such fluctuations, making them a valuable and reliable data source for image processing.

The NIR data provides a promising avenue for the development of a novel methodology aimed at analyzing multi-modal images. This innovative approach holds the potential for augmenting accuracy in pattern recognition within the domain of image processing. Its efficacy is particularly pronounced in situations where labeled training data is limited or unattainable, thereby addressing the constraints imposed by specific regions or circumstances. Moreover, this approach exhibits versatility across various data sources, encompassing satellite or Unmanned Aerial Vehicle (UAV) imagery, which facilitates the identification of intricate patterns and anomalies without necessitating human labeling or intervention.

In this research paper, we introduce a pioneering application of image-to-image translation techniques targeted at fire detection. Our focus centers on the translation of initial RGB images into Near InfraRed (NIR) images, which can serve as a preprocessing step for fusion or other types of methods that can benefit for such information. Our approach can be used to generate NIR images from existing RGB images, thus creating an additional source of data. The RGB and NIR data sources can then be combined or fused to increase the resilience of the algorithm for the increased benefit of pattern recognition. Such algorithms can leverage several key advantages, including heightened resilience to variations in environmental illumination and improved spatial resolution. These benefits facilitate the pre-

cise identification of areas impacted by fires, thereby enabling more efficient and timely wildfire detection. Such advancements play a pivotal role in safeguarding human life, preserving the environment, and mitigating the deleterious consequences associated with wildfires.

In the context of wildfire detection, RGB-NIR data can be particularly beneficial. It enhances the spatial resolution of the images and allows for the differentiation of vegetation types based on their reflectance patterns in the NIR spectrum. This capability enables a better identification of areas affected by flames versus those that are not. In recent literature, there are very limited data sources for NIR images, as opposed to the RGB fire and/or flame images. Thus, this paper tries to address this limitations and provide valuable insights into generating new data sources. We believe that this approach can be used for other applications as well.

The rest of this work is organized as follows. The related work is provided in Section II. The proposed method is described in section III. In Section IV, the evaluation of the proposed method and results are described. Finally, Section V concludes the paper.

2 Related Work

As previously mentioned, the inclusion of Near InfraRed (NIR) information has been shown to significantly enhance and streamline the performance of various image processing and computer vision tasks. This has been exemplified in several applications documented in the literature. For instance, in studies such as (Rüfenacht et al., 2013) and (Salamati et al., 2011), NIR information was utilized to effectively remove shadows from images. The elimination of shadows in images holds the potential to enhance the accuracy and efficacy of tasks such as tracking, segmentation, and object detection. Shadow boundaries often result in confusion with different surfaces or objects, thus it becomes advantageous to eliminate them entirely from the image.

In (Brooksby et al., 2003), the authors investigate the potential of integrating Magnetic Resonance Imaging (MRI) and Near-Infrared (NIR) imaging modalities to achieve noninvasive, high-resolution maps of optical properties. In (Schaul et al., 2009) and (Feng et al., 2013), two image dehazing approaches that leverage the combination of color and NIR images are proposed. In (Clarke, 2004), the authors explore two methods for extracting process-related information from NIR microscopy data cubes. Additionally, NIR has found applications in image enhancement (Matsui et al., 2011), image registration (Firmenichy et al., 2011), image color restoration (Lv et al., 2022), prediction of quality attributes (Kamruzzaman et al., 2012), and many others.

In the context of image-to-image translation, both NIR-to-RGB and RGB-to-NIR conversions have been explored. The NIR-to-RGB translation has been investigated in (Dou et al., 2019), where it was applied to face image translation. In (Sun et al., 2019), the authors proposed a novel asymmetric cycle GAN for NIR to RGB domain translation, specifically focusing on NIR colorization. While (Yan et al., 2020) introduced a new method for NIR-to-RGB translation, utilizing a U-net based neural network to learn texture information and a CycleGAN based neural network to extract color information. In (Shukla et al., 2022), the authors presented a technique for high-resolution NIR prediction from RGB images, targeting plant phenotyping applications. While both approaches exist, the literature tends to have a relatively greater emphasis on the NIR-to-RGB translation.

The RGB-to-NIR approach has gained attention in agricultural applications for determining crop parameters that are not visible to the human eye, as discussed in (Aslahishahri et al., 2021). The authors investigate image-to-image translation techniques to generate a NIR spectral band solely from a RGB image in aerial crop imagery. Similarly, (Sa et al., 2022) synthesize NIR information from RGB input using a data-driven, unsupervised approach without the need for manual annotations or labels. Their focus is on enhancing the system for fruit detection. (Yuan et al., 2020) explore the use of a conditional Generative Adversarial Network (cGAN) structure to generate a NIR spectral band conditioned on the input RGB image. (Ciprián-Sánchez et al., 2021) provides a quantitative demonstration of the feasibility of applying deep learning-based fusion methods to infrared imagery from wildfires. They introduce a novel artificial IR and fused image generator called FIRe-GAN. The goal of this method is to fuse RGB with NIR information while trying to preserve the consistent color.

The goal of this paper is to find a way to translate an RGB image and generate a corresponding NIR image as output. An inference module is created to generate a domain image-to-image translation, e.g. translation of RGB to NIR images.

2.1 Image-to-image Translation

Image-to-image translation refers to a task of converting an input image from one domain to a different domain image while preserving its essential content. In the case of RGB to NIR image translation, the goal is to transform an RGB image into its corresponding NIR representation. RGB images are composed of three color channels: red, green, and blue, which are visible to the human eye. On the other hand, NIR images capture light in the near-infrared spectrum, which is beyond the range of human vision.

In recent literature, image-to-image translation techniques are often based on deep learning models like generative adversarial networks or variational autoencoders (VAEs) (Liu et al., 2017; Zhu et al., 2017; Pang et al., 2021). For training a model, a dataset that con-

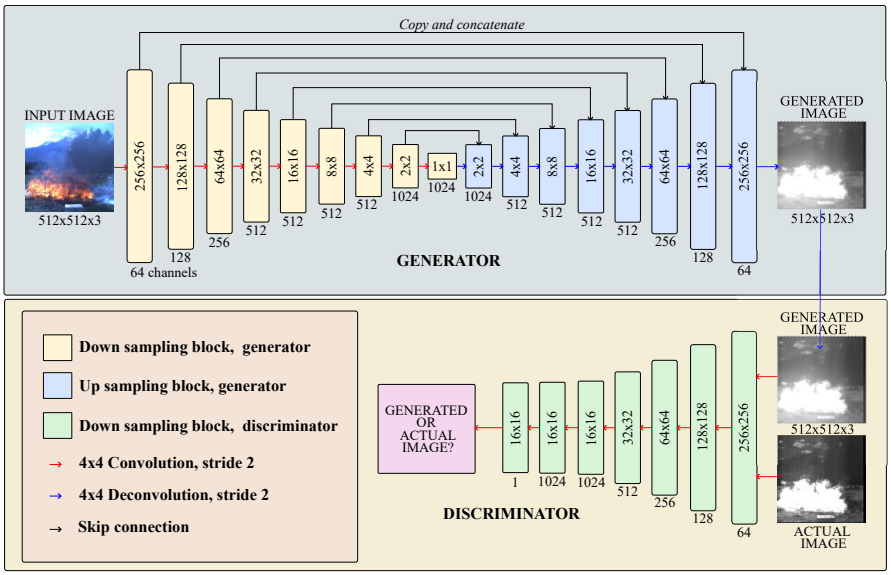


Figure 1: An example configuration of Pix2Pix as is used in our experiments.

tains pairs of aligned RGB and NIR images is required. The model is trained to minimize the difference between the generated NIR images and the ground truth NIR images in the dataset. This enables the model to learn the underlying patterns and correlations between RGB and NIR images, allowing it to generate plausible NIR representations when given an RGB input during the testing phase.

In our case, the tranformation from RGB to NIR image is approximated by a conditional generative adversarial network, which is introduced in the following section. This transformation is highly nonlinear which stems from many factors, such as lighting sources, surface reflections, intristic and extrinsic camera characteristics. Here, the challenge is to estimate the global optimal solution that guarantees convergence.

2.2 Pix2Pix Generative Adversarial Network

In this paper, we employ an existing image-to-image translation method known as Pix2Pix (Isola et al., 2017). Pix2Pix is a type of cGAN, which was introduced in 2017 by the researchers from Berkeley AI Research (BAIR) Laboratory, UC Berkeley. It can learn to map an input image from one domain to an output image in a different domain. Typically, Pix2Pix uses a paired dataset consisting of input images and their corresponding output images, while traditional GANs generate images from random noise. An example applications is the conversion of a black-and-white image into a colored image. Ideally, the network learns to generate realistic output images by training on these paired examples.

Similar to a traditional GANs, the architecture of Pix2Pix consists of two main components: a generator and a discriminator. In Fig 1. an example configuration of Pix2Pix is illustrated as is used in our experiments. In this paper, we investigate the use of Pix2Pix to understand the benefits the paired images can bring to this task. Below we briefly introduce the main components which are used for experiments.

2.2.1 Generator

The generator takes the input RGB image and attempts to transform it into an output NIR image. It is an encoder-decoder model with U-Net architecture (Ronneberger et al., 2015). This is accomplished by first downsampling (encoding) the input image down to a bottleneck layer, and then upsampling (decoding) the bottleneck representation to the size of the output image. In U-Net architecture, the skip-connections are added between the encoding layers and corresponding decoding layers, as denoted in Fig 1. The architecture consists of standardzed blocks of convolutional, batch normalization, dropout, and activation layers.

During training, the generator progressively improves its ability to generate realistic output images by minimizing the difference between the generated and real images, as perceived by the discriminator. It is trained via the discriminator model. The goal is to minimize the loss predicted by the discriminator for generated images which are predicted as real images.

2.2.2 Discriminator

The discriminator's role is to distinguish between the generated output image and the real output image from the dataset, thus the discriminator is trained on the real input and generated NIR images (Ronneberger et al., 2015; Schonfeld et al., 2020). The generator and discriminator are trained in an adversarial manner via adversarial loss, where the generator aims to produce output images that fool the discriminator, while the dis-

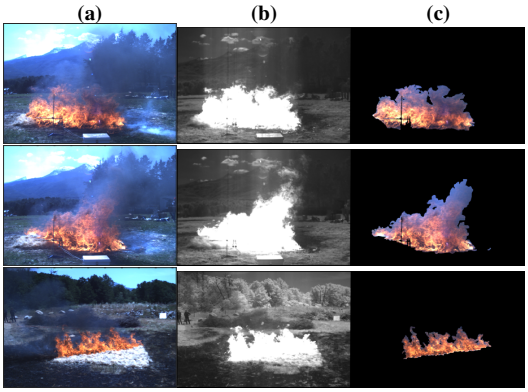


Figure 2: Example of images of the Corsican Fire Database taken with a multi-spectral camera. (a) visible (RGB) image, (b) Near Infrared (NIR) image, (c) ground truth flame pixels based on the visible image. Ground truth images are not used in this paper, however, they are shown here as a reference of what is a goal of detection from the original image.

criminator aims to correctly classify the real and generated images. The discriminator tries to become more accurate in distinguishing between real and generated images. This adversarial training process helps the generator learn to generate output images that are visually similar to the target domain. The update of the discriminator is unrelated to the update of the generator.

2.2.3 GAN

The generator and discriminator are connected together to create a composite model (Creswell et al., 2018; Goodfellow et al., 2020). Adam is used as an optimization method, with 0.0002 as a learning rate, while binary crossentropy and mean apsolute error (MAE) are used as loss funtions.

2.3 Dataset preparation

The Corsican Fire Database (CFDB) is used as our dataset (Toulouse et al., 2017). The CFDB comprises a collection of 500 visible images depicting wildfires captured worldwide, and 95 paired images consisting of visible and infrared modalities. These images are captured under realistic outdoor conditions. They are obtained using the JAI AD-090GE camera, which utilizes a 2-CCD multi-spectral camera system. The camera captures both visible and near-infrared spectra (700-900 nm) using the same optical setup. Both the visible and infrared images have dimensions of 1024×768 pixels. Each image in the dataset is accompanied by a corresponding ground-truth image. In this paper, the ground truth images are not used. Fig 2. illustrates some examples from the Corsican Fire Database. The CFDB dataset is openly accessible for research purposes.

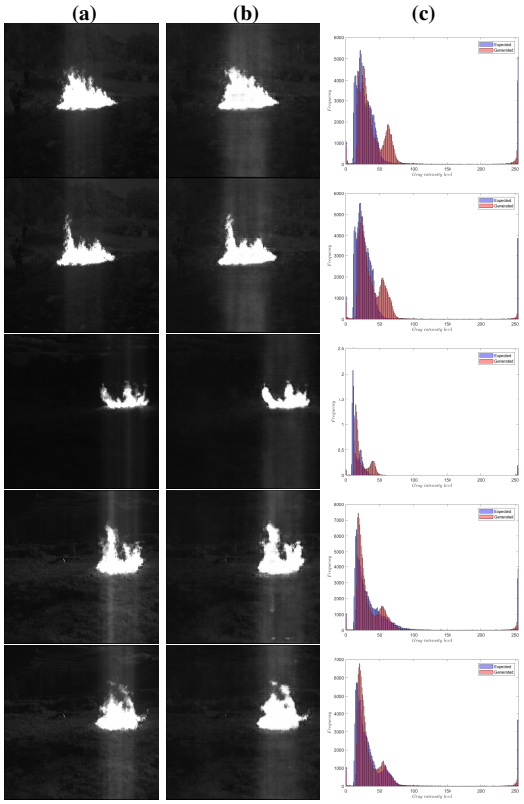


Figure 3: Results of our approach: (a) original NIR image from the CFDB, (b) generated NIR image, (c) and their histograms.

In this paper, we use 95 paired images from the CFDB, where 75 images are used (approximatelly 80%) during training, and the other images are used for testing purposes (20%). The images are first loaded and resized into a target size of 512×512 pixels to prepare them as inputs to Pix2Pix architecture.

2.4 The mapping procedure

The original RGB images are three-channels images, while NIR images are one-channel images. To learn the mapping function between RGB and NIR, we transform NIR images into three-channel images, where the content of the original channel is just copied to others. Then, we employ the classical Pix2Pix network where we experiment with diffrent configurations, and number of epochs and batch sizes. The generated images contain three channels, where information between channels differs slightly between 5-10%. To measure the performance of our approach, the generated NIR images have to be converted to the one-channel images. This procedure can be accomplished with several different methods, such as by computing the max, min, median or mean value of each pixel in three channels and then creating the one-channel images. We experiment with these results with respect to metrics and present the results in Section 3.

2.5 Metrics

The peak signal-to-noise ratio (PSNR) is used as a quality measurement between the original and a generated image (Korhonen and You, 2012). PSNR represents a measure of the peak error. The higher the PSNR, the better the quality of the generated image. To define the mathematical formulation, let us assume I and G are the original input and generated images, respectively. Then, the PSNR between I and G is given by Eq. (1).

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{I^2}{G} \right) \tag{1}$$

The Mean Squared Error (MSE) measures the average squared difference between the expected and the generated image of a dataset (Wang and Bovik, 2009). The lower the value of MSE, the lower the error. The MSE between I and G is given by Eq. (2).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (I_i - G_i)^2 \tag{2}$$

The Root Mean Squared Error (RMSE) is calculated by taking the square root of the Mean Squared Error (MSE).

MSE and PSNR focus on measuring the absolute differences between the predicted and true values or the reconstructed and original images, without considering the perceptual characteristics of human vision. They provide a quantitative assessment of the fidelity or accuracy of the signals or images by considering the mean squared error or the signal-to-noise ratio.

On the other hand, SSIM (Structural Similarity Index) is a perception-based model that takes into account the structural information and important perceptual phenomena related to human vision (Wang et al., 2004). It considers the idea that pixels in an image are interdependent, particularly when they are spatially close. SSIM measures the similarity between two images by assessing the perceived change in structural information, taking into consideration luminance masking and contrast masking effects. The Structural Similarity Index (SSIM) is calculated as shown in Eq. (3).

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3}$$

3 Results and Discussion

The Pix2Pix model is trained from scratch using an available open-source implementation[1]. The training process consists of a fixed number of iterations, with 100 epochs and a batch size of 1. Given the presence of 75 images in the training dataset, each epoch comprises 75 iterations, resulting in a total of 7500 training steps for the entire process. During each training step, a batch of real examples is selected as the initial step. The generator is then employed to generate a corresponding batch of samples based on the input source images. Subsequently, the discriminator is updated using both the batch of real images and the batch of generated samples.

Afterwards, the generator model undergoes an update, wherein the real source images are used as input, and the expected outputs consist of the real target images, accompanied by class labels of 1 (indicating real images). The loss calculation is based on these inputs. The generator produces two loss scores, along with a weighted sum score. We focus on the weighted sum score, as it is instrumental in updating the model weights.

To monitor the training progress and ensure timely evaluation, images are generated every 10 epochs and subsequently compared to the corresponding expected target images. This approach provides a means of regularly assessing the advancement of the training process. Upon completion of training, the remaining 20 images in the test subset are utilized to generate NIR images, which are then compared with the original images. To gauge the quality of the generated images, four metrics are employed for the comparison between the original (expected) and generated images. The obtained results are presented in Table **??**.

Table 1: The metrics results for test NIR images from CFDB.

Metrics	Results
PSNR	21.95
MSE	516.09
RMSE	22.72
SSIM	0.8242

The mapping procedure from a three-channel to a one-channel NIR image involved determining the maximum, minimum, median, or mean value of each pixel across the three channels. After evaluating the results, it was found that the minimum function yielded the best outcomes. Hence, the minimum function was employed to generate the final metric results. It is worth noting that the disparity between the maximum, minimum, median, and mean functions was minimal, with variances falling within the range of 1-2%.

Fig. 3. displays a set of five original RGB images alongside their corresponding generated NIR images, accompanied by their respective histograms. The histograms provide clear evidence that the model has effectively learned the translation process from RGB to NIR, capturing the desired characteristics. Furthermore, the structural similarity index demonstrates a remarkable similarity between the generated NIR images

[1]Pix2Pix Github page: `https://github.com/phillipi/pix2pix`

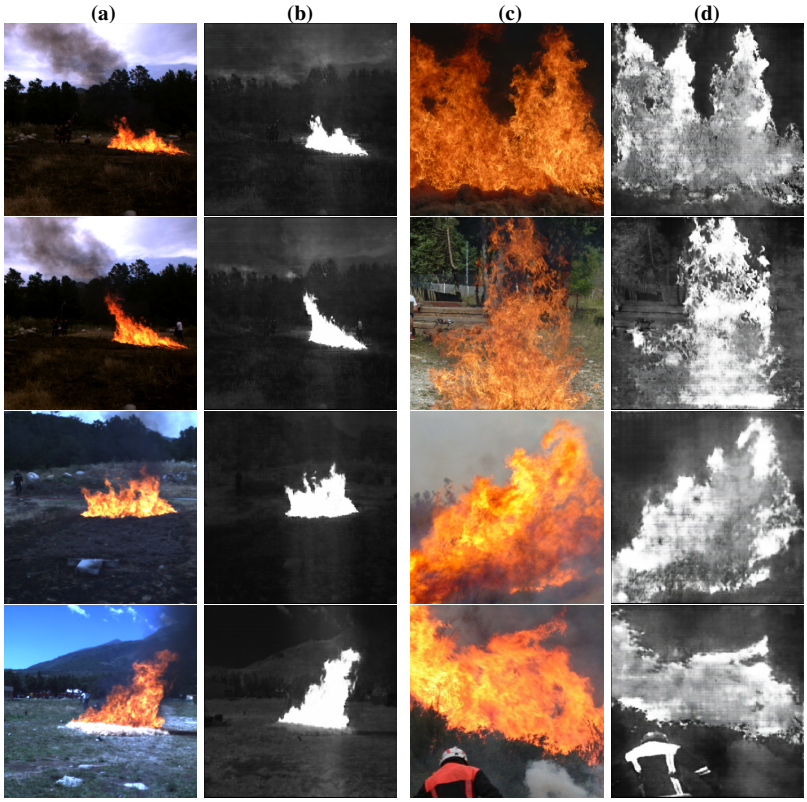


Figure 4: Results of our approach on a sample of 500 CFDB RGB images with satisfactory results: (a) original RGB image from the CFDB, and (b) generated NIR image; and with less satisfactory results: (c) original RGB image from the CFDB, (d) generated NIR image.

and the original ones, with an average value of 82.42%. Visual inspection confirms that the model excels in detecting the position and shape of the fire to a significant extent. However, it is worth noting that some details on the edges of the frames appear less prominent in the generated images.

The model was subsequently utilized to generate NIR images from the original 500 images from the CFDB dataset. In Fig 4., examples of both successful and less satisfactory translations are showcased. Since there are no corresponding original NIR images available for comparison, the evaluation is based solely on visual assessment. From the obtained results, it is evident that images resembling the ones encountered during training are effectively translated into their NIR counterparts, as seen in Fig 4. in (a) and (b). However, images that exhibit a higher density of fire pixels tend to have less satisfactory translations to their NIR representations, as seen in Fig. 4. (c) and (d). Although the model generally succeeds in accurately determining the position and shape of the fire, the internal structure appears blurred and requires additional processing.

To address this limitation, one potential approach would involve augmenting the training subset with similar samples and subsequently reiterating the training procedure. By incorporating a more diverse range of fire instances, the model can potentially improve its ability to capture finer details and enhance the translation of images with higher fire pixel density to NIR images.

4 Conclusion

In this paper, we demonstrate that the Pix2Pix model can be used to successfully perform image to image translation for the case of RGB to NIR translation. The model demonstrated its ability to successfully translate RGB images to NIR images, as evidenced by the clear similarities observed in the histograms and the high structural similarity index with an average value of 82.42%. The model effectively detected the position and shape of the fire, although some details on the edges of the frames were less pronounced in the generated images.

Applying the model to the CFDB dataset showed promising results for images resembling those encountered during training. The Pix2Pix model demonstrated its potential for generating NIR images and capturing key fire-related characteristics. Further refinements and augmentations could enhance its ability to produce high-quality NIR images in various scenarios.

References

Akagic, A. and Buza, E. (2022). Lw-fire: A lightweight wildfire image classification with a deep convolutional neural network. *Applied Sciences*, 12(5):2646.

Aslahishahri, M., Stanley, K. G., Duddu, H., Shirtliffe, S., Vail, S., Bett, K., Pozniak, C., and Stavness, I. (2021). From rgb to nir: Predicting of near infrared reflectance from visible spectrum aerial images of crops. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1312–1322.

Brdjanin, A., Dardagan, N., Dzigal, D., and Akagic, A. (2020). Single object trackers in opencv: A benchmark. In *2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6. IEEE.

Brooksby, B. A., Dehghani, H., Pogue, B. W., and Paulsen, K. D. (2003). Near-infrared (nir) tomography breast image reconstruction with a priori structural information from mri: algorithm development for reconstructing heterogeneities. *IEEE Journal of selected topics in quantum electronics*, 9(2):199–209.

Buza, E. and Akagic, A. (2022). Unsupervised method for wildfire flame segmentation and detection. *IEEE Access*, 10:55213–55225.

Ciprián-Sánchez, J. F., Ochoa-Ruiz, G., Gonzalez-Mendoza, M., and Rossi, L. (2021). Fire-gan: A novel deep learning-based infrared-visible fusion method for wildfire imagery. *Neural Computing and Applications*, pages 1–13.

Clarke, F. (2004). Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. *Vibrational Spectroscopy*, 34(1):25–35.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65.

Dardagan, N., Brđanin, A., Džigal, D., and Akagic, A. (2021). Multiple object trackers in opencv: a benchmark. In *2021 IEEE 30th international symposium on industrial electronics (ISIE)*, pages 1–6. IEEE.

Dou, H., Chen, C., Hu, X., and Peng, S. (2019). Asymmetric cyclegan for unpaired nir-to-rgb face image translation. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1757–1761. IEEE.

Feng, C., Zhuo, S., Zhang, X., Shen, L., and Süsstrunk, S. (2013). Near-infrared guided color image dehazing. In *2013 IEEE international conference on image processing*, pages 2363–2367. IEEE.

Firmenichy, D., Brown, M., and Süsstrunk, S. (2011). Multispectral interest points for rgb-nir image registration. In *2011 18th IEEE international conference on image processing*, pages 181–184. IEEE.

Ghiass, R. S., Arandjelović, O., Bendada, A., and Maldague, X. (2014). Infrared face recognition: A comprehensive review of methodologies and databases. *Pattern Recognition*, 47(9):2807–2824.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Kamruzzaman, M., ElMasry, G., Sun, D.-W., and Allen, P. (2012). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Analytica Chimica Acta*, 714:57–67.

Kapo, M., Šabanović, A., Akagic, A., and Buza, E. (2023). Early stage flame segmentation with deep learning and intel's openvino toolkit. In *2023 XXIX International Conference on Information, Communication and Automation Technologies (ICAT)*, pages 1–6. IEEE.

Korhonen, J. and You, J. (2012). Peak signal-to-noise ratio revisited: Is simple beautiful? In *2012 Fourth International Workshop on Quality of Multimedia Experience*, pages 37–38. IEEE.

Liu, M.-Y., Breuel, T., and Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30.

Lv, S., Huang, X., Cheng, F., and Shi, J. (2022). Color restoration of rgb-nir images in low-light environment using cyclegan. In *3D Imaging - Multidimensional Signal Processing and Deep Learning: 3D Images, Graphics and Information Technologies, Volume 1*, pages 1–9. Springer.

Matsui, S., Okabe, T., Shimano, M., and Sato, Y. (2011). Image enhancement of low-light scenes with near-infrared flash images. *Information and Media Technologies*, 6(1):202–210.

Mishra, P., Passos, D., Marini, F., Xu, J., Amigo, J. M., Gowen, A. A., Jansen, J. J., Biancolillo, A., Roger, J. M., Rutledge, D. N., et al. (2022). Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry*, page 116804.

Pang, Y., Lin, J., Qin, T., and Chen, Z. (2021). Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia*, 24:3859–3881.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.

Rüfenacht, D., Fredembach, C., and Süsstrunk, S. (2013). Automatic and accurate shadow detection using near-infrared information. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1672–1678.

Sa, I., Lim, J. Y., Ahn, H. S., and MacDonald, B. (2022). deepnir: Datasets for generating synthetic nir images and improved fruit detection system using deep learning techniques. *Sensors*, 22(13):4721.

Šabanović, A., Ahmetspahić, N., Kapo, M., Buza, E., and Akagic, A. (2023). Early stage flame segmentation with deeplabv3+ and weighted cross-entropy. In *2023 XXIX International Conference on Information, Communication and Automation Technologies (ICAT)*, pages 1–6. IEEE.

Salamati, N., Germain, A., and Siisstrunk, S. (2011). Removing shadows from images using color and near-infrared. In *2011 18th IEEE International Conference on Image Processing*, pages 1713–1716. IEEE.

Schaul, L., Fredembach, C., and Süsstrunk, S. (2009). Color image dehazing using the near-infrared. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1629–1632. IEEE.

Schonfeld, E., Schiele, B., and Khoreva, A. (2020). A u-net based discriminator for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8207–8216.

Shukla, A., Upadhyay, A., Sharma, M., Chinnusamy, V., and Kumar, S. (2022). High-resolution nir prediction from rgb images: Application to plant phenotyping. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4058–4062. IEEE.

Sun, T., Jung, C., Fu, Q., and Han, Q. (2019). Nir to rgb domain translation using asymmetric cycle generative adversarial networks. *IEEE Access*, 7:112459–112469.

Toulouse, T., Rossi, L., Campana, A., Celik, T., and Akhloufi, M. A. (2017). Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Safety Journal*, 92:188–194.

Wang, Z. and Bovik, A. C. (2009). Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

Yan, L., Wang, X., Zhao, M., Liu, S., and Chen, J. (2020). A multi-model fusion framework for nir-to-rgb translation. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 459–462. IEEE.

Yuan, X., Tian, J., and Reinartz, P. (2020). Generating artificial near infrared spectral band from rgb image using conditional generative adversarial network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:279–285.

Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E. (2017). Toward multimodal image-to-image translation. *Advances in neural information processing systems*, 30.

Optimizing Point Cloud Processing for Facial Reconstruction Using Open-Source Tools: A Preliminary Study

Ena Barčić, Petra Grd
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, Varaždin, Croatia
{enbarcic, petra.grd}@foi.unizg.hr

Abstract. *In this preliminary study, we present the first step in point cloud processing for the creation of bone and skin models for facial reconstruction, while focusing on what can be achieved using open-source tools. Through our experiment, we achieved an average reduction of 91.13% in point cloud models size for the skull and 80.35% for the skin model. The results demonstrate the feasibility of optimizing point cloud data, enabling more efficient storage, processing, and analysis for facial reconstruction. This study lays the foundation for further research in point cloud processing to enhance the efficiency and accuracy of facial models.*

Keywords. digital facial reconstruction, point clouds, data size, open-source tools

1 Introduction

Digital facial reconstruction (DFR), the process of recreating the facial features of an individual based on skeletal remains (Miranda et al., 2018), plays a critical role in forensic anthropology and the identification of unknown individuals. With the advance of digital technologies, there has been a shift towards the utilization of 3D modeling and visualization techniques. Point cloud models, even though a very new approach, have become increasingly prevalent in these applications due to their ability to accurately represent the surface geometry of facial structures. However, the growing complexity and size of point cloud data poses a challenge in terms of storage, processing, and computational efficiency. By reducing the size of point clouds, while preserving essential anatomical details, the presented method aims to optimize computational resources and streamline the facial reconstruction workflow.

Point clouds have distinct advantages over mesh techniques in forensic facial reconstruction, particularly when an accurate representation of both skeletal and facial anatomy is essential. Point clouds, being sets of three-dimensional points that capture the shape and position of facial features, allow for a more precise depiction of facial morphology and skeletal landmarks. Unlike meshes, which rely on surface interpolation and can sometimes oversimplify complex anatomical structures.

DFR offers numerous advantages over traditional methods, including speed, automation, and standardization (Miranda et al., 2018). Today, head CT (computed tomography) scans are most commonly utilized as a base for creating new DFR models. From CT images, point clouds can be created using Digital Imaging and Communications in Medicine (DICOM) data, representing a dense collection of individual measurement points with x, y, and z coordinates (Ulinuha et al., 2019). The flexibility of point cloud-based models allows for the precise placement of points within the 3D space, enabling extremely accurate localization of facial landmarks and features. This approach not only provides a more realistic visualization but also allows for quantitative measurements and comparisons.

Open-source tools have emerged as valuable resources in various scientific domains, and DFR is no exception. These tools offer researchers and forensic professionals access to freely available software libraries, algorithms, and frameworks that can be customized and extended for specific applications (Simmons-Ehrhardt et al., 2018). Leveraging open-source tools in this preliminary study enables the exploration of their potential in generating and analyzing point cloud data for DFR.

In the following sections, we will discuss the methodology employed for generating point cloud-based models from CT scans, the techniques used for point cloud creation, the analysis and evaluation of the created point clouds, the open-source tools, and the implications and limitations of this preliminary study.

2 Related work

DFR has been a topic of significant interest and research within the fields of forensic anthropology, computer vision, and medical imaging. This section provides an overview of related works and advancements in the field, highlighting key contributions and method-

ologies employed in the pursuit of more effective and precise DFR.

One of the first papers on DFR was (da Costa Moraes et al., 2014), where the authors present a protocol for computer-aided DFR using free software and photogrammetry. While they initially used a point cloud model, they converted it into a mesh using MeshLab. They focused on anatomical modeling of the facial features which resulted in the process being rather long and relying on virtual soft tissue pegs that were placed by hand for their final reconstruction. Similarly, the authors in (Simmons-Ehrhardt et al., 2018) present their approach to dense facial tissue depth mapping (DFTDM) of CT models using open-source tools. The authors generated DFTDM on hollowed, cropped face shells using the Hausdorff sampling. They used point clouds to perform the tissue depth measurements and found that minimum depth values ranged from 1.2 mm to 3.4 mm. Maximum depths were found in the buccal region and neck, excluding the nose.

Some of the earliest and most notable examples of CT scan usage can be seen in (Ulinuha et al., 2019) where the authors focus on creating a protocol for the extraction of skull and face surfaces from CT images. They focus on the problem of CT image size and were among the first authors to identify point clouds as a potential solution. Their extraction process produces point clouds with an average file size of 9.1% of the size of the original CT image file. In (Hargreaves et al., 2021) the authors present a generative deep learning approach for DFR directly from the bone with limited data and no background expertise. Their model is trained on 665 CT head scans that have been cleaned of noise, rotated to the correct orientation, and then filtered by density to find the bone and soft tissue to be used as the input and label respectively. They were able to achieve a facial detection score of 88.32%. Lastly, the authors (Duan et al., 2015) focus on 3D face reconstruction from the skull by regression modeling in shape parameter spaces. DFR is realized by using the relationship between the real face and the face statistical shape model. The skull and face are divided into five corresponding feature regions, and a mapping from each skull region to the corresponding face region is established. Their dataset consisted of 114 whole-head CT scans from the Han ethnic group in the North of China. They use dense meshes to represent the face and skull.

A popular approach is the creation of DFR models based on specific populations and STT. These are also the papers that put the largest emphasis on creating large-scale and population-specific databases. In (Ma et al., 2011), the authors focus on building a Vietnam-specific DFR based on STT and geometric proportions. Their database consists of full-head CT scans of 959 individuals. However, they still rely on manual landmarks and STT marker placement. They present one of the first approaches with deformable landmark-based templates and the addition of their own landmarks. The authors are able to achieve errors between 1.17mm to 1.24mm in two cases, which are some of the lowest errors achieved at the time. Additionally, the authors (Shui et al., 2010) present a China-specific DFR and build their own database, which consists of more than 200 full-head CT scans of Chinese people from both north and south China. They use thin-plate spline (TPS) to achieve non-rigid registration between the unidentified skull and reference skull and then choose a reference template. They also use a region-based approach reconstructing the face, nose, and mouth separately. In the end, the authors were able to achieve high-accuracy results in the coronal and mandible regions.

3 Creation of point cloud models

In this section, we present the main challenges to DFR, together with the tools and the dataset used in this study.

3.1 Challenges

While looking at the previously described research papers, we were able to identify several key challenges in the field of DFR:

One of the biggest challenges is the low availability of CT datasets. Obtaining a diverse and representative dataset that includes individuals from different populations and age groups is crucial to ensure the accuracy and generalizability of the reconstructed faces. In addition, the available datasets are all very small in size. To address this challenge, efforts should be made to expand the existing datasets by collaborating with forensic institutions, medical facilities, and research organizations to collect and share relevant CT scans and associated metadata. Finally, there are no "clean" publicly available datasets for DFR. Most of the available datasets fall into three categories: brain CT scans that tend not to contain the lower jaw, head and neck scans that tend not to contain the top part of the skull and mixed sets.

In addition, point clouds are not well-studied in the field of DFR. While point cloud data has gained recognition in various fields, including computer vision and 3D modeling, its application in DFR is relatively unexplored. The lack of in-depth studies and established methodologies for handling point cloud data specifically for DFR presents a challenge in optimizing the use of this data format.

Another challenge we identified was the limited capacity of open-source tools for point cloud processing. Open-source tools, such as MeshLab, have been widely used in the field of 3D modeling and reconstruction. However, when it comes to point cloud processing, these tend to have limited capabilities when compared to mesh models. Point cloud data requires specialized

algorithms for processing, registration, and manipulation, which tend to not be fully supported by existing open-source tools.

Furthermore, a relatively small number of research publications focused specifically on point cloud-based facial reconstructions. The field is still evolving, and while there is a growing interest in utilizing point cloud data, the number of published studies addressing its challenges and advancements remains relatively small.

A prominent problem is also the lack of standardized protocols and guidelines in the field of DFR, specifically for point cloud protocols. The absence of widely accepted standardized protocols and guidelines in DFR can lead to inconsistencies and discrepancies in the reconstruction process. Establishing a unified set of guidelines, including data acquisition, landmark placement, and facial approximation techniques, is essential to enhance the reliability and replicability of DFR across different forensic cases.

Addressing these challenges will require collaborative efforts from researchers, forensic practitioners, and software developers. Overcoming these obstacles will not only enhance the accuracy and efficiency of DFR, but also contribute to the development of standardized methodologies and tools that can be adopted by the forensic community at large.

3.2 Computed tomography (CT)

CT images have proven to be invaluable in the field of DFR, providing a wealth of anatomical information for accurate facial analysis. They allow detailed visualization of the craniofacial region, including the bones, soft tissues, and underlying structures. This imaging method captures high-resolution cross-sectional images, enabling forensic investigators to examine the intricate details of the skull and facial features. By leveraging advanced software tools, present in open-source programs, such as segmentation algorithms and 3D visualization, forensic professionals can manipulate the CT image data to extract the facial structures and generate accurate digital models for further analysis. The use of CT images in DFR enhances the ability to recreate the individual's appearance and supports investigative efforts in identifying the deceased or missing persons, contributing to the resolution of criminal cases and providing closure to families and communities.

3.3 Tools

The tools used in this research were chosen based on analyzed research in the field of DFR. The tools had to be free to use and verified in at least one other research paper.

3D Slicer (3D Slicer image computing platform, 2023) is a widely utilized open-source tool in the field of medical imaging and DFR, particularly used for CT images. With its user-friendly interface and powerful capabilities, 3D Slicer enables efficient segmentation, visualization, and manipulation of CT images, facilitating the extraction of meaningful information from complex datasets. By leveraging advanced algorithms and interactive visualization tools offered by 3D Slicer, forensic investigators can accurately assess facial features, perform bone/skin segmentation, and much more. This enables aiding in the identification and comparison components for individuals in forensic investigations.

MeshLab (Cignoni et al., 2008) is a powerful open-source tool that has found significant applications in DFR. It is also one of the few open-source tools that support the analysis of point cloud data. With its comprehensive set of tools and features, MeshLab provides a valuable platform for processing, editing, and analyzing 3D geometric models, including point clouds obtained from various sources, such as laser scanners or CT segmentations. By utilizing its advanced algorithms, users can refine and manipulate the models, enhancing the level of detail and accuracy necessary for facial analysis. Moreover, MeshLab facilitates measurements and comparisons of facial landmarks, aiding in the identification of individuals or the assessment of facial asymmetry and morphological variations.

CloudCompare (CloudCompare - Open Source project, 2023) is an open-source tool that plays a crucial role in the comparison of facial and bone point clouds. CloudCompare enables researchers and practitioners to perform detailed analysis and alignment of 3D point cloud data. In the context of facial and bone point clouds, this tool proves invaluable for assessing and quantifying the degree of correspondence between these two sets of data. By utilizing various algorithms and techniques provided by CloudCompare, such as registration, surface comparison, and distance measurements, researchers can precisely evaluate the alignment accuracy and spatial relationships between facial and bone point clouds.

3.4 Dataset

The CQ500 head CT dataset (CQ500 - V7 Open Datasets, 2023), provided by the Centre for Advanced Research in Imaging, Neurosciences and Genomics (CARING) in New Delhi, India, serves as a valuable clinical validation dataset. It contains 491 head CT scans and is one of the largest freely available datasets made for research purposes. The dataset consists of head CT scans obtained using various CT scanner models. The CT scanners used in these centers had a slice per rotation ranging from 16 to 128, offering a diverse range of image data. The dataset was obtained by extracting anonymized data from local Picture Archiving and Communication System (PACS) servers in compliance with internally defined Health Insurance Portability and Accountability Act (HIPAA) guidelines. One notable feature of the CQ500 dataset is the availability

of clinical radiology reports associated with the scans, providing valuable context and additional information. The dataset was collected in two batches, with Batch B1 encompassing all head CT scans taken at the specified centers over a 30-day period starting from November 20, 2017. Batch B2 was selected using a natural language processing (NLP) algorithm, which identified specific pathologies.

4 Materials and methods

In this section, we present our experiment.

4.1 CT scan processing and segmentation

In our study, we employed 3D Slicer version 5.2.1 for CT image processing. This choice was driven by the extensive features and capabilities offered by 3D Slicer, making it suitable for our research objectives. The CT images used in our analysis were in DICOM format and were obtained from the CQ500 dataset. As the dataset consisted of both brain scans and slice scans, we had to manually refine it and choose 15 of the most complete head scans for our analysis with ten of those being considered as larger (contained the orbital and nasal area) and five being considered smaller scans (containing only the supraorbital region).

To extract relevant structures (bone and face skin) from the CT images, we employed thresholding techniques using Hounsfield units (HU) in 3D Slicer. HU values represent the density of tissues and materials in CT scans (Ulinuha et al., 2019). To achieve the best results, we started by applying automatic thresholding supported within 3D Slicer first for bone, and then for soft tissue. For the best results to be achieved, we then adjusted the values to best suit each individual scan for their respective bone and soft tissue values. To ensure the quality and accuracy of our segmentations, we took additional steps to filter out any impurities or residual equipment artifacts. This involved employing various techniques such as noise reduction and manual removal of unwanted elements. These preprocessing steps were crucial for obtaining clean and precise segmentations, minimizing any potential interference or inaccuracies in subsequent analyses. Once the segmentations were completed, we saved the resulting structures as Standard Tessellation Language (STL) files. STL is a commonly used file format for 3D models, making it compatible with a wide range of software and facilitating further analysis and visualization of the segmented structures. The preliminary results are presented in Figure 1.

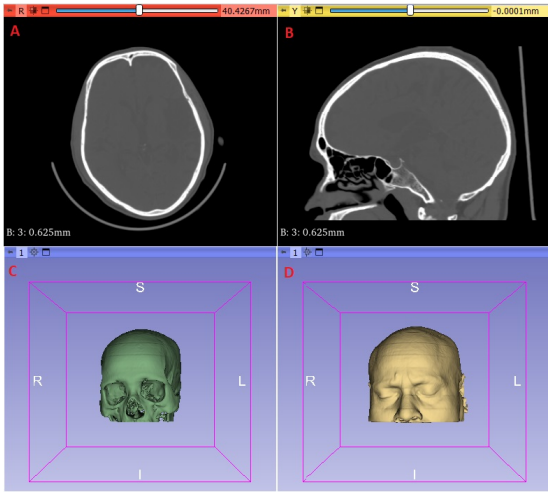


Figure 1: A and B original CT scan, C and D segmented 3D models

4.2 Point cloud creation and refinement

Next, we utilized MeshLab 2022.02 software for STL image processing, enabling comprehensive analysis and manipulation of the data. The STL images were loaded into MeshLab, providing a solid foundation for further processing steps. To ensure the accuracy and reliability of the data, we performed additional cleaning procedures, removing any artifacts or noise that may have been present in the original files. After the initial cleaning, the STL files were converted into point cloud format, enabling efficient handling and extraction of geometric information. The resulting data consisted of an average skull point cloud containing 980,939 points and an average face point cloud comprising 1,251,540 points. The preliminary point cloud models were saved as Polygon File Format or Stanford Triangle Format (PLY) files. The PLY format is a file format commonly used for storing 3D point cloud data. This representation provided a detailed and comprehensive depiction of the underlying geometry. An important advantage of point cloud data is the ability to precisely read and save the location of each point of the point cloud. This allows us to significantly reduce the number of points while still being able to reconstruct the original if it is needed.

To streamline the point cloud data, we removed all duplicate vertices. Duplicate vertices are multiple points that share the same coordinates. They can occur during data acquisition or processing, some of the most common reasons are errors in the data collection process, inaccuracies in scanning devices, or noise introduced during point cloud generation. They need to be removed to reduce data redundancy, improve processing speed and enhance the overall accuracy. To enhance the quality of the point cloud data, we applied a denoising process using the surface reconstruction screened poisson option in MeshLab. This algorithm utilized the oriented point sets to reconstruct watertight surfaces, effectively reducing noise and improving the overall accuracy of the point cloud data. The denoising step played a crucial role in refining the geometry and eliminating any unwanted artifacts or irregularities.

For further optimization, we employed point cloud simplification techniques, reducing the number of points while preserving the essential characteristics of the facial geometry. To confirm the model's accuracy we used the Compute Geometric Measures feature in Meshlab, making sure that the point clouds matched the original STL images using distance, area, and curvature measures. Through this process, we obtained an average simplified bone point cloud consisting of 72,039 points and an average simplified face point cloud comprising 199,747 points. This simplification step improved the computational efficiency and facilitated smoother data analysis. To ensure consistent alignment and comparison, we evaluated the normal vectors of the point sets. a normal vector represents the direction perpendicular to a surface at a specific point on that surface. When working with point clouds that represent surfaces (such as facial scans), normals can be evaluated at each point based on the surrounding points. Specifically, we compared the normals for each point with those of its neighboring points, with the number of neighbors set to 10. This step helped ensure accurate alignment and provided a reliable basis for subsequent measurements and comparisons.

4.3 Point cloud comparison and analysis

Finally, we utilized CloudCompare v2.13 for point cloud comparison. We wanted to determine whether the newly processed and original point clouds were properly aligned, if any distortions or outliers were present, and finally where the largest and smallest distances were measured, so that we can further focus on those areas in our continued research. We selected CloudCompare due to its renowned accuracy and reliability in measuring distances between point clouds. This choice ensured that our analysis was based on robust and dependable measurements. To begin the comparison process, we imported the point clouds (PLY files) into CloudCompare and aligned them accordingly. This alignment step was crucial to ensure proper spatial registration and enable accurate distance measurements. We also verified that the number of points in the point clouds was consistent and confirmed that it matched the numbers stated in the previous processing steps performed in MeshLab. This validation ensured consistency and provided confidence in the subsequent analysis.

Using the built-in compare distance functions in CloudCompare, we first calculated the distances between the newly processed point clouds and their original point cloud models and then conducted a comprehensive comparison of point cloud distances between the skull and skin point cloud models. This function allowed us to precisely measure the distances between corresponding points in the two point clouds, providing valuable insights into the variations and discrepancies between them. By leveraging the advanced capabilities of CloudCompare, we were able to conduct a detailed and accurate comparison of the point clouds, highlighting both the areas of agreement and areas of discrepancy. This information provided valuable insights into the spatial relationships and structural differences between the point clouds, enabling a comprehensive understanding of the variations in facial features and bone structures.

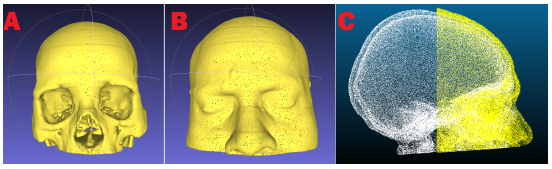


Figure 2: A and B preliminary dense point cloud models, C refined models

4.4 Point cloud acceptance criteria

In this study, we have set out 3 acceptance criteria. They were determined by criteria and best practices as seen in (Ulinuha et al., 2019, da Costa Moraes et al., 2014, Simmons-Ehrhardt et al., 2018, Hargreaves et al., 2021, Duan et al., 2015).

Visual appearance - the simplified point clouds had to undergo a visual inspection to assess their quality.

Noise and artifacts - noise, outliers, or artifacts in the simplified point cloud will be analyzed and needs to be nonexistent or kept at an acceptable minimal rate.

Alignment - newly created point clouds had to be aligned to the original mesh, preserving the shape of the originals.

5 Results

In our study, we aimed to present a preliminary point cloud reduction technique and analyze the variations in distance measurements when aligning, as well as compare the reduced point clouds. We were able to achieve a remarkable reduction in the sizes of the point clouds after processing. The results for the 15 analyzed CT scans are presented in Table 1.

Looking at the reduction percentages, we observe a consistent trend across all scans. For the skull, the reduction percentages range from 94.66% to 85.44%, while for the skin, they range from 87.69% to 68.10%. The findings demonstrate the effectiveness of the applied processing methods in achieving more streamlined and manageable point cloud models while preserving the essential anatomical details necessary for accurate DFR.

Examining the differences between the larger brain CT scans (scans 1-9 and 16) which contained both the orbital and nasal region and the smaller brain CT scans (scans 10-12 and 14, 15) which contained only the supraorbital region, we can observe that the reduction

Table 1: Results of point cloud reduction

Scan ID	Bone org.	Skin org.	Bone proc.	Skin proc.	Bone reduc.	Skin reduc.
1	1.159.590	2.045.188	61.924	281.910	94.65%	86.21%
2	1.120.242	2.028.364	77.528	277.887	93.07%	86.29%
3	1.102.358	1.445.892	76.585	213.907	93.05%	85.20%
4	1.192.294	1.417.956	80.645	205.808	93.23%	85.48%
5	1.339.260	1.481.048	87.839	199.328	93.44%	86.54%
6	1.360.758	1.459.832	79.263	186.797	94.17%	87.20%
7	1.361.004	1.530.458	79.196	194.076	94.18%	87.31%
8	1.322.850	1.409.420	70.568	200.176	94.66%	85.79%
9	1.328.444	1.644.510	72.701	202.355	94.52%	87.69%
10	454.076	484.502	65.712	152.372	85.52%	68.55%
11	444.036	474.766	64.411	150.623	85.49%	68.27%
12	478.280	489.290	68.142	153.354	85.75%	68.65%
13	1.156.004	1.888.552	66.217	267.902	94.27%	85.81%
14	443.306	479.806	64.502	153.039	85.44%	68.10%
15	451.582	493.520	65.347	156.673	85.52%	68.25%

percentages for the skull and skin are relatively consistent across both groups. This suggests that the processing techniques were equally effective in reducing point cloud sizes regardless of the size of the analyzed CT scan.

We compared the processed point clouds to the originals as well as the original mesh models to test for deviations. All skull and skin point cloud models showed perfect alignment and no outliers or deviations were present. When analyzing the point cloud models more closely we found that the average maximum error for the skull model was 0,730368 and 0,805882 for the skin model, being well within the acceptable range for DFR. In addition when comparing the point clouds with the original mesh segmentations we found the average distance between the skin mesh and point cloud was 0,00331060 and between the skull mesh and point cloud was 0,00260404. The largest distances were found in the areas of the back of the head and in the largest and most complex CT scans. This can be attributed to factors such as segmentation accuracy, simplification process limitations, and alignment considerations. Specifically, as the point clouds underwent an additional cleaning process when compared to the original STL files. In addition, a mitigating factor is the fact that the back of the head is not used for facial reconstruction, and is cut out for easier processing.

In addition, upon aligning and comparing the reduced point clouds and comparing them to the originals, we observed interesting patterns in the distance measurements across different facial regions between the bone and skin. The nasal area exhibited the largest distances, indicating potential differences in the shape and alignment of the nose between the point clouds. This finding suggests that the nose may have distinct variations and deviations in different individuals or scanning sessions, which could impact subsequent DFR or analyses.

Conversely, we found that the forehead area, superorbital area, and certain parts of the cheeks exhibited the smallest distances. These regions displayed closer alignment and similarity between the point clouds, indicating consistent facial features in these areas. This observation suggests that these regions may have less variability and are more reliable for precise DFR and analysis.

Overall, our results highlight the effectiveness of the point cloud reduction techniques employed, as well as the importance of accurately aligning and comparing point clouds for reliable DFR. These findings contribute to the advancement of DFR methodologies and provide valuable insights into the variations and consistencies within point cloud data.

6 Discussion

The results of our study demonstrate a significant reduction in the size of the point cloud models while maintaining clarity and preserving a high level of detail. The smaller size of the point clouds enables faster processing and analysis, as computational operations can be performed more efficiently. This can lead to improved turnaround times and increased productivity in the field of DFR.

It is observed that larger and more complex skull and facial skin models tend to respond better to the processing, resulting in a larger reduction of points of the point cloud, while still preserving important details. This could be because larger point clouds inherently contain more data points, capturing a higher level of detail and intricacies of the scanned object. This increased level of detail allows for a more accurate representation of the bone and skin structures, resulting in improved results. Additionally, larger point clouds provide a denser

Table 2: Acceptance criteria and alignment check

Scan ID	Avg. Dist. Skin	Avg. Dist. Skull	Error Skin	Error Skull
1	0.00281830	0.00256553	0.857123	0.776819
2	0.01277510	0.00469573	0.836600	0.754045
3	0.00394878	0.00265517	0.825804	0.753992
4	0.00106718	0.00321285	0.843168	0.753699
5	0.00332792	0.00288330	0.823054	0.721549
6	0.00234217	0.00609075	0.861356	0.720610
7	0.00200455	0.00609103	0.887325	0.720610
8	0.00358644	0.00336779	0.843581	0.820974
9	0.01041560	0.00344334	0.871147	0.820132
10	0.00090295	0.00016398	0.741171	0.668090
11	0.00045204	0.00013618	0.723477	0.667988
12	0.00155575	0.00022674	0.727078	0.670085
13	0.00319508	0.00314697	0.808749	0.770929
14	0.00054517	0.00025358	0.717866	0.668009
15	0.00072191	0.00012763	0.720730	0.667982

sampling of the surface, which leads to a more precise reconstruction. The additional points enable the capture of finer features, such as subtle variations in curvature or small surface irregularities.

Furthermore, larger point clouds provide a more robust foundation for subsequent processing steps, such as alignment and comparison. The increased number of points enhances the stability and reliability of these operations, allowing for more accurate registration and analysis of the models.

However, it is noted that there are certain areas that could be improved, the data suggest that problems with segmentation mainly occurred when segmenting the left side of the faces due to distortion caused by the subjects being pressed against the scanner. Addressing these issues and refining the segmentation process could result in more precise and accurate models.

Additionally, the data reveals the presence of residual soft tissue in some skin scans, which could not be segmented and was subsequently filtered out during the processing in MeshLab. This implies that there might be a need for more advanced segmentation techniques or additional manual intervention to overcome challenges related to soft tissue segmentation.

Furthermore, some smaller bone fragments located inside the skull also needed to be filtered out in a similar manner. This highlights the importance of carefully assessing the specific requirements of the DFR application and tailoring the processing steps accordingly to remove any unwanted or irrelevant structures while maintaining the integrity of the main bone and skin models.

In addition, there is still room for further improvement in the processing of the point cloud models. Advanced point cloud refinement techniques, such as noise reduction, outlier removal, and surface reconstruction, can be implemented to enhance the quality and accuracy of the data. To further improve processing we plan on experimenting with more prominent tools, such as MATLAB.

Another avenue for improvement lies in feature-based compression methods. By selectively preserving important anatomical features and prioritizing the representation of critical facial landmarks and regions, we can maintain the necessary precision for DFR while reducing the overall data size. This approach allows us to strike a balance between size and detail, optimizing the efficiency of storage and processing resources.

Finally, the construction of a dedicated head CT dataset for DFR would be an important tool. Having a dedicated dataset for DFR would ensure that the data is standardized and optimized for accurate and reliable reconstruction outcomes. Such a dataset could encompass a diverse range of head CT scans with varying anatomical features, STT, and pathologies, providing a comprehensive representation of the human population. Additionally, it would enable researchers and practitioners to focus their efforts on refining algorithms and techniques specifically tailored for DFR, leading to advancements in accuracy, speed, and efficiency. Moreover, a well-curated dataset would facilitate benchmarking and comparison of different reconstruction approaches, promoting collaboration and knowledge sharing within the research community. Ultimately, the availability of a dedicated head CT dataset would greatly contribute to the development and improvement of DFR methods, enabling enhanced clinical applications, forensic investigations, and facial prosthetics.

Overall, while the presented data demonstrates the successful reduction of point cloud size and the preservation of important details in skull and facial skin models, there are areas that warrant further investigation and refinement. Improving CT image segmentation,

addressing challenges related to distorted scans and soft tissue segmentation, and adapting the processing steps to filter out unwanted structures will contribute to the continued enhancement of the accuracy and precision of the point cloud models, ultimately advancing the field of DFR.

7 Conclusion

In conclusion, our paper has covered various aspects of DFR, CT image processing, and point cloud comparison. We explored different open-source tools and techniques used in these processes, including 3D Slicer, MeshLab, CloudCompare, and their respective functionalities. We also discussed the importance of data formats, such as DICOM and STL.

Furthermore, we examined the outcomes of the image processing steps performed in 3D Slicer, MeshLab, and the reduction in point cloud size achieved through simplification and cropping. The reduction in point cloud size was substantial, with the bone point clouds being reduced by 91.13% on average and the skin point clouds being reduced by 80.35%. We also analyzed the alignment and comparison of the processed point clouds with the original point clouds and found perfect alignment with no outliers or deviations. Finally, we were looking at the relationship between the skull and skin point clouds and found that the largest distances were observed in the nasal area, while the smallest distances were recorded in the forehead, superorbital, and cheek regions.

Moreover, we discussed potential areas for improvement, including employing advanced segmentation and point cloud refinement techniques, feature-based compression methods, and adaptive level of detail approaches to enhance the precision of point cloud models while keeping them at a small size. These strategies can contribute to more accurate DFR, faster processing, and improved efficiency in facial analysis.

Overall, the presented results highlight the complexity and potential of DFR, CT image processing, and point cloud comparison. We demonstrated the importance of utilizing appropriate tools, considering data formats, and implementing optimization techniques to achieve precise and efficient results. Continued advancements in these areas will pave the way for significant applications in fields such as forensic reconstruction, medical imaging and beyond.

References

Miranda, G. E., Wilkinson, C., Roughley, M., Beaini, T. L., & Melani, R. F. H. (2018). Assessment of accuracy and recognition of three-dimensional computerized forensic craniofacial reconstruction. *PLoS One*, 13(5), e0196770.

Ulinuha, M. A., Yuniarno, E. M., Hariadi, M., & Purnama, I. K. E. (2019, March). Extraction of skull and face surfaces from CT images. In 2019 *International Conference of Artificial Intelligence and Information Technology (ICAIIT)* (pp. 37-40). IEEE.

Simmons-Ehrhardt, T., Falsetti, C., Falsetti, A. B., & Ehrhardt, C. J. (2018). Open-source tools for dense facial tissue depth mapping of computed tomography models. *Human biology*, 90(1), 63-76.

da Costa Moraes, C. A., Dias, P. E. M., & Melani, R. F. H. (2014). Demonstration of protocol for computer-aided forensic facial reconstruction with free software and photogrammetry. *Journal of Research in Dentistry*, 2(1).

Hargreaves, M., Ting, D., Bajan, S., Bhavnagri, K., Bassed, R., & Chang, X. (2021, November). A Generative Deep Learning Approach for Forensic Facial Reconstruction. In 2021 *Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-7). IEEE.

Duan, F., Huang, D., Tian, Y., Lu, K., Wu, Z., & Zhou, M. (2015). 3d face reconstruction from skull by regression modeling in shape parameter spaces. *Neurocomputing*, 151, 674-682.

Ma, T. C., Nguyen, D. T., Dinh, Q. H., & Bui, T. D. (2011, October). 3D facial reconstruction system from skull for Vietnamese. In 2011 *Third International Conference on Knowledge and Systems Engineering* (pp. 120-127). IEEE.

Shui, W., Zhou, M., Deng, Q., Wu, Z., & Duan, F. (2010, November). 3D craniofacial reconstruction using reference skull-face database. In 2010 *25th International Conference of Image and Vision Computing New Zealand* (pp. 1-7). IEEE.

3D Slicer image computing platform (2023). Retrieved from https://slicer.org/

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., & Ranzuglia, G. (2008, July). Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference* (Vol. 2008, pp. 129-136).

CloudCompare - Open Source project (2023). Retrieved from https://www.danielgm.net/cc/

CQ500 - V7 Open Datasets (2023). Retrieved from https://www.v7labs.com/open-datasets/cq500

Bridging Blockchain Technology and Humanitarian Demining: A Novel Concept for Decentralized Storage of Landmine and UXO Locations

Marko Horvat, Klara Krmpotić
University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Applied Computing
Unska 3, HR-10000 Zagreb, Croatia
{marko.horvat3, klara.krmpotic}@fer.hr

Andrija Krtalić
University of Zagreb
Faculty of Geodesy
Kačićeva 26, HR-10000 Zagreb, Croatia
kandrija@geof.hr

Amila Akagić
University of Sarajevo
Faculty of Electrical Engineering
University Campus, Zmaja od Bosne bb, 71000 Sarajevo, Bosnia and Herzegovina
aakagic@etf.unsa.ba

Abstract. *This paper presents a proposal to explore decentralized storage solutions and the potential of blockchain technologies, in particular, Filecoin network and Storage.NFT framework, for the secure and reliable storage of landmine locations and descriptive information for non-technical survey in humanitarian demining. The presented innovative approach offers numerous benefits such as decentralized and secure storage, tamper-proof data, trust, and transparent data access by implementing reliable blockchain technology in the field of demining. Scalability, transparency, reliability, integration with existing systems, legal compliance, and cost-effectiveness are issues that need to be further explored. Using blockchain technology to store information in minefield records can significantly improve the efficiency, accessibility, and safety of humanitarian demining operations.*

Keywords. Distributed Storage Networks, Filecoin Blockchain, Non-Fungible Tokens, Non-Technical Survey, Mine Action

1 Introduction

Humanitarian demining is a critical process in post-conflict areas that aims to remove landmines and explosive remnants of war and ultimately save lives (Dorn, 2019; Camacho-Sanchez, Yie-Pinedo & Galindo, 2023). The Non-Technical Survey (NTS) in demining involves the collection, analysis, and storage of minefield records, which can be critical to successful mine clearance operations (Bajić et al., 2021).

Advances in remote sensing technologies such as magnetic detectors, ground penetrating radars, and hyperspectral aerial images captured by Unmanned Aerial Vehicles (UAVs), aircraft, and satellites have greatly improved the efficiency and accuracy of humanitarian demining (Krtalić & Bajić, 2019). These cutting-edge technologies, together with the application of novel Artificial Intelligence (AI) and Deep Learning (DL) algorithms, have greatly improved identification of landmines and Unexploded Ordnance (UXO) with a high degree of certainty (Baur et al., 2020; Horvat et al., 2022; Osmankovic et al., 2022; Horvat et al., 2023).

However, sharing and managing the data generated by these sensor technologies and non-technical survey analysis remains a significant challenge, particularly when it comes to secure access, trustworthiness, and collaboration among a diverse group of stakeholders, including governments, nongovernmental organizations, and international organizations (Horvat et al., 2023).

To address these concerns, a novel blockchain-based concept was proposed for data management for non-technical survey in mine-action. Specifically, in the proposed concept minefield records documents with information on secondary indicators about Suspected Hazardous Area (SHA) are stored on the blockchain. This innovative solution could securely and reliably store geolocation, type, presence probability and other relevant description of suspected UXO in an expressive and semantically rich user-defined format. This approach, as a continuation of our

previous work (Krtalić & Bajić, 2019; Bajić et al., 2021; Horvat et al., 2022; Horvat et al., 2023), could enable further processing of such data, including data mining, automated reasoning, and decision support in mine action. The core benefit of the proposed concept lies in its ability to utilize inherent properties of blockchain technology in a novel manner to provide a transparent, immutable, and decentralized system for managing important demining data. This not only increases the security and reliability of data storage, but also facilitates seamless collaboration between stakeholders, allowing for efficient execution of humanitarian demining operations.

The reminder of the paper is organized as follows; Section 2 presents the novel concept of decentralized management of mine records on blockchain, focusing on the Filecoin and NFT.Storage technologies for its implementation. Section 3 explains on a real-life example how these technologies can be applied to minefield records blockchain storage. The benefits of the presented concept are listed in Section 3.1 and Section 3.2 discusses potential areas for future research. Finally, Section 4 concludes the paper and reiterates the most important implications of the proposed concept.

2 The Concept of Decentralized Mine Action Data Management using Blockchain Technologies

Blockchain has revolutionized the way we think about data storage and management transaction (Iansiti & Lakhani, 2017). One of the key benefits of blockchain technology is its ability to provide an immutable and transparent ledger, which can secure transactions without relying on a central authority.

This decentralized approach provides several benefits, including decentralized control, distributed architecture and applications, data immutability, and network consensus to approve any transaction (Wang et al., 2019). By leveraging these inherent advantages, blockchain technology enables streamlined and secure data management for businesses across various industries.

A block is a fundamental component that plays a crucial role in ensuring the security, immutability, and decentralization of the blockchain network. It is a data structure that contains a set of transactions and additional information, which typically includes block header and transactions. A header contains metadata about the block, while transactions are actual data entries or operations that are being recorded on the blockchain. Blocks are connected in a chain, hence the name blockchain.

The size of the blocks is predefined. Once the block is filled, it is "closed" and connected to the previous block called Parent block, thus creating a chain. Each transaction that is written into the block has its own timestamp when it was added to the block. This property makes blockchain a secure information storage technology as it is almost impossible to change the timestamp. All blocks are written to all devices connected in the network, which means that if a change is to be made, it must be achieved by consensus, i.e., by unanimous decision of all related stakeholders. There are various open blockchain networks, one of them is Filecoin, it is primary decentralized, peer-to-peer storage network (Vyzovitis et al., 2020).

2.1 Filecoin Blockchain

Filecoin (FIL) is a decentralized storage network built on a blockchain-based market, designed to enable users to store and retrieve data in a secure, reliable, and efficient manner (Bauer, 2022a; Psaras & Dias, 2020). The Filecoin network's primary function is to provide storage services, relying on a native token as an incentive mechanism for storage providers (Benet & Greco, 2018). Its unique consensus mechanism, Proof-of-Replication (PoRep), and Proof-of-Spacetime (PoSt), ensures that the data is stored securely and can be retrieved when needed (Dai et al., 2020). In this regard, the Filecoin network trades storage like a commodity. Its role is to ensure the safe and reliable storage of data by involving various storage providers for this purpose. The Filecoin network thus acts as a marketplace for data storage, bringing together users who need storage with parties willing to provide storage capacity.

This storage network is built on peer-to-peer principle, using cryptography to provide security of files over longer periods. Users pay for their files to be stored on available storage providers, which guarantee safe storage over time (Bamakan et al., 2021). Filecoin facilitates open markets for available storages for storing and retrieving files that anyone can participate in, it is not controlled by single company, hence decentralization (Fisch et al., 2018).

Filecoin can also store Non-Fungible Tokens (NFT). NFT is type of a cryptocurrency, that originated from token standard of Ethereum. Each NFT is created unique and irreplaceable. They have distinctive source code and are bound with virtual/digital properties, called metadata, making it possible to individually distinguish between each other. These tokens can be any digital object. NFT is a type of cryptocurrency derived by Ethereum using smart contracts. Every NFT contains details of ownership, safe mechanisms, and metadata, which is the main difference between normal currency and NFTs. This feature comes from ERC-721 standard which implements application programming interface (API) inside smart contracts (Bauer, 2022b). It also offers trades with tokens between stakeholders, overview of an account, shows current owner of each token and list of available tokens on the network (Wang, Wang & Qin, 2021).

An overview of Filecoin architecture is shown in a diagram (Figure 1). The diagram explains user-system interaction, where users upload or download data through a client, which then interfaces with the NFT.Storage API. The NFT.Storage API is responsible for data storage. It stores metadata in a dedicated Metadata Database and the actual data in an InterPlanetary File System (IPFS) cluster. Then the data is pinned from the IPFS cluster to the Filecoin network. The final step of the process is the Filecoin network, which guarantees the storage of the data by different storage providers.

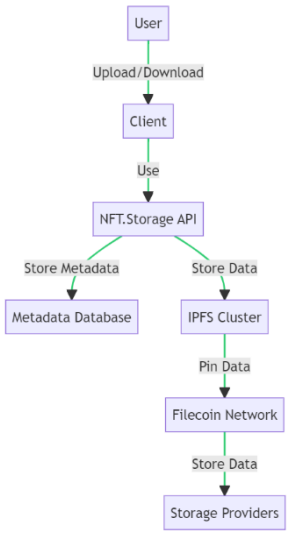


Figure 1. Simplified flowchart of the Filecoin and NFT.Storage network architecture.

As previously mentioned, Filecoin can store NFTs, which is accomplished through specialized service, NFT.Storage. By using this service, the storage of tokens on a decentralized network is simple and reliable for stakeholders, without the obligation of setting up and maintaining their own infrastructure.

2.2 NFT.Storage Framework

NFT.Storage is an open software framework that enables the creation and management NFTs representing digital assets and their storage on decentralized networks ("NFT.Storage Docs", 2023; Wyatt, 2022a; Wyatt, 2022b). NFTs are unique tokens representing ownership and provenance of digital assets, ensuring their uniqueness and traceability (Herian et al., 2021). By combining NFTs with decentralized storage networks, NFT.Storage provides a secure, verifiable, and tamper-proof storage solution for digital assets, with a goal to store all NFT data as public goods. As such, NFT.Storage presents and ideal choice for the implementation of the presented concept of storing demining documents as digital assets on the blockchain.

NFT data is stored using IPFS as a Uniform resource identifier (URI), e.g., „`ipfs://…`", making the connection between token and its data transaction ("Introducing NFT.Storage", 2021). Once the data is uploaded, stakeholder will receive IPFS hash which represents data that is to be stored, Content IDentifier (CID). This can be accessed through URLs, such as „`ipfs://<CID>`". As mentioned before, blockchain technology makes a copy on every node in the network, which is here used to store multiple copies of data on network. The most important copies are the ones on IPFS server (monitored by NFT.Storage) and on Filecoin. To access any of the data for certain NFT, all that is needed is to provide the CID. Data can be retrieved through IPFS gateway or directly from IPFS command line. This process is illustrated in Figure 2.

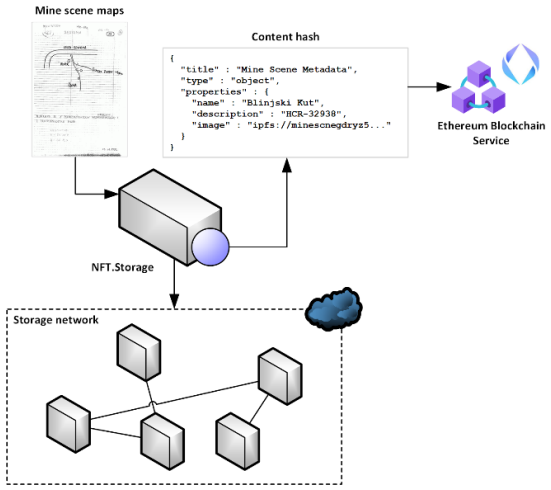


Figure 2. NFT.Storage architecture overview incorporating content addressing, provable storage and resilient retrieval features. Adapted from ("Introducing NFT.Storage", 2021).

3 Utilizing Filecoin and Storage.NFT for Minefield Records Storage

Minefield records are very important demining documents. As can be seen in Figure 3, such documents are hand-drawn and may be difficult to ready, yet they contain important information needed to determine the type, shape and location of secondary mine indicators. The maps contain various information written on this document – for instance when it was made, which mines were found in this are – that helps in describing Suspected Hazardous Area (SHA) (Bajić et al., 2011).

It is important to note that minefield records should be made available and easily accessible to interested stakeholders while ensuring the integrity of the information. For this reason, the use of the proposed concept is beneficial to both government agencies and the public.

In the proposed concept all pertaining information in the minefield records are stored as metadata in NFT, with NFT being the digital object of the scanned minefield record itself. To store this metadata and the demining document as NFT, there must exist function to create a file from uploaded image and append its metadata.

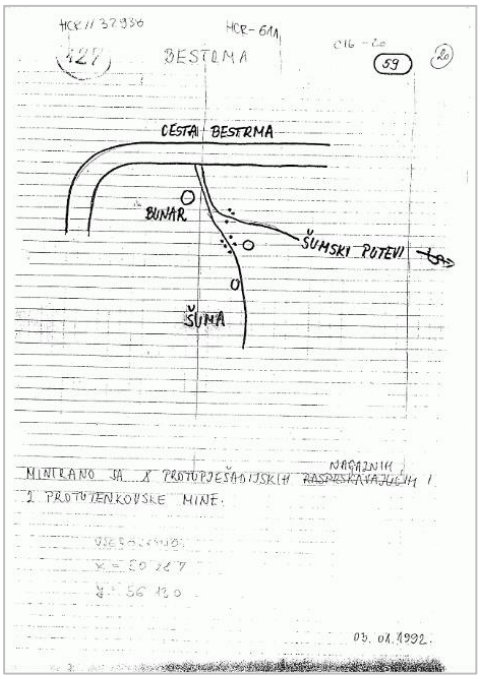


Figure 3. A real-life example of a typical hand-drawn minefield record document (kindly provided by the Croatian Mine Action Centre – HCR) that is stored on blockchain as NFT.

Figure 4 shows a source code for creating a new NFT and storing it in FIL. The code imports `NFTStorage` class from `NFT.Storage` package, that takes users' API key and uses it to store and connect NFT to its rightful owner.

```
async function storeNFT(imagePath, name,
description, DocType, Iink, GKZona, GKE, GKN,
UNSektor, OznakaUN, Oznaka, Zupanija, Opcina,
Naselje, VojskalD, Date, POM, PPM, Ostala,
Provjere, TipPolja) {
// load the file from disk
    const image = await fileFromPath(imagePath)

// create a new NFTStorage client using our API
key
    const nftstorage = new NFTStorage({token:
NFT_STORAGE_KEY })

// call client.store, passing image & metadata
    return nftstorage.store({image, name,
description,DocType, Iink, GKZona, GKE, GKN,
UNSektor, OznakaUN, Oznaka, Zupanija, Opcina,
Naselje, VojskalD, Date, POM, PPM, Ostala,
Provjere, TipPolja
    })
}
```

Figure 4. The usage of `storeNFT()` function.

The `main` function calls `storeNFT` function and passes attributes for the metadata processing. Also, as Figure 4 shows, from given parameter `imagePath`, program calls `fileFromPath` function which then returns object of class `File`. This is crucial because functions that are called after can work with objects of type File or Blob. After this function call has successfully finished, the program will try to execute function `store` from `NFTStorage` class object. The `store` function takes the metadata, passed in JSON format according to theERC-1155 standard, and creates a new NFT data file. Then the function returns the address IPFS URL from which the user can access the NFT data file and retrieve the metadata.

In Figure 5 is an example of how to call this code and provide all the required parameters and what the outcome will be. Every argument written behind the `upload.mjs` is a document data for NFT's metadata.

```
node upload.mjs 32938.gif 'dd32938' 'Deminig
document no 32938' 'DeminingDocument' '32938' '5'
'611360' '5028700' 'None' '0' '59' 'SM' 'Sunja'
'Bestrma' 'None' '03/01/1992' '2' '8' '0' 'False'
'Mix'
```

Figure 5. Example of `upload.mjs` file call.

In Figure 5 can be seen how all the necessary parameters are provided when `upload.mjs` is called. In this certain example it is import that user provides all the parameters because otherwise the program will return error and ask the user to provide new, complete data. This will return a token, which has two elements, `ipnft` and `url`. `Ipnft` is used in URL, and this URL is IPFS URL from which the user can access metadata. It is crucial to enclose all parameters intended to be stored as strings within quotation marks. This practice enables parsers to accurately identify the beginning and ending of each parameter, ensuring proper data processing. An example of JSON metadata for creating NFT with a minefield record is shown in Figure 6.

There are certain components of metadata that cannot be changed such as the CID and contract address, but other information in metadata can be changed. The metadata may be changed in case if new field of information is needed or a more precise location coordinates are defined. Figure 7 shows TypeScript code of how this can be done using Metaplex service ("Metaplex Docs", 2023; Camacho-Sanchez, Yie-Pinedo & Galindo, 2023; Peterfay, 2022).

Metaplex is an open framework and high-level application protocol designed for creating, minting and deploying NFTs ("Metaplex Docs", 2023; Camacho-Sanchez, Yie-Pinedo & Galindo, 2023; Peterfay, 2022). It consists of a set of tools and smart contracts that provide NFT developers with a comprehensive set of services that simplifies the process of creating and deploying NFTs.

```
{
    "Date": "03/01/1992",
    "DocType": "DeminingDocument",
    "GKE": "611360",
    "GKN": "5028700",
    "GKZona": "5",
    "Iink": "32938",
    "Naselje": "Bestrma",
    "Opcina": "Sunja",
    "Ostala": "0",
    "Oznaka": "59",
    "OznakaUN": "0",
    "POM": "2",
    "PPM": "8",
    "Provjere": "False",
    "TipPolja": "Mix",
    "UNSektor": "None",
    "VojskalD": "None",
    "Zupanija": "SM",
    "description": "Deminig document no 32938",
    "image": {
        "/":
"bafybeictkpxa4siq4yk7e5aakih5uzmhayob3ou2z3rcysu
jistrxcx4vu"
    },
    "metadata.json": {
        "/":
"bafkreigdqt2ic6piktw56kb6oxwgyceyiyhw7kq44vwrtzr
ccm2cwr4nim"
    },
    "name": "dd32938",
    "type": "nft"
}
```

Figure 6. An example of JSON configuration file for storing minefield metadata.

For the code provided in Figure 7, user is required to create JSON for metadata, which will be uploaded by `NFTClient` object which has a method `uploadMetadata`. After this is successfully executed, the URI of the uploaded JSON metadata will be linked to NFT and new, updated metadata will be added.

```
const NEW_METADATA = {
    description: 'New description!',
    attributes: [
        {trait_type: 'GKN', value: '5028701'},
        {trait_type: 'GKZone', value: '4'},
        {trait_type: 'GKE', value: '611361'}
    ]
};

async function uploadMetadata(imgUri: string,
description: string, attributes: {trait_type:
string, value: string}[]) {
    const { uri } = await METAPLEX
        .nfts()
        .uploadMetadata({
            name: "My new name",
            description: description,
            image: imgUri,
            attributes: attributes
        });
    return uri;
}

async function updateNft(nft:Nft|Sft,
metadataUri: string, newName: string) {
    await METAPLEX
        .nfts()
        .update({
            name: newName,
            nftOrSft: nft,
            uri: metadataUri
        });
}
```

Figure 7. Upload of new minefield metadata using Metaplex open framework.

NFT and its minefield metadata can be deleted from user's account, but despite this the nodes in the IPFS storage network can still contain a copy of this data indefinitely. This is why anything that is stored on blockchain is considered permanent. To delete an NFT from blockchain, it is important to provide correct CID. An example of this function is shown in Figure 8.

```
async function deleteNFT(cid) {
const nftstorage = new NFTStorage({ token:
NFT_STORAGE_KEY })
    res = await nftstorage.delete(cid)
}
```

Figure 8. The usage of `deleteNFT()` function.

The UML use case diagram in Figure 9 shows the three basic operations of the proposed concept for the NFT blockchain-based mine record storage service: creation, updating, and deletion of NFTs enabled by interactions between the user, the Metaplex framework, and the Filecoin network.

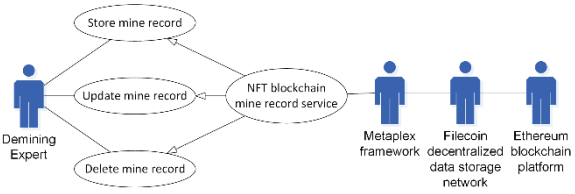


Figure 9. UML use-case diagram representing the interactions between the demining expert and the proposed system for NFT blockchain-based mine record storage service using Metaplex and Filecoin for the creation, updating, and deletion of NFTs containing minefield records.

3.1 Benefits of the Proposed Concept

The integration of Filecoin and Storage.NFT for minefield record storage in humanitarian demining provides several advantages including: (*i*) decentralized and secure storage, (*ii*) immutability, and (*iii*) transparency and transparent data access.

First, concerning decentralized and secure storage, the Filecoin network is important for guaranteeing data redundancy and distributed storage. This reduces the risk of data loss significantly while also offering enhanced protection against unauthorized access (Benet & Greco, 2017).

Second, data immutability and tamper resistance are also very important features of the proposed approach. Blockchain technology ensures that once data is added to the blockchain, it cannot be altered or deleted. This provides a reliable and tamper-proof record of landmine and UXO locations enhancing trust in the data. By using Storage.NFT, minefield records are represented as unique NFTs, ensuring their authenticity and provenance, and preventing unauthorized changes (Herian et al., 2021).

The third advantage is transparency and transparent access to data. The decentralized nature of the blockchain allows all stakeholders involved to access the same data set, which promotes transparency and facilitates collaboration in demining efforts. The blockchain-based nature of Filecoin and Storage.NFT allows for transparent and verifiable data access, which facilitates collaboration and trust between various demining stakeholders such as Mine Action Centers (MACs), companies conducting the demining and the public.

One of the major drawbacks of adding digital documents to the blockchain is that uploads to the blockchain are not real-time. Since consensus must first be reached, the allowable wait time is 48 hours. This is more of a disadvantage for the Blockchain than for demining and uploading demining documents, as demining occurs after the end of hostilities and the non-technical aspect of demining (i.e., collecting the data) requires more time.

In addition, since Filecoin is based on the Etherum blockchain network, there are certain file size limits. For each upload, the user is only allowed to upload a total of 31 GiB. This is still a lot of space, but every user should be aware of this limitation.

Furthermore, for future use, instead of just the demining documents, there should be an option to also save the edge points of polygons after multiple analysis of all the collected data from the suspected threat area. Moreover, to make uploading demining documents easier, there should be a web page with user friendly GUI, where end user could input all necessary metadata and image, and after successful upload get JSON with this metadata. Also, a possibility for delete and update options for NFTs which will call functions described in this chapter.

3.2 Future Research

Considering future work, several challenges need to be addressed when implementing blockchain technologies for minefield record storage in humanitarian demining.

Primarily, these include data validation and verification, scalability, integration with existing mine action systems and databases, legal and regulatory compliance (i.e., GDPR and other potential legal issues related to the use of decentralized storage networks in humanitarian demining), and cost effectiveness. Developing a mechanism to validate and verify the landmine and UXO data entered into the blockchain may involve processing and cross-referencing sensor data from multiple sources, correlating the information with existing databases, or using artificial intelligence algorithms to identify potential discrepancies.

Furthermore, in terms of scalability it is necessary to assess the blockchain network's ability, such as the Filecoin, to efficiently store and retrieve large volumes of minefield records and ensure seamless integration with existing humanitarian demining tools.

Thirdly, to fully realize the potential of the blockchain in humanitarian demining efforts, future research should focus on developing a proof-of-concept system, conducting comprehensive performance evaluations, assessing security, exploring interoperability, and evaluating usability. A proof-of-concept implementation will demonstrate the feasibility of integrating blockchain based NFTs for minefield records storage, retrieval, and management. A performance evaluation will determine the suitability of the proposed storage solution for humanitarian demining, while the security analysis will ensure the protection of minefield records and the safety of demining personnel. Interoperability with other decentralized technologies, such as distributed ledger systems, can improve collaboration and data sharing among demining stakeholders.

Finally, evaluation of user experience and usability will facilitate adoption and integration into existing workflows and allow researchers to further explore the potential of blockchain-based storage solutions to improve the efficiency, security, and reliability of minefield record storage in humanitarian demining operations.

4 Conclusion

This paper proposes a novel concept of using blockchain technologies and NFTs for secure and reliable storage of minefield records in humanitarian demining operations. The technical feasibility of the proposed concept has been explained and demonstrated on a real-world example using existing blockchain technologies: Filecoin blockchain platform and Storage.NFT framework.

Other advantages of this approach, such as decentralized and secure storage, tamper-proof data, incentives for storage providers, and transparent data access, were considered. Challenges such as scalability, integration with existing systems, legal and regulatory compliance, and cost-effectiveness should be considered in future research.

The use of blockchain technologies for distributed storage of minefield records on the Internet has the potential to significantly improve the accessibility, efficiency, and security of humanitarian demining operations and should be fully explored in the future.

References

Bajić, M., Matić, Č., Krtalić, A., Čandjar, Z., & Vuletić, D. (2011). Research of the mine suspected area. *HCR Centre for Testing, Development and Training Ltd*.

Bajić, M., Gold, H., Horvat, M., Krtalić, A., Laura, D., & Muštra, M. (2021). The novel paradigm for a decision support system of the aerial non-technical survey. In *Book of papers of 17th*

International Symposium "MINE ACTION 2021", 62-68.

Bamakan, S. M. H., Nezhadsistani, N., Bodaghi, O., & Qu, Q. (2021). A decentralized framework for patents and intellectual property as nft in blockchain networks.

Bauer, D. P. (2022). Filecoin. In *Getting Started with Ethereum: A Step-by-Step Guide to Becoming a Blockchain Developer* (pp. 97-101). Berkeley, CA: Apress.

Bauer, D. P. (2022). ERC-721 Nonfungible Tokens. In *Getting Started with Ethereum: A Step-by-Step Guide to Becoming a Blockchain Developer* (pp. 55-74). Berkeley, CA: Apress.

Baur, J., Steinberg, G., Nikulin, A., Chiu, K., & de Smet, T. S. (2020). Applying deep learning to automate UAV-based detection of scatterable landmines. *Remote Sensing*, *12*(5), 859.

Benet, J., & Greco, N. (2018). Filecoin: A decentralized storage network. *Protoc. Labs*, 1-36.

Benisi, N. Z., Aminian, M., & Javadi, B. (2020). Blockchain-based decentralized storage networks: A survey. *Journal of Network and Computer Applications*, *162*, 102656.

Camacho-Sanchez, C., Yie-Pinedo, R., & Galindo, G. (2023). Humanitarian demining for the clearance of landmine-affected areas. *Socio-Economic Planning Sciences*, *88*, 101611.

Cui, S., Zhao, G., Gao, Y., Tavu, T., & Huang, J. (2022, November). VRust: Automated Vulnerability Detection for Solana Smart Contracts. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (pp. 639-652).

Dorn, A. W. (2019). Eliminating hidden killers: How can technology help humanitarian demining?. *Stability: International Journal of Security and Development*, *8*(1).

Fisch, B., Bonneau, J., Greco, N., & Benet, J. (2018). Scaling proof-of-replication for filecoin mining. *Report. Protocol Labs Research*.

Herian, R., Di Bernardino, C., Chomczyk Penedo, A., Ellul, J., Ferreira, A., von Goldbeck, A., ... & Siedler, N. L. (2021, July). NFT–Legal Token Classification. EU Blockchain Observatory & Forum.

Horvat, M., Krtalić, A., Akagić, A., Krmpotić, K., & Skender, S. Humanitarian Demining Using Data Observatories and Data Lakes. In *Book of papers of 19th International Symposium "MINE ACTION 2023"*, 47-51.

Horvat, M., Krtalić, A., Bajić, M., Muštra, M., Laura, D., & Gold, H. (2022). MINEONT: A proposal for a core ontology in the aerial non-technical survey domain. In *Book of papers of 18th International Symposium "MINE ACTION 2022"*, 31-35.

Iansiti, M., & Lakhani, K. R. (2017). The truth about blockchain. *Harvard business review*, *95*(1), 118-127.

Introducing NFT.Storage: Free Decentralized Storage for NFTs. (2021). Retrieved from https://filecoin.io/blog/posts/introducing-nft.storage-free-decentralized-storage-for-nfts/

Kalhoro, A., Wagan, A. A., Khan, A. A., Lin, J. M., Ku, C. S., Por, L. Y., & Yang, J. (2023). Rewarding Developers by Storing Applications on Non-Fungible Tokens. *Mathematics*, *11*(11), 2519.

Krtalić, A., & Bajić, M. (2019). Development of the TIRAMISU advanced intelligence decision support system. *European Journal of Remote Sensing*, *52*(1), 40-55.

Metaplex Docs, (2023). Retrieved from https://docs.metaplex.com

NFT.Storage Docs. (2021). Retrieved from https://nft.storage/docs/

Osmankovic, D., Akagic, A., Krivic, S., Uzunovic, T., & Velagic, J. Towards Safe and Explainable Humanitarian Demining with Deep Learning. In *Book of papers of 18th International Symposium "MINE ACTION 2022"*, 52-56.

Peterfay, E. (2022). Creating a website for a programmatically generated NFT collection: solana network, Häme University of Applied Sciences.

Psaras, Y., & Dias, D. (2020, June). The interplanetary file system and the filecoin network. In *2020 50th Annual IEEE-IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S)* (pp. 80-80). IEEE.

Vyzovitis, D., Napora, Y., McCormick, D., Dias, D., & Psaras, Y. (2020). Gossipsub: Attack-resilient message propagation in the filecoin and eth2. 0 networks. *arXiv preprint arXiv:2007.02754*.

Wang, Q., Li, R., Wang, Q., & Chen, S. (2021). Non-fungible token (NFT): Overview, evaluation, opportunities and challenges. *arXiv preprint arXiv:2105.07447*

Wang, W., Hoang, D. T., Hu, P., Xiong, Z., Niyato, D., Wang, P., ... & Kim, D. I. (2019). A survey on consensus mechanisms and mining strategy management in blockchain networks. *Ieee Access*, *7*, 22328-22370.

Wyatt, B. (2022). Improving document management with NFTs. Retrieved from https://blog.icme.io/improving-document-management-with-nfts/

Wyatt, B. (2022). Protecting creators with on-chain NFTs. Retrieved from https://blog.icme.io/protecting-creators-with-on-chain-nfts/

Localization of intruder UAV operator using Unmanned Aerial Vehicles

Ivan Magdalenić, Elvis Popović, Boris Tomaš, Neven Vrček, Lovro Posarić
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, Varaždin, Croatia
{ivan.magdalenic, elvpopovi, boris.tomas, neven.vrcek, lposaric}@foi.unizg.hr

Abstract. *The paper presents the results of an experiment where multiple Unmanned Aerial Vehicles (UAVs) were used to locate intruder UAV operator. UAVs were equipped with hardware of such dimension and weight that can be mounted on them. The hardware is then used to detect, identify, and track UAV signal to enable localization of intruder UAV operator. The measured signal strength is used to calculate distance between UAVs and intruder UAV operator to determine its exact location. Several problems need to be overcome to achieve that and they are addressed in the paper.*

Keywords. Localization, Unmanned Aerial Vehicles, UAV

1 Introduction

In many countries, there exist regions where the operation of Unmanned Aerial Vehicles (UAVs) is prohibited. These regions may include, but are not limited to, airspace above airports, military installations, and government buildings. Enforcing flight restrictions in these areas, particularly with respect to UAVs, presents a significant challenge. One major obstacle is the difficulty in identifying the location of the UAV operator, who may be concealed several hundred meters away. While it is possible to remotely disable a UAV, for instance by emitting a strong electromagnetic pulse that induces short circuits, this does not reveal the location of the operator. In this paper, we refer to a UAV engaged in unauthorized activities as an "intruder UAV", and the individual controlling it as the "intruder UAV operator".

Our proposed approach involves locating the intruder UAV operator by measuring the strength of their wireless control signal at multiple locations simultaneously and using this data to trilaterate their position. To accomplish this, we employ UAVs equipped with specialized hardware. This approach offers several advantages, including rapid adaptability to changing conditions on the ground and the ability to move our UAVs closer to the intruder UAV operator's location for more accurate positioning. However, there are also limitations to our approach, including the short battery life of our UAVs and hardware, as well as constraints on antenna size and processing power.

The remainder of this paper is organized as follows: Chapter 2 presents related work; Chapter 3 describes our proposed method for localizing intruder UAV operators using UAVs; Chapter 4 details our initial experiment; and the final section discusses our findings, conclusions, and directions for future work.

2 Related work

Localization of signal sources by measuring signal strength is not novel and it is used for military purposes long ago. Signal strength was used in (Kokić, 2017) to introduce secondary information source for enhanced GPS positioning, thus reducing GPS error. Signal strength was converted to distance between UAVs. Their GPS location was corrected to accommodate calculated distances of complex UAV constellations. In the terms of multipaths authors state that this is not an issue because UAVs were operating in clear line of sight, as this is usual UAV application.

In another research (Tomaš, 2013) author used similar technique to localize sources of the signal in urban environments. Sources of information were moving vehicles. Experiment conducted has managed to introduce less than 5m precision of localization of signal sources. This result is significant considering urban environments and multipathing.

To detect and identify intruder UAV operator we have to intercept communication between him and his UAV. (Jawhar, 2017) and (Hayat, 2016) provide an overview of the communications and architectures used in UAVs. It was described in the papers that a smaller number of UAVs use the pure IEEE 802.11 standard for communication, while most UAVs use their own modified version of the IEEE 802.11 standard. We single out (Bisio, 2018) and (Ezuma,

2020) that deal with the detection and classification of UAVs with regard to information available from captured communication between the UAV and the UAV-controlling system. One approach to locate UAV is described in detail in (Ezuma, 2020). The equipment used is 6 GHz bandwidth Keysight MSOS604A oscilloscope with a maximum sampling frequency of 20 GSa / s, 2 dBi omnidirectional antenna (for short-range detection), and 24 dBi Wi-Fi network antenna (for detection) at greater distances). Antennas are used in the 2.4 GHz spectrum.

What is common to those approaches is the usage of large equipment that is stationary or mounted on vehicles. Such equipment is not ideal for the hunt of small and fast UAVs. That's way we decided to take an alternative approach and use UAVs.

3 Localization of intruder UAV operator using UAVs

Our approach to localizing the intruder UAV operator is illustrated in Figure 1. The first step involves visually identifying the intruder UAV, after which several of our UAVs are dispatched to its vicinity to measure the strength of its radio signal, which should be the strongest in the area. The next step entails searching for a signal with the same pattern, or in our case, the same MAC address, as that of the intruder UAV operator. The final step involves calculating the position of the intruder UAV operator based on these measurements.

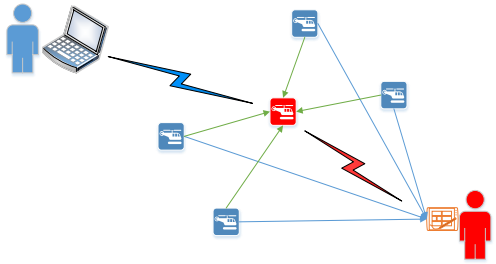


Figure 1. Localization of the intruder UAV operator

In Figure 2 is shown one of the UAVs with scanning equipment mounted and the intruder UAV.

One challenge we faced was developing hardware that was both lightweight and compact enough to be mounted on a UAV. Our initial setup consisted of a Raspberry Pi 4 Model B with 8 GB RAM, an Alfa AWUS036ACH wireless network adapter, a Ublox NEO M8U GPS module, and a HUAWEI E3372 LTE modem (as shown in Figure 2). However, after conducting several experiments, we determined that this hardware was not suitable for our approach due to two main issues: high power consumption, which significantly reduced the UAVs' flight time, and communication problems resulting from insufficient data transmission from the UAVs. We hypothesized that these issues were caused by mechanical vibrations affecting the components and wiring. As a result, we decided to switch to an integrated hardware solution (shown in Figure 3).

Figure 2. UAV with mounted hardware and the intruder UAV

The specification of the hardware are as follows: LoRa and WiFi packet capturing (3 devices): LILYGO TTGO T-Beam v1.1 ESP32 868 MHz, 0,96“ OLED screen, SSD 1306 driver, battery holder for Li-Ion 18650, MCU: ESP32, 8 MB PSRAM, 4 MB flash, LoRa: Semtech SX1276 14 dBm, NEO-6M (GPS) modul, WiFi/Bluetooth antenna 3D WiFi + IPEX. The transmission rate is 300 kb/s, and the radio signal strength sensitivity is up to -148 dBm.

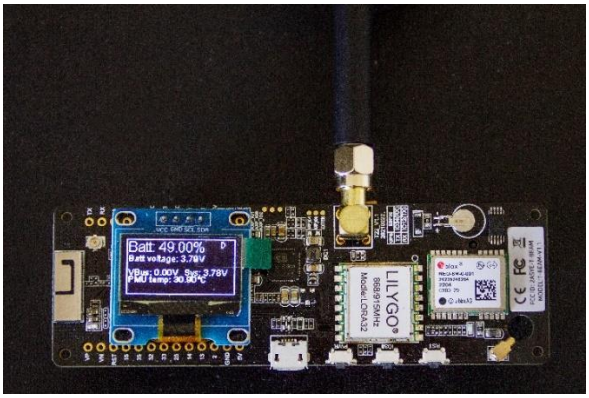


Figure 3. Hardware for intruder UAV operator detection and communication with home operator

In Figure 4 is presented a block diagram of the firmware used on each hardware mounted on UAVs. The firmware is developed that it captures all Wi-Fi packets on specific channel and every second it send the data with measured signal strength together with its present GPS coordinates to base station.

In Figure 5 is shown a block diagram of the firmware at the base station. The role of the base station is to collects data of all UAVs and to forward those data to server where all data are used to calculate the location of the intruder UAV operator. Our algorithm uses all collected data to determine with certain probability the location of the intrude UAV operator.

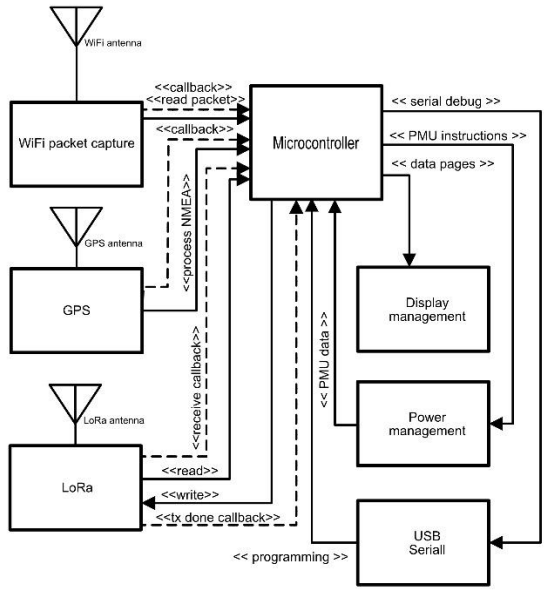


Figure 4. The block diagram of the firmware used on the hardware mounted on UAV

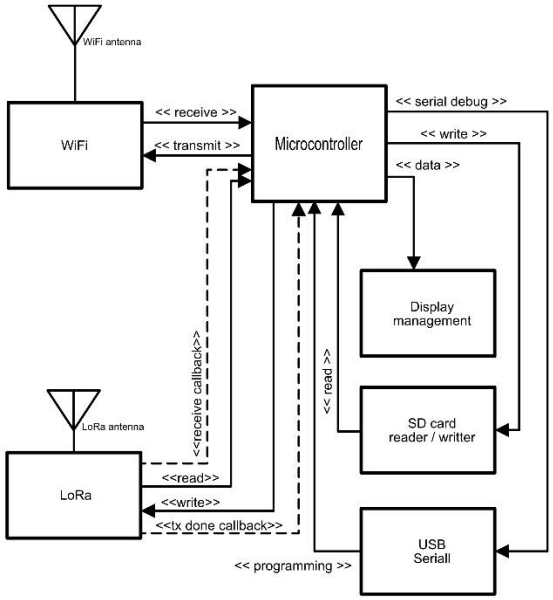


Figure 5. The block diagram of the firmware at the base station

We have divided our project into two phases. In the first phase, we focus exclusively on intruder UAVs that communicate using the standard 802.11 protocol. This phase is elaborated in this paper. In the second phase, we plan to employ machine learning techniques to identify the intruder UAV's signal pattern within the radio spectrum.

During the first phase, both the intruder UAV and its operator can be identified by their respective MAC addresses, which facilitates differentiation of their radio signals. Each of our UAVs can scan a different Wi-Fi channel to expedite the search for the intruder UAV's signal. Once this signal has been identified, all of our UAVs switch to listening on the same Wi-Fi channel and transmit their measurements to a central server.

One challenge we face is determining the optimal constellation of our UAVs to accelerate the search for the intruder UAV's radio signal. Once this signal has been detected and identified, another challenge arises in managing the constellation of our UAVs to move closer to the intruder UAV operator's location.

4 Experiment

The objective of this experiment was to evaluate the equipment and software depicted in Figures 3, in order to assess the viability of our methodology within a controlled environment with a known channel characteristic (center frequency of 2412MHz). The intruder UAV operator utilized an omnidirectional antenna with a constant output power of 20 dBm.

The intruder UAV operator actively controlled the intruder UAV from a fixed location, the GPS coordinates of which were determined at the onset of the experiment for subsequent distance calculations. The equipment and software were activated and gradually moved away from the intruder UAV operator in a linear formation.

The experiment was repeated multiple times, and the relationship between received signal strength and distance is illustrated in Figures 6, 7, and 8.

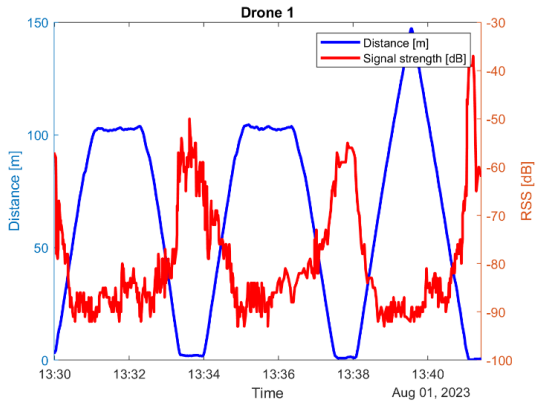


Figure 6. UAV 1 – signal strength/distance dependence

Data was collected at intervals of 1 s from all three UAVs and used to calculate the probability of the intruder UAV operator's location using the Friis transmission equation (Friis, 1946), with the frequency set to the midpoint of channel 1 (2412MHz) and the power component set to a constant value (20dBm) and disregarded due to multiple UAVs receiving the same signal.

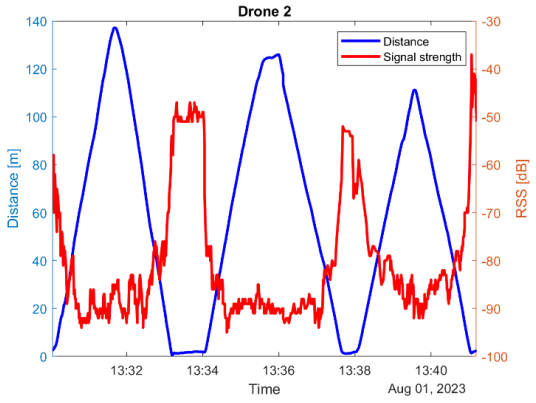


Figure 7. UAV 2 – signal strength/distance dependence

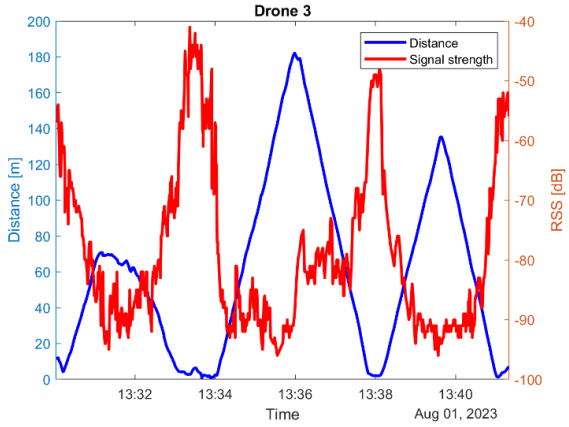


Figure 8. UAV 3 – signal strength/distance dependence

A tool and algorithm were developed to calculate the location of the intruder UAV operator, which computes the probability of location for selected points in space as depicted in Figure 9.

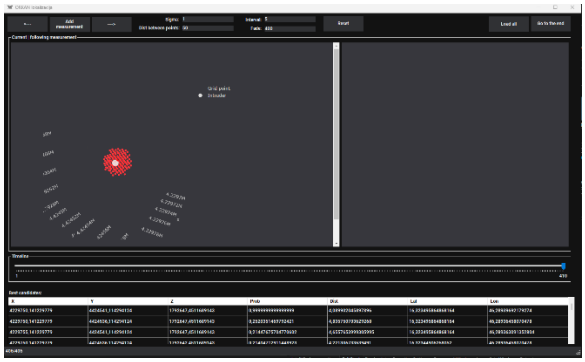


Figure 9. The tool for calculating the location of the intruder UAV operator.

The tool calculates the probability for all points within a defined region of space where the intruder UAV operator is expected to be located, as shown in Figure 10.

Probability for each is calculated based on location estimation using Friis transmission equation for every single measurement. Probability is multiplied from data originating for every UAV that is surveying airspace in real-time.

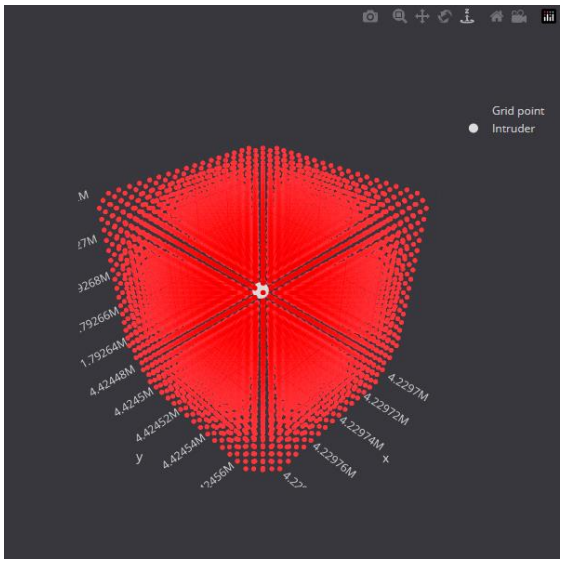


Figure 10. 3D space of localization points

Subsequently, a probability threshold was established so that only points with a probability above a certain value were displayed. Figure 9 displays points with a probability above 80% for the location of the intruder UAV operator. It should be noted that a white dot in the center of Figure 11 represents the actual location of the intruder UAV operator.

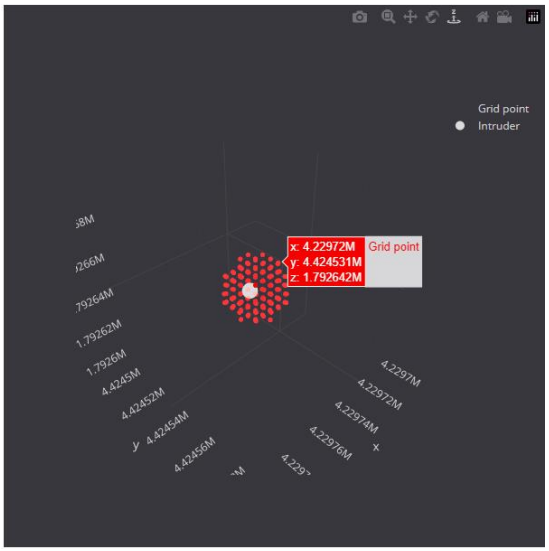


Figure 11. 3D visualization of the calculated probability of the location of the intruder UAV operator with threshold set 80%

Under ideal conditions, with measurements taken in open space with clear line-of-sight between our hardware and the radio signal source, our experiment achieved an accuracy within one meter for locating the intruder UAV operator, which is a highly satisfactory result.

5 Discussion

The utilization of multiple UAVs to locate the intruder UAV operator offers several advantages, including rapid target identification and tracking, as well as the flexibility to adapt to changing conditions within the area. However, this approach also presents several disadvantages, including limited flight time for the UAVs, the size and weight of the signal receiver, and processing power constrained by battery capacity. While more powerful receivers are available on the market, many are too large and heavy to be mounted on a small UAV.

The experiment demonstrated the feasibility of measuring distance using RF signal attenuation. However, challenges remain in terms of proper parameter setup and potential environmental influences on the radio signal. GPS error is another potential source of issues that will be addressed in future work. The effects of RF noise and GPS error may be mitigated by the open space of the experimental setup and multiple measurements that could theoretically minimize errors over time.

Similar to RF noise, different shapes and antenna power outputs could be mitigated by multiple measurements from various locations. Indeed, radiation patterns could be shaped like triangled torus, system could be capable of localizing the sources.

It should be noted that in real-world scenarios, signals may be partially obstructed by buildings, trees, and other obstacles. This introduces imprecision into the relationship between actual distance and that calculated using the Friis transmission equation. Even multipath scenarios might not pose significant challenge to constellation of rapidly moving aerial signal scanners as depicted in this research. Future work will focus on investigating such use cases and enhancing the algorithm capable of compensating for signal obstruction and multipaths.

There is also a clear limitation in terms of the maximum distance between the receiver and signal source imposed by the hardware used, which is approximately 250 meters. Additionally, in the initial phase of our project, only UAVs utilizing pure Wi-Fi protocol can be detected.

6 Conclusion

This paper presents the results of an infield initial experiment in which specialized hardware was utilized to measure the signal strength of an intruder UAV operator, with the aim of assessing the feasibility of using these measurements to determine their precise location. The results of the experiment demonstrate that the hardware mounted on the UAVs functioned as intended, exhibiting sufficient reception and processing capabilities to detect, identify, and track the radio signal emitted by the intruder UAV operator.

Future work on this project will focus on several key areas: i) real-time calculation of the intruder UAV operator's location; ii) investigation of the impact of environmental factors and GPS errors; iii) evaluation and benchmarking of the system against various radiation patterns emitted by intruder UAV controllers; and iv) determination of the optimal constellation and flight path of UAVs for efficient localization of the intruder UAV operator.

Acknowledgments

This work has been fully supported by Croatian Science Foundation under the project IP-2019-04-4864.

References

Kokić P., Tomaš B. (2017) Enhanced drone swarm localization using GPS and trilateration based on RF propagation model. *Central European Conference on Information and Intelligent Systems.* Varaždin, 2017. pp. 259-264

Tomaš B. (2013) Wifi roaming access point optimum assignment in urban multi-sensor. *Research papers Faculty of Materials Science and Technology Slovak University of Technology in Trnava*, pp 109-115

Jawhar I. et al. (2017) Communication and Networking of UAV-Based Systems: Classification and Associated Architectures. *Journal of Network and Computer Applications*, February, DOI: 10.1016/j.jnca.2017.02.008

Hayat S. et al. (2016) Survey on Unmanned Aerial Vehicle Networks for Civil Applications: A Communications Viewpoint. *IEEE Communications Surveys & Tutorials*, Volume: 18 , Issue: 4 , Fourthquarter, DOI: 10.1109/COMST.2016.2560343

Bisio I., Garibotto C., Lavagetto F., Sciarrone A., and Zappatore S. (2018) Unauthorized amateur UAV detection based on WiFi statistical fingerprint analysis. *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 106–111

Ezuma M., Erden F., Kumar Anjinappa C., Ozdemir O. and Guvenc I. (2020) Detection and Classification of UAVs Using RF Fingerprints in the Presence of Wi-Fi and Bluetooth Interference. *IEEE Open Journal of the Communications Society*, vol. 1, pp. 60-76, doi: 10.1109/OJCOMS.2019.2955889

Friis H. T. (1946) A Note on a Simple Transmission Formula. *Proceedings of the IRE*, vol. 34, no. 5, pp. 254-256, doi: 10.1109/JRPROC.1946.234568

Comparison of Machine Learning Algorithms on Noisy Data

Dijana Oreski, Dunja Visnjic, Nikola Kadoic
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, Varazdin
{dijana.oreski, dunja.visnjic, nikola.kadoic}@foi.unizg.hr

Abstract*. Having noisy data in datasets is not a rare situation. Noisiness in data can influence the results and accuracy of machine learning algorithms applications. This paper focuses on the question of which machine learning algorithm will perform the best when identifying noisy data. Answering this question brings us one step closer to a meta-learning recommendation system, the main goal of project SIMON. To answer this question, experiments are conducted on the publicly available datasets Bot-IoT, which consists of real and simulated IoT network traffic. Several approaches to implementing behavioral IoT botnet attack detection have been explored in the literature, including machine learning. In this paper, we are exploring four different types of machine learning algorithms: (i) machine learning algorithms based on error, (ii) machine learning algorithms based on information, (iii) machine learning algorithms based on similarity, and (iv) machine learning algorithms based on probability. The results show that the highest accuracy is achieved by a machine learning algorithm based on error – an artificial neural network that achieved the highest accuracy. However, even though the decision tree achieved slightly lower results, there is no statistically significant difference between the artificial neural network and the decision tree.*

Keywords. Machine learning algorithms, noisy data, meta-learning, network forensics, IoT devices

1 Introduction

This paper is prepared under the scope of project SIMON: Intelligent system for automatic selection of machine learning algorithms in social sciences. The main goal of the project SIMON is to develop an intelligent system for the automatic selection of machine learning algorithms in the social sciences that perform better on a given data set, taking into account the specific characteristics of the data. The research on the project involves a comparative analysis of a large number of machine learning algorithms on a large number of datasets.

In the last few years, there is a significant increase in the usage of meta-learning to enable the selection of the best machine learning algorithm for a given data set. Meta-features are used to explain the properties of the datasets and the performance of machine learning algorithms. One of the meta-features is noise. Noise is a measurement disparity between a setting that is empirical and what a dataset claims about it. Empirical data may be noisy for a variety of reasons, such as a lack of accuracy in the dataset's construction or the addition by mistake of extra attributes to some items, or the omission of certain objects while characterizing the scope of an attribute. Machine learning algorithms' ability to predict outcomes can suffer with the presence of noise in data sets. Bot-IoT data is characterized by noise.

In recent years, the wide adoption of the modern Internet of Things (IoT) paradigm has led to the development of fog computing, which improves the collection and processing of enormous amounts of data, when cloud computing features, such as networking, data storage, administration, and analytics, are placed very close to the edge of networks. IoT development has also led to an increase in security challenges. Many IoT devices have various security flaws in their implementation and design, making them a prime target for botnet attacks. One of the best options is intrusion detection systems, especially those created with artificial intelligence.

The billions of physical items that IoT consists of, can connect with each other, including only a little human involvement. IoT has developed into one of the most widespread technologies and a fascinating area of study in the business and research sector. IoT is becoming increasingly popular and in demand. Numerous organizations are providing funding in this area for their purposes and as a service to other organizations. Although the IoT is expanding rapidly due to technological advancements, the proliferation of IoT devices, and the activation of services, there are significant security threats and financial harm caused by the actions of numerous botnets. A botnet is a group of software robots, or bots, that operate automatically

and autonomously. Botnets function on networks of zombie computers under the remote control of attackers. IoT has been impacted and infected by intelligent botnet activity, including distributed denial of service (DDoS) attacks, spamming, and phishing. There hasn't been a network forensics technique that can perfectly classify, detect, and trace the activities of sophisticated botnets to date, even though botnets with such attacks have posed a severe security concern to the Internet infrastructure for years. Consequently, it's critical to correctly identify these botnets' actions and there is a strong need to create new approaches for detecting attacks, predicting attacks, and preventing attacks.

Machine learning algorithms can be applied to protect data from cyber security threats. Machine learning algorithms are used in a variety of ways to stop and spot network outbreaks and security holes. In recent years, numerous studies have trained and validated predictive models developed employing machine learning algorithms to define these botnet attacks. Artificial neural networks are among the machine learning algorithms which were mostly used so far across various research papers.

Prior research focused primarily on achieving the highest level of accuracy in separating legitimate from malicious IoT communications, with little attention paid to identifying the specific sort of attack that was being performed, or investigating data characteristics, neither comparing various machine learning approaches nor investigating their applicability for IoT dataset and specifies of such data.

Bot-IoT data is characterized by noise. Recently, a few research papers proposed pre-processing techniques which support data noise cleaning on IoT data. Jane and Arockiam (Jane & Arockiam, 2021) proposed a technique called Detection and Removal of Noise (DaRoN) that removes “the null values, error values, repeated values, incomplete values, and irrelevant values”. By doing that, the authors removed noisy IoT data. Bobulski and Kubanek (Bobulski & Kubanek, 2022) discussed the need for automatization of data cleaning and suggested a data cleaning method.

Although such initiatives provided a promising avenue to deal with noise in data, cleaning approaches imply deleting data. Part of the data disappears during the cleaning phase, which results in the loss of information from the initial dataset. This is especially important when dealing with small datasets, so the decision maker/data analyst may decide not to clean the data so that potentially important information would not be lost.

In this paper we did not perform data cleaning, instead, we are employing four machine learning-based predictive models on raw data to preserve information and discover which machine learning approaches work best on such data.

An aim is to identify IoT botnet assaults that not only aid in separating legitimate traffic from malicious traffic but also identify the specific IoT botnet attack type. To accomplish this, four different types of machine learning algorithms were applied to publicly available IoT datasets. Such a dataset is shown to be noisy, with numerous erroneous attribute values and missing values. From a data mining point of view, noisy data requires careful examination and preparation before the development of predictive models. By exploring and analyzing that dataset, we are striving to answer the question: which of the four different approaches to machine learning gives the most accurate predictive models on noisy IoT data?

The rest of the paper is organized as follows. In Section 2, a related literature review is provided. Section 3 explains the data and research methodology and gives research results focusing on predictive model accuracy. Finally, Section 4 concludes the article and provides guidelines for future research activities.

2 Literature review

Finding the intelligent intrusion detection system in IoT-based environments for many sorts of applications has been the subject of a vast body of work and study so far in the literature. Hereinafter, are listed some of them.

Bagui, Stevens and Bagui (Bagui et al., 2021) present a thorough examination of the benchmark dataset NSL-KDD to create a powerful network-based intrusion detection system. The uniqueness of this study lies in the discovery of the bare minimum set of information necessary for the automated classification of each assault as well as each attack type in the NSL-KDD dataset. There hasn't yet been any analysis done on the specific attack's level. Following the usage of Information Gain for feature selection are machine learning techniques like J48 Decision Tree and Naive Bayes. With each method, high classification accuracy is attained.

Hyun (Hyun, 2021) conducted a comparative study on how well some machine learning algorithms perform at spotting botnet activities. k-nearest neighbours (k-NN) is the most effective and efficient machine learning method for DDoS, DoS, and reconnaissance attack detection, according to experimental findings utilizing the Bot-IoT dataset.

Malathi and Padmaja (Malathi & Padmaja, 2023) performed research to provide security using a variety of machine learning methods, which are primarily intended to detect an attack on an interconnected (IoT) network right away. Different recognition algorithms are estimated using specific metadata, or Bot-IoT. Several different machine learning algorithm types were handled during this execution step, and the majority of them achieved outstanding results. Metadata from the Bot-IoT was used to collect novel factors, but implementation and the new features created produced more trustworthy results.

Wiyono and Cahyani (Wiyono & Cahyani, 2020) were motivated to develop a novel classification

algorithm for network forensics that could monitor suspected botnet activity in the compromised network and is based on the identification of network traffic. Based on the performance study and experimental findings, the authors concluded that decision tree C4.5 algorithm and network flow identification combined with feature selection and proper classification approaches are sufficient to identify and classify attacks and help track botnet activity in the IoT environment.

Motylinski et.al. (Motylinski et al., 2022) presented a method for classifying the different attack types that were included in the IoT-Bot dataset and pre-processing phase. The authors compared the results obtained with the GPU-accelerated versions of the cuML library's Random Forest, k-NN, Support Vector Machine (SVM), and Logistic Regression classifiers, as well as the pre-processing steps taken to prepare the data for training. The best-trained models achieved 0.99 scores for accuracy, precision, recall, and f1-score, by using their approach. Additionally, the training and estimation times were greatly shortened by using feature selection and training models on GPU.

Alothman, Alkasassbeh, and Al-Haj (Alothman et al., 2020) tested various classifiers, and the results from the best three: J48, Random Forest (RF), and Multilayer Perceptron (MLP) networks were explained. The outcomes demonstrated the superiority of the RF and J48 classifiers over MLP networks and other cutting-edge technologies. The best binary classifier revealed in this study had an accuracy of 0.99, while the best classifications of main attacks and subcategories had accuracy values of 0.96 and 0.93, respectively. In this study, authors evaluated results also in terms of False Negative (FN) rates. This time around, J48 and RF classifiers performed better than the MLP network classifier and were able to classify subcategories with a maximum micro FN rate of 0.076.

A study by Bagui et.al (Bagui et al., 2019) classifies cyberattacks in the UNSW-NB15 dataset using a hybrid feature selection process and classification techniques. In order to identify the best subset of features, k-means clustering, and correlation-based feature selection were combined. After feature selection, two classification methods - one probabilistic, Naive Bayes (NB), and the other based on decision trees (J48), were used in the modeling phase. According to their findings, the NB model in combination with a hybrid feature selection approach was able to increase the classification accuracy of the majority of attacks, particularly the unusual attacks. With this feature selection and NB model combination, the false alarm rates were lower for the majority of attacks, especially the unusual attacks. Although the J48 decision tree model's classification rate for all attack families was already quite high, with or without feature selection, it did not perform any better.

Naaz (Naaz, 2021) used the IoT dataset to test the effectiveness of the machine learning algorithms random forest classifier, support vector machine, and logistic regression for the detection of phishing attacks. The results were then compared to earlier studies that had used the same dataset. Based on the accuracy, error rate, precision, and recall, the outputs of these algorithms have then been compared.

Yudhana, Riadi, and Ridho (Yudhana et al., 2018) performed the classification of DDoS attacks by utilizing naive Bayes and neural networks to analyze network traffic. Based on the results, it was discovered that naive Bayes had a 99.9% accuracy rate while artificial neural networks had a 95.23% accuracy rate. The conclusions of the experiments demonstrate that the naive Bayes approach outperforms the neural network. The authors conclude that their analysis along with experimental results can be utilized as proof in the trial process.

Alrashdi et.al. (Alrashdi et al., 2019) propose an Anomaly Detection-IoT (AD-IoT) system, which is an intelligent anomaly detection based on a Random Forest machine learning algorithm, to handle IoT cybersecurity concerns in a smart city. At dispersed fog nodes, the suggested approach may successfully detect hacked IoT devices. The authors used a current dataset to demonstrate the model's accuracy and evaluate it. Their research demonstrates that the AD-IoT is capable of achieving the maximum classification accuracy of 99.34% with the lowest false positive rate.

Almiani et.al. (Almiani et al., 2020) provided a fully automated, artificially created intrusion detection system against cyberattacks. The suggested model makes use of multi-layered recurrent neural networks that are intended to be used for security in fog computing, which is implemented very close to end users and IoT devices. Using a balanced version of the hard dataset, NSL-KDD, the authors demonstrated their suggested model. The experimental findings and computer simulations indicated the stability and robustness of the suggested model in terms of a range of performance measures.

Lutta, Sedky, and Hassan (Lutta et al., 2021) recently carried out a Systematic Literature Review (SLR) of the most recent IoT forensics research developments. One of their guidelines for future research indicates the need for machine learning algorithms implemented in this domain.

Numerous fields deals with the problem of data with noise and IoT forensics is no exception. There are two main types of noise in data: class noise and attribute noise. Attribute noise is considered to be less harmful to modeling than class noise. Robust machine learning algorithms have proven to be a good approach to dealing with imperfect data. According to the literature review, those algorithms are k-NN (Liu & Zhang, 2012; Moosavi et al., 2010), artificial neural networks (Folleco et al., 2009; Miranda et al., 2009), classification, and regression trees (Folleco et al., 2009; Khoshgoftaar et al., 2011; Miranda et al., 2009).

3 Empirical analysis

The used dataset is described in the first section of this chapter, and the methodologies for data analysis are explained in the second section. The third section provides insights into research results.

3.1 Data description

Data used in this research was created by Koroniotis, Moustafa, Sitnikova, and Turnbull (Koroniotis et al., 2019). The authors created a new Bot-IoT dataset using realistic IoT networks. In (Koroniotis et al., 2019), a full description of simulating IoT sensors is provided along with data understanding. There are 3 668 522 cases in the dataset. It should be mentioned that it is the reduced dataset since the original one contains 72 000 000 instances and its analysis was extremely time and computationally demanding. That is the reason why just 5% of the original dataset has been extracted. In our study, we also used such a reduced dataset. 46 features are present in the dataset for each instance. Some of these features are characterized as noisy. Thus, hereinafter we are dealing with attribute noise.

Since Koroniotis et al. (Koroniotis et al., 2019) performed feature selection and extracted only 10 features from the original dataset, 10 features are also chosen in this research. The ten best features which were determined by Koroniotis et al. (Koroniotis et al., 2019) are as follows: rate, drate, srate, state number, max, mean, min, stddev, gs number, seq. The ten best features from our research are in Table 1, along with noise level measured as amount of irrelevant information.

Table 1. Feature description

Feature	Feature Explanation	Noise level
drate	Destination-to-source packets per second	0.65
Flags number	Numerical representation of feature flags	0.54
max	Maximum duration of aggregated records	0.32
mean	Average duration of aggregated records	0.29
min	Minimum duration of aggregated records	0.34
N IN Conn P DstIP	Number of inbound connections per destination IP.	0.41
N IN Conn P SrcIP	Number of inbound connections per source IP.	0.11
seq	seq Argus sequence number	0.71
State number	Numerical representation of feature state	0.26
stddev	The standard deviation of aggregated records	0.74
attack	Class label: 0 for Normal traffic, 1 for Attack Traffic	0.44

For feature selection in this study, we used a contrast set mining based on the STUCCO algorithm: SfFS (STUCCO for Feature Selection). In our earlier research papers, SfFS application in feature selection produced the best results (Oreski & Androcec, 2018; Oreški & Andročec, 2020). The concept of SfFS is initially introduced in (Oreški & Kliček, 2015). The decision on the implementation of SfFS for feature selection is due to the noisy character of the used data. Missing values along with the outliers bring noise to the data. SfFS has shown to be a good approach for dealing with noisy data (Oreški & Kliček, 2015).

3.2 Methods overview

Koroniotis et al. (Koroniotis et al., 2019) evaluated the performance of network forensic methods by applying three machine learning algorithms. The models that were trained were: Support Vector Machine (SVM), Recurrent Neural Network (RNN), and Long-Short Term Memory Recurrent Neural Network (LSTM-RNN). All those algorithms present similar approaches to the development of predictive models. In this paper, we are expanding the number of different approaches by investigating the effectiveness of four different approaches to learning. Thus, we are comparing algorithms belonging to four different groups of machine learning algorithms: (i) machine learning algorithms based on error, (ii) machine learning algorithms based on information, (iii) machine learning algorithms based on similarity, (iv) machine learning algorithms based on probability. An artificial neural network is used as a machine learning algorithm based on error, the decision tree is used as a machine learning algorithm based on information, the k-NN is used as a machine learning algorithm based on similarity, and Naïve Bayes classifier is used as machine learning algorithm based on probability.

The artificial neural network algorithm develops a predictive model by correcting the weights of links between neurons to reduce the error of the model. The decision tree algorithm seeks the most informative features to develop a predictive model based on such features. k-NN seeks similar instances and classifies new instances in the same class as the nearest feature. The Naïve Bayes classifier is based on the Bayesian theorem of conditional probability.

After the four algorithms were applied to the dataset, several two-matched sampling t-tests were used to compare the algorithms to answer the following research questions:

1. Is there a statistically significant difference in accuracy between artificial neural network (ANN) and decision tree (DT)?

2. Is there a statistically significant difference in accuracy between artificial neural network (ANN) and k-NN?
3. Is there a statistically significant difference in accuracy between artificial neural network (ANN) and Naïve Bayes (NB)?
4. Is there a statistically significant difference in accuracy between decision tree (DT) and k-NN?
5. Is there a statistically significant difference in accuracy between decision tree (DT) and Naïve Bayes (NB)?
6. Is there a statistically significant difference in accuracy between k-NN and Naïve Bayes (NB)?

4 Research results

In data analysis, we initially carried out feature selection and identified relevant features for separating the attack and normal classes (ten features enlisted in Table 1). The proposed methodology's classification accuracy is then evaluated.

Predictive model accuracy can be tested and assessed in a variety of ways. The *k*-fold cross-validation is employed here. The data set is split into *k* subsets using *k*-fold cross-validation. The test set is always one of the *k* subsets, whereas the training set is always the other *k-1* subsets. 10 folds are used in this research.

Table 2 presents the results of four machine algorithms in terms of accuracy.

Table 2. Models accuracy

Algorithm	Accuracy
Artificial neural networks	99,83
Decision tree	99,24
k-NN	98,65
Naïve Bayes classifier	96,33

We may infer several conclusions by using the performance measure from Table 3. Question to be asked, is it possible to generalize the outcomes or are they the result of chance? The goal of statistical significance testing is to determine how well evaluation measures reflect classifier behavior. Two matched sampling t-tests were used to compare the algorithms we tested on one domain. The significance of the mean difference is investigated at the significance level of 0.05. To test whether we can reject the null hypothesis, the assumption in Table 3 is that there is no difference between the mean values of algorithms performances.

The presumptions of the t-test were met.

Table 3. T-test

Hypothesis	Model	T-test
H0: ANN = DT	ANN	0.07
	DT	
H0: ANN = k-NN	ANN	0.04
	k-NN	
H0: ANN = NB	ANN	0.02
	NB	
H0: DT = k-NN	DT	0.06
	k-NN	
H0: DT = NB	DT	0.02
	NB	
H0: k-NN =NB	k-NN	0.02
	NB	

As seen in Table 3. there are statistically significant differences in the performances of artificial neural networks and k-NN and NB. However, differences in performances between ANN and DT are not statistically significant. DT performs statistically significantly better than NB, but not when compared with k-NN. In the end, k-NN performs better than NB and the difference in performances between these machine learning algorithms is statistically significant.

5 Conclusion

In this paper, we have proposed a machine learning-based intrusion detection predictive model for IoT network security. The proposed model adopts four different approaches to machine learning-based development of predictive models. The results of the performance evaluation reveal the effectiveness of the artificial neural networks approach compared to the decision tree, k-NN, and Naïve Bayes classifier.

This paper gives three scientific contributions: i) in the field of machine learning, by investigating how different machine learning approaches handle noisy IoT data, (i) in network forensics, by comparing different machine learning approaches and demonstrating which one achieves the best predictive model in this domain, and (iii) in the field of IoT field, because the models are more accurate than the previous models applied to the same dataset.

There are several limitations of the research presented here. First, only one dataset is used in algorithms comparison. In future research, we will upgrade several datasets. Second, the dataset was highly unbalanced, which could lead to bias in the learner's favor of the attack class. This prejudice arises from the fact that the attack class is much overrepresented in comparison to the normal class. We

will employ several strategies to address the class imbalance in future research. Some of them include comparative analysis of different class imbalanced strategies combined with machine learning algorithms on several datasets to determine if there is a connection between certain machine learning algorithms' performance in terms of accuracy and the level of imbalance and noisiness in data. Third, we will perform data cleaning in the pre-processing phase and compare the results of our models gained on raw data with the results of cleaned data.

Acknowledgments

This paper is supported by Croatian science foundation under the project SIMON: Intelligent system for automatic selection of machine learning algorithms in social sciences, UIP-2020-02-6312.

References

Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., & Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory*, *101*, 102031. https://doi.org/10.1016/j.simpat.2019.102031

Alothman, Z., Alkasassbeh, M., & Al-Haj Baddar, S. (2020). An efficient approach to detect IoT botnet attacks using machine learning. *Journal of High Speed Networks*, *26*(3), 241–254. https://doi.org/10.3233/JHS-200641

Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M., & Ming, H. (2019). AD-IoT: Anomaly detection of IoT cyberattacks in smart city using machine learning. *2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019*, *January*, 305–310. https://doi.org/10.1109/CCWC.2019.8666450

Bagui, S., Kalaimannan, E., Bagui, S., Nandi, D., & Pinto, A. (2019). Using machine learning techniques to identify rare cyber-attacks on the UNSW-NB15 dataset. *Security and Privacy*, *2*(6). https://doi.org/10.1002/spy2.91

Bagui, S., Stevens, J., & Bagui, S. (2021). Analyzing and Classifying Network Attacks Using Machine Learning on the NSL-KDD Dataset. *The Journal of Computing and Technology*, *2*(1), 24–36.

Bobulski, J., & Kubanek, M. (2022). A method of cleaning data from IoT devices in Big data systems. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, *February 2023*, 6596–6598. https://doi.org/10.1109/BigData55660.2022.10020651

Folleco, A. A., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2009). Identifying learners robust to low quality data. *Informatica (Ljubljana)*, *33*(3), 245–259.

Hyun, M.-J. (2021). Hyun, M. (2021). A comparative study of the performance of machine learning algorithms to detect malicious traffic in IoT networks. *Journal of Digital Convergence*, *19*(9), 463–468. https://doi.org/10.14400/JDC.2021.19.9.463

Jane, V. A., & Arockiam, L. (2021). Daron: A technique for detection and removal of noise in IoT data by using central tendency. *Annals of the Romanian Society for Cell Biology*, *25*(2), 3197–3203.

Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *41*(3), 552–568. https://doi.org/10.1109/TSMCA.2010.2084081

Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. *Future Generation Computer Systems*, *100*(November), 779–796. https://doi.org/10.1016/j.future.2019.05.041

Liu, H., & Zhang, S. (2012). Noisy data elimination using mutual k-nearest neighbor for classification mining. *Journal of Systems and Software*, *85*(5), 1067–1074. https://doi.org/10.1016/j.jss.2011.12.019

Lutta, P., Sedky, M., Hassan, M., Jayawickrama, U., & Bakhtiari Bastaki, B. (2021). The complexity of internet of things forensics: A state-of-the-art review. *Forensic Science International: Digital Investigation*, *38*, 301210. https://doi.org/10.1016/j.fsidi.2021.301210

Malathi, C., & Padmaja, I. N. (2023). Identification of cyber attacks using machine learning in smart IoT networks. *Materials Today: Proceedings*, *80*(xxxx), 2518–2523. https://doi.org/10.1016/j.matpr.2021.06.400

Miranda, A. L. B., Garcia, L. P. F., de Carvalho, A. C. P. L. F., & Lorena, A. C. (2009). Use of Classification Algorithms in Noise Detection and Elimination. *4th International Conference, HAIS*, 417–424.

Moosavi, M. R., Fazaeli Javan, M., Zolghadri Jahromi, M., & Sadreddini, M. H. (2010). An adaptive nearest neighbor classifier for noisy environments. *Proceedings - 2010 18th Iranian Conference on Electrical Engineering, ICEE 2010*, 576–580. https://doi.org/10.1109/IRANIANCEE.2010.5507

005

Motylinski, M., MacDermott, Á., Iqbal, F., & Shah, B. (2022). A GPU-based machine learning approach for detection of botnet attacks. *Computers and Security*, *123*(September), 102918. https://doi.org/10.1016/j.cose.2022.102918

Naaz, S. (2021). Detection of phishing in internet of things using machine learning approach. *International Journal of Digital Crime and Forensics*, *13*(2), 1–15. https://doi.org/10.4018/IJDCF.2021030101

Oreski, D., & Androcec, D. (2018). Hybrid Data Mining Approaches for Intrusion Detection in the Internet of Things. *Proceedings of International Conference on Smart Systems and Technologies 2018, SST 2018*, 221–226. https://doi.org/10.1109/SST.2018.8564573

Oreški, D., & Andročec, D. (2020). Genetic algorithm and artificial neural network for network forensic analytics. *43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1200–1205.

Oreški, D., & Kliček, B. (2015). A novel feature selection techniques based on contrast set mining. *14th International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'15)*.

Wiyono, R. T., & Cahyani, N. D. W. (2020). Performance Analysis of Decision Tree C4.5 as a Classification Technique to Conduct Network Forensics for Botnet Activities in Internet of Things. *2020 International Conference on Data Science and Its Applications, ICoDSA 2020*, 1–5. https://doi.org/10.1109/ICoDSA50139.2020.9212932

Yudhana, A., Riadi, I., & Ridho, F. (2018). DDoS classification using neural network and naïve bayes methods for network forensics. *International Journal of Advanced Computer Science and Applications*, *9*(11), 177–183. https://doi.org/10.14569/ijacsa.2018.091125

YOLOv7 Model for Small Object Handling in Maritime Images

Miran Pobar
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, 51000 Rijeka, Croatia
mpobar@uniri.hr

Abstract. *Around the world many maritime surveillance cameras exist whose functionality could be expanded with computer vision-based object detection in order to monitor traffic and provide automated statistics or to increase safety. To this aim, we train two versions of the Yolov7 object detection model on a suitable custom dataset with four object categories and evaluate their detection performance. In order to handle small objects such as buoys or objects that appear visually small in the frame such as distant boats, we examine two different configurations of input to the model.*

Keywords. computer vision, neural networks, small objects, ship detection

1 Introduction

Object detection in images is a fundamental step in many realistic applications of computer vision, e.g. surveillance, autonomous driving, industrial automation etc. Although most recent research in object detection uses large datasets of common objects such as MS COCO (Lin et al. 2014) for both training and performance benchmarking, specialized tasks require training on task-specific datasets that include object categories of interest and in realistic application context. In this paper, the focus is on detection of maritime vessels and small objects in images from shore-mounted surveillance and panoramic cameras, with possible applications for increasing the safety by detecting objects that may be difficult to detect via other means such as radar, or monitoring traffic for vessels that are not equipped with the automatic identification system (AIS). Maritime surveillance cameras already exist in many places, and a whose functionality could be better utilized by automated object detection.

For object detection, we use transfer learning from an established convolutional neural network-based model to adapt the original model to the domain of interest.

To train and evaluate the model, an original dataset with annotated images of maritime scenes is used. The size of the dataset is quite small but is designed to cover a diverse set of locations and maritime scenes (e.g. large commercial port, marina, canal, small fishing boat port, shore near a tourist destination) and to enable transfer learning from models pretrained on large generic datasets.

Although great progress has been made in recent years with neural network-based methods for detecting objects in images, detection of small or tiny objects has remained challenging (Tong & Wu, 2022), especially on busy backgrounds or when objects, in addition to being small, have similar colour as the background.

Here, to deal with smallest objects such as buoys that may appear only a few pixels large in the video frame, a simple input slicing approach is evaluated. If the input image is larger than the size that the model expects, the input image is sliced into rectangular segments of the expected size and detection is performed on each slice. In this way, rescaling is avoided, which could obliterate smallest objects and make detection impossible.

The contributions of this paper include the detection performance comparison of two YOLOv7 model variants for the task of object detection on a novel maritime image dataset, and the comparison of how two image input strategies, namely rescaling or slicing affect the detection performance, especially of small objects.

The rest of the paper is organized as follows: in Section 2, related work is introduced, Section 3 describes the proposed dataset and models for maritime object detection with details of experimental setup. Results are presented and discussed in Section 4. Finally, a conclusion and suggestion for further work are given in Section 5.

2 Related Work

The majority of recent object detection methods such as SSD (Liu et al., 2016), Yolo family of detectors

(Redmon et al., 2016, Redmon & Farhadi, 2018, Bochkovskiy, Wang & Liao, 2020, Bochkovskiy & Liao, 2022) and EfficientDet (Tan, Pang & Le, 2020) are focused on detection of objects in single images and are based on convolutional neural networks (CNNs).

While universal object detectors trained on general object datasets perform well for many tasks, new application domains often require modifications and specialized datasets to adapt the models to the new tasks, such as detection in thermal imaging (Krišto, Ivašić-Kos & Pobar, 2020), or when dealing with small objects, such as detecting bees (Stojnić et al., 2021) or persons (Sambolek & Ivasic-Kos, 2021) in UAV videos.

Detection of small objects is an active topic, and some research is directed at more task-specific methods, for example (Stojnić et al., 2021), while some aim for more general solutions with modification of existing architectures such as YOLO-Z (Benjumea et al. 2023). Although in many practical applications of object detection, the true input source are not individual independent images but frames of video, relatively few methods focus on exploiting the information from multiple frames to improve object detection. For example, in (Broad, Jones & Lee, 2018) an architecture of single-image CNN-based object detector is modified with a recurrent layer so that data from multiple frames is fused on the level of extracted features and further processed in the network.

In this paper a simple input image slicing strategy is explored to avoid reducing the image size and preserve small objects in the input images, similar to (Unel, Ozkalayci & Cigla, 2019)

Datasets for detection of small object in the sea such as SeaDroneSee (Varga et al., 2022) have appeared recently. The SeaDroneSee is tailored for the search-and-rescue scenario and thus focuses on the detection of persons in the water. Reflecting the application, the images were collected using unmanned aerial vehicles with video quality being one of prime concerns.

The SPSCD dataset proposed in (Petković et al. 2023) contains images from a stationary camera in a port and is focused on detection of several classes of ships that may enter the port. It covers different times of day and weather conditions but doesn't specifically address small objects and is limited to one port.

The dataset proposed in this paper targets similar use as the SPSCD dataset but also focuses on small objects in the maritime domain, and aims to cover more variety of locations, however on a smaller scale and with fewer object classes. Since the source videos are broadcast streams from coastal cameras the video quality is highly compressed, which makes detection of small objects even more challenging.

3 Methods and Data

In this experiment, two different Yolov7-based models (Wang, Bochkovskiy & Liao, 2023) for detection of maritime objects in images were trained and evaluated on a custom dataset. Yolov7 is an iteration of the Yolo (Redmon et al., 2016) family of convolutional neural network-based object detectors that predict the bounding boxes of objects in images and their corresponding class labels in a single pass through the network. On the MS COCO object detection dataset, it achieves excellent results in terms of both speed and accuracy (Wang, Bochkovskiy & Liao, 2023). It has also recently been applied to the problem of detection of maritime objects (Zhu et al., 2023), and the authors note the mean average precision (MAP) of 75.66% at 50% intersection-over-union (IoU) threshold on a dataset for maritime search-and-rescue task (Gasienica-Jozkowy, Knapik & Cyganek, 2021). The authors proposed a modified YOLOv7 architecture, and achieved an increase small object detection mAP score of 4.2%, and an improvement of detection accuracy from 0.578 to 0.754 for the *boat* class specifically and from 0.615 to 0.643 for the buoy class in one network variant.

3.1 Dataset

The object detection dataset was prepared from 20 video streams available on the internet streaming from stationary cameras from various ports, canals, marinas, and other points of interest (Fig. 1), covering a diverse range of scenes and weather conditions. Broadly grouped, 10 streams were from small ports or marinas, where usually many smaller boats and buoys appear simultaneously in frame, with lots of occlusion due to boats being close together. Five streams were of general vistas of tourist interest, where usually fewer boats appear simultaneously but most move, and five streams were from large ports or shipping canals where normally large commercial ships appear. Most videos were streamed in 1920x1080 resolution, and a minority in 1280x720, with varying degrees of quality and compression strength. Some of the cameras were completely stationary, while some used panning or zooming motion to cover a larger area. The videos were gathered from April to May 2022.

From the videos, individual frames were extracted at time points where the scene subjectively appeared significantly changed, and saved in .jpg format. In some cases, e.g. when the camera pans or zooms, the interval between extracted images may be just a few seconds, while in some cases, e.g. in calm weather and in a port that is not busy, the interval may be several hours or days.

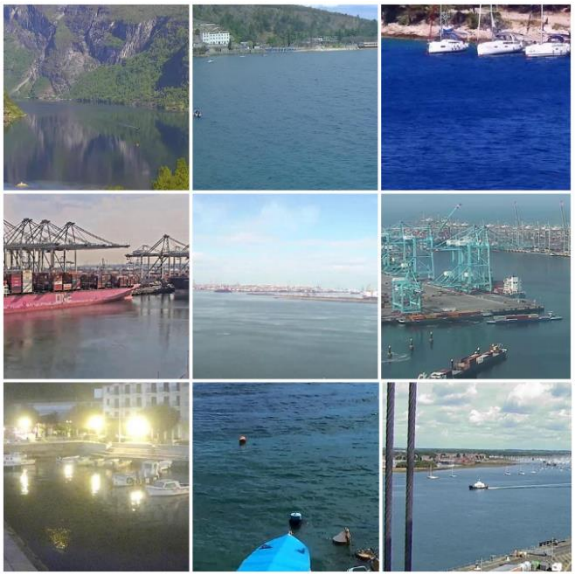


Figure 1. Examples of frames in source videos.

The frames were then manually annotated with bounding boxes and class labels in four classes: *Ship*, *Boat*, *Buoy*, *Swimmer* and *Unknown* for unidentified objects in the sea. For labeling purposes, boats were roughly defined as sailboats, small fishing boats, small passenger boats, small and medium yachts while ships cargo vessels, ferries, medium and large passenger ships, large yachts etc. were labeled as ships. Since there were very few instances of the *swimmer* class, in the final dataset this class was merged with the *unknown* class. The final dataset contains 135 images with 2755 object instances (Table 1). Regarding the object sizes, boat class was ususally the largest, due to position of several cameras that were in marinas where boats passed close by. Ship class is usually visually smaller in the image because large ships were usuall much farther away from the camera. Buoy and unknown classes were the smallest regardless of the camera position, with buoy instances most often taking up under 0.01% of total image area, sometimes taking up only a few pixels.

Table 1. Number of object instances per class and median bounding box area as % of image area per class.

Class	Instances	Median area (% of image)
Boat	1537	0.93
Ship	486	0.13
Buoy	672	0.01
Unknown	60	0.02

Some examples of objects in the four classes are shown in Figure 2.

Figure 2. Example instances of object classes.

3.2 Experiment Setup

Two variants of the YOLOv7 architecture (Wang, Bochkovskiy & Liao, 2023) were used to train the models for the task of maritime object detection, with only the output number of classes modified according to the number of classes in the dataset.

The two model configurations used in this experiment correspond to the full YOLOv7 network and to the much simpler and computationally less demanding YOLOv7-tiny network. The standard YOLO7 architecture is tailored for use on desktop GPUs with 415 layers and about 37 million parameters and requires about 104.7 billion floating point computations for inference. The second variant used in this experiment is the Yolov7-tiny, which comprises 263 layers and about 6 million parameters, and is much less computationally demanding, requiring about 13.8 billion floating point operations. The input image size for both models was 640x640 pixels.

The set task is to detect objects of *ship*, *boat*, *buoy*, and *unknown* classes in images.

The training was done using transfer learning from the model weights pretrained on the MS COCO (Lin et al., 2014) dataset and available at the YOLOv7 repository (Wang, 2023).

The dataset was first randomly split into training, validation, and test sets in 60:15:25 ratio. Then, the training and validation sets were augmented with random 640x640 crops of original images, so that the training set consisted of a mixture of 81 original images and 276 crops, 357 images in total.

The training was performed for 3000 epochs on the training set with learning rate set to 0.01. In comparison to the default training settings in the Github repository (Wang, 2023), some data augmentations were turned off, namely, image translation and scaling, mosaic, mixup and paste probabilities were set to 0.

After 3000 epochs, the weights that performed best on the validation set were tested on the test set.

Models were first evaluated on the rescaled original test images that are typically sized 1920x1080 pixels (scaling strategy), and then on the extracted 640x640 slices that correspond to the native network input size (slicing strategy). To recalculate the ground truth bounding boxes in the sliced images, the yolo-tiling script was used (Neskorozhenyi, 2023)

All experiments were performed in the PyTorch implementation from the YOLOv7 repository using PyTorch 1.10.1, Python 3.8.10 and Ubuntu 20.04 operating system, on a Nvidia RTX 3090 GPU.

Standard Average precision (AP) and mean average precision (mAP) measures (Padilla, Netto & Da Silva, 2020) at 50% IoU threshold and in 50% to 95% IoU range were used to evaluate the performance of the models.

4 Results and Discussion

The results of detection with the trained YOLOv7 and YOLOv7-tiny models are shown in Table 2.

Table 2. Detection results for YOLOv7 and YOLOv7-tiny using scaled and sliced input strategies.

		YOLOv7-tiny		**YOLOv7**	
Class	mAP	Input Scaled	Input Sliced	Input Scaled	Input Sliced
Boat	IoU>0.5	0.687	**0.846**	0.287	0.699
	IoU .5:.95	0.355	**0.639**	0.111	0.488
Ship	IoU>0.5	**0.894**	0.839	0.741	0.555
	IoU .5:.95	0.6	**0.605**	0.414	0.357
Buoy	IoU>0.5	0.532	**0.725**	0.191	0.659
	IoU .5:.95	0.215	**0.426**	0.064	0.39
Un-known	IoU>0.5	**0.97**	0.666	0.187	0.335
	IoU .5:.95	0.38	**0.533**	0.119	0.302
Mean	IoU>0.5	**0.771**	0.769	0.140	0.422
	IoU .5:.95	0.388	**0.551**	0.128	0.362

Although the full YOLOv7 model is much more complex, the obtained performance was in this much better with the simpler YOLOv7-tiny model. This might be due to the quite modest dataset size, and the much larger number of parameters of the full YOLOv7 model that require more training data. Further discussion thus mostly focuses on the YOLOv7-tiny model.

For the *Boat* and *Buoy* classes, the slicing strategy yields significantly better results for both mAP@0.5 and mAP@0.5:0.95 measures, for both models, while for the Ship class there are no significant differences in case of the YOLOv7-tiny model, and for the full model actually performs worse.

The *Unknown* class is underrepresented in the dataset yet is a catch-all class for all unrecognized objects so it may exhibit a great variance in appearance. The main purpose of the class is to prevent the classifier from assigning another class label to parts of image that have a quality of "objectness", but can't be classified into the other classes. Further investigation should be made whether this actually affects the precision of detection of other classes.

Another view of the object detection performance is shown with a confusion matrix (Tables 3 and 4).

Notably, there is very little confusion between the ship and boat classes, and the greatest source of errors are either completely missed detections (background false negative) or false positive detections (background false positive). Missed detections, or false negatives, are as expected largest for the *Buoy* class which appears smaller in the images than the other content classes (*Ship* and *Boat*), although distant ships and boats may also appear small. The false positive problem is present for both Buoy and Boat classes, possibly due to large variety of appearances of boats and small size of buoys.

Table 3. Confusion matrix for the YOLOv7-tiny model and scaling input strategy

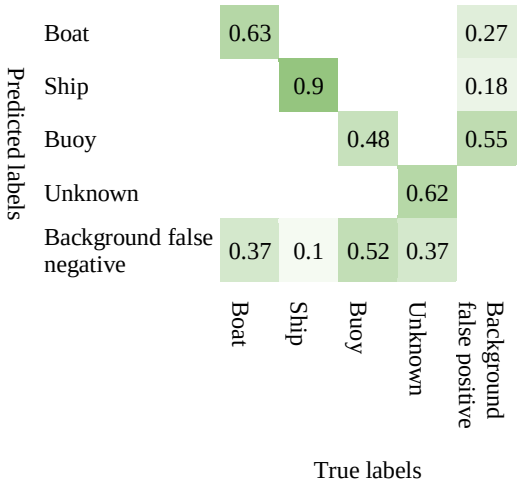
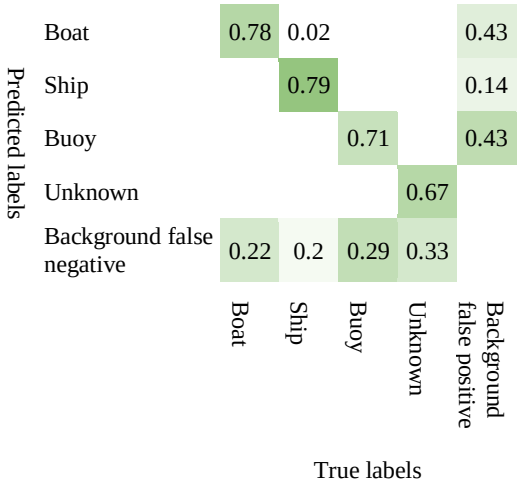


Table 4. Confusion matrix for the YOLOv7-tiny model and slicing input strategy.



Comparing the confusion matrices for the scaling and slicing input strategies (Tables 3 and 4), again it is apparent that slicing helps detection of small objects, where for the Buoy class the percent of true positive detections increased from 48 to 71, and for the Boat class slightly less dramatic from 63 to 78 percent. The true positive rate for the Ship class however fell 10 percent, possibly due to slicing off parts of ship that the network relied on for detection.

This confirms that for the task of detecting small objects, slicing the input image into tiles should be worthwhile despite the added complexity and the need to process the same image several times (equal to the

number of slices), especially if high recall (low missed detection rate) is needed.

A few examples of detection with both models using the scaling strategy are shown in Figure 3, and using the slicing strategy in Figure 4.

Figure 3. Example detection results for the scaling input strategy. Top row: Ground truth, middle row: YOLOv7, bottom row: YOLOv7-tiny.

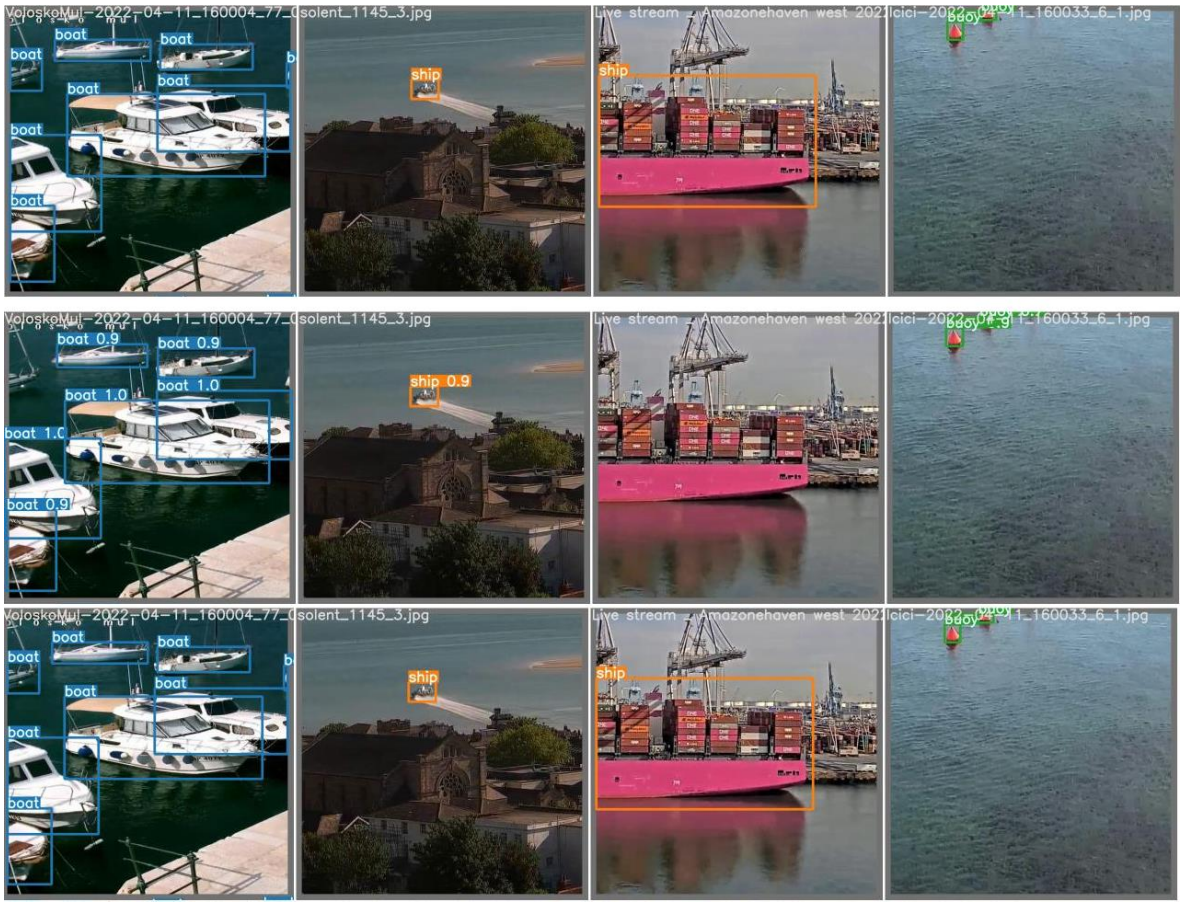


Figure 4. Example detection results for the slicing input strategy. Top row: Ground truth, middle row: YOLOv7, bottom row: YOLOv7-tiny.

5 Conclusion

In this paper, we explored the task of object detection in maritime images on a custom dataset prepared from surveillance cameras. The simpler and faster YOLOv7-tiny performed better than full YOLOv7 model in this experiment, pointing to both suitability of that model for the task and the likely necessity of gathering more training data to fully utilize the more complex model architecture. However, the YOLOv7-tiny architecture is well-suited to the task as is less computationally demanding than the full model and is aimed at the edge GPU devices, that may conceivably be deployed on site near the already installed cameras, avoiding the penalty of high video compression if streaming video is used as input for object detection.

To deal with detection of small objects, which may be important for hazard avoidance or rescue situations, an alternative input strategy to simply rescaling the input image at the network input was considered. Here, the input image was sliced into segments of exact size of network input, without first resizing the image. The experimental results show that this greatly improves detection of small objects such as buoys, raising the true positives from 41 to 71 percent on the test set using the YOLOv7-tiny model.

The detection performance of small objects is still far from perfect, and additional improvements should be considered in future work, for example integration of detections from multiple consecutive frames.

In addition, the dataset will be expanded to include even more variety of weather conditions and locations, and the method tested on other maritime datasets such as SPSCD or SeaDroneSee.

Acknowledgments

This work was fully supported by the EU Horizon project "INNO2MARE: Strengthening the Capacity for Excellence of Slovenian and Croatian Innovation Ecosystems to Support the Digital and Green Transitions of Maritime Regions" (101087348) and University of Rijeka project uniri-drustv-18-288.

References

Benjumea, A., Teeti, I., Cuzzolin, F., & Bradley, A. (2023). *YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles* (arXiv:2112.11798). arXiv. http://arxiv.org/abs/2112.11798

Broad, A., Jones, M., & Lee, T. Y. (2018, September). Recurrent Multi-frame Single Shot Detector for Video Object Detection. In *BMVC* (p. 94).

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Gasienica-Jozkowy, J., Knapik, M., & Cyganek, B. (2021). An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. *Integrated Computer-Aided Engineering*, *28*(3), 221-235.

Krišto, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using YOLO. *IEEE access*, *8*, 125459-125476.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (pp. 740-755). Springer International Publishing.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.

Neskorozhenyi, R. (2023, June). yolo-tiling [GitHub repository]. GitHub. https://github.com/slanj/yolo-tiling

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. In 2020 international conference on systems, signals and image processing (IWSSIP) (pp. 237-242). IEEE.

Petković, M., Vujović, I., Lušić, Z., & Šoda, J. (2023). Image Dataset for Neural Network Performance Estimation with Application to Maritime Ports. *Journal of Marine Science and Engineering*, *11*(3), 578.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Sambolek, S., & Ivasic-Kos, M. (2021). Automatic person detection in search and rescue operations using deep CNN detectors. *Ieee Access*, *9*, 37905-37922.

Stojnić, V., Risojević, V., Muštra, M., Jovanović, V., Filipi, J., Kezić, N., & Babić, Z. (2021). A Method for Detection of Small Moving Objects in UAV

Videos. *Remote Sensing*, *13*(4), 653. https://doi.org/10.3390/rs13040653

Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781-10790).

Tong, K., & Wu, Y. (2022). Deep learning-based detection from the perspective of small or tiny objects: A survey. *Image and Vision Computing*, *123*, 104471.

Unel, F. O., Ozkalayci, B. O., & Cigla, C. (2019). The Power of Tiling for Small Object Detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 582–591. https://doi.org/10.1109/CVPRW.2019.00084

Varga, L. A., Kiefer, B., Messmer, M., & Zell, A. (2022). SeaDronesSee: A Maritime Benchmark for Detecting Humans in Open Water. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3686–3696. https://doi.org/10.1109/WACV51458.2022.00374

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7464-7475).

Wang, C. Y. (2023, June). Yolov7 (Version 0.1) [GitHub repository]. GitHub. https://github.com/WongKinYiu/yolov7/

Zhu, Q., Ma, K., Wang, Z., & Shi, P. (2023). YOLOv7-CSAW for maritime target detection. *Frontiers in neurorobotics*, *17*, 1210470. https://doi.org/10.3389/fnbot.2023.1210470

Explainable Artificial Intelligence Meets Active Learning: A Novel GradCAM-based Active Learning Strategy.

Karel Križnar*
University of Ljubljana
Faculty of mathematics and physics
Jadranska ulica 19, Ljubljana, Slovenija
kk4876@student.uni-lj.si

Jože M. Rožanec†
Jožef Stefan International Postgraduate School
Jamova cesta 39, 1000 Ljubljana, Slovenia
joze.rozanec@ijs.si

Blaž Fortuna
Qlector d.o.o.
Rovšnikova 7, Ljubljana, Slovenija
blaz.fortuna@qlector.com

Dunja Mladenić
Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
dunja.mladenic@ijs.si

Abstract. *Active learning is a subfield of machine learning that studies how to identify data instances that contribute most to the learning of a given learner and requests some oracle to provide complementary information to enhance the learner's learning process towards a specific goal. This paper proposes a novel active learning strategy for image classification tasks. Furthermore, we obtain insights that lay the ground for developing a novel early-stopping criteria to maximize the model's learning while reducing the required instances. The proposed active learning technique uses insights obtained via GradCAM activation maps to understand the cognition process triggered by each image in the learner and select those that trigger the most diverse patterns, presupposing that diverse GradCAM patterns point to new learning opportunities. To our knowledge, the approach is among the first to use insights from explainable artificial intelligence to drive data selection and annotation. Nevertheless, our results show that it does not achieve the same quality of results as the random and uncertainty sampling techniques we compared to. More research is required to understand how to enhance the models' performance with such a strategy.*

Keywords. Active Learning; Automated Visual Inspection; Early Stopping; Machine Learning; Explainable Artificial Intelligence; Neural Networks; Quality Assurance and Maintenance

1 Introduction

The Industry 4.0 paradigm fosters using advanced technologies to optimize manufacturing further by increasing the flexibility and efficiency of the manufacturing process and the product value over the whole manufacturing lifecycle (Frank et al., 2019). Among the enabling technologies, we find (i) the Internet of Things, (ii) Big Data, (iii) Cloud Computing, (iv) Digital Twin, (v) Additive Manufacturing, (vi) Autonomous Robots, (vii) Simulation, (viii), Cybersecurity, (ix) Augmented Reality, and (x) Artificial Intelligence (Suleiman et al., 2022). These technologies are used to improve manufacturing sustainability in multiple aspects, such as employee productivity, improved profit margin, intelligent production planning and control, manufacturing agility, and efficiency, among others (Ching et al., 2022).

Product quality is relevant for businesses as it fosters customer trust, enhances loyalty, and strengthens the brand's reputation (Yang et al., 2020). Therefore, quality control plays a key role in manufacturing, given it ensures compliance with standards and specifications. During the visual inspection, integrity, surface finish, and geometric dimensions can be assessed (Newman and Jain, 1995). The automated visual inspection aims to mitigate common issues that relate to the subjective nature of human inspection (e.g., operator-to-operator inconsistency and quality dependence on the employees' experience and well-being (See, 2012)), and scalability issues (e.g., speed of inspection and continuous execution of the inspection process among others (Chouchene et al., 2020)).

Machine vision is a sub-field of artificial intelligence that develops means to process images with different purposes. It is used in the context of visual inspection to detect and flag faulty products while recognizing their defects and allowing rapid intervention to remove such products, address failure root causes, and mitigate the void that results from the estimated and actually delivered manufactured products in the production pipeline (Javaid et al., 2022). The decreased cost of sensors, the use of machine learning algorithms,

*Karel Križnar and Jože M. Rožanec are co-first authors with equal contribution and importance

†Corresponding author: Jože M. Rožanec. Email: joze.rozanec@ijs.si

and other technologies are key enablers for the broader adoption of automated visual inspection in the context of Industry 4.0. Furthermore, an automated visual inspection can be considered an essential element of manufacturing and can have a broader impact on the overall organization under the Industry 4.0 paradigm (Konstantinidis et al., 2021).

Supervised machine learning models require labeled data to learn specific defect types. Labeling data is expensive. Furthermore, machine learning model training times have associated costs that must be considered to determine the value for money (Justus et al., 2018). Reducing training times has a direct impact on reducing such costs. Active learning provides means to select data instances to speed up the learning of a machine learning algorithm and therefore reduce training times, given fewer data instances are required to achieve certain discriminative performance.

The main contribution of this research is the development of a novel active learning technique and early stopping criteria. While the active learning criteria aims to achieve a steep learning curve for the machine learning algorithm, the early stopping criteria allows to identify when training should be stopped, given that the algorithm will not learn much more. Combining them, we aim to achieve an enhanced learning process of the machine learning algorithms while reducing the amount of labeled data that must be fed to the algorithm and the training times required to train a model. The experiments were performed considering two real-world use cases: a dataset provided by *Philips Consumer Lifestyle BV* corporation and an open dataset from Kolektor (Božič et al., 2021), which has become one of the standard datasets for automated visual inspection tasks.

To evaluate machine learning models, we measure the discriminative capability with the AUC ROC metric. Furthermore, we compare how many data instances are required by each active learning strategy to achieve their best performance and how many samples must be shown to the model before early stopping.

The rest of this paper is structured as follows: Section 2 presents related work, Section 3 introduces a novel active learning strategy, while the Section 4 describes the *Kolektor* and *Philips Consumer Lifestyle BV* use cases and datasets. Section 5 introduces the methodology and experiments. Section 6 describes and analyzes the results obtained. Finally, Section 7 offers our conclusions and outlines future work.

2 Related Work

Since deep learning algorithms for machine vision have achieved super-human performance on certain tasks (Ciregan et al., 2012), much research has been invested in how to enhance them. Machine learning models for automated visual inspection have been widely adopted, and state-of-the-art performance has been achieved with deep learning models (Pouyanfar et al., 2018; Beltrán-González et al., 2020). Nevertheless, training such models from scratch is computationally expensive. Furthermore, acquiring enough quality labeled data for training such models requires a considerable effort (Kocaguneli et al., 2012; Whang et al., 2023). Therefore, approaches such as transfer learning or few-shot learning have been proposed to reduce both (Zhuang et al., 2020; Wang et al., 2020). While both approaches exist independently of active learning, active learning can be used to enhance both.

While active learning has been researched in the past, attention to it has been gradually increasing as a means to reduce the labeling effort (Ren et al., 2021). To perform data selection, active learning methods mostly considered either the inherent characteristics of the data (e.g., informativeness, representativeness, and diversity (Wu, 2018)) or the model's behavior w.r.t. the data (e.g., the closeness of a datapoint to the decision boundary). Such methods considered the data selection process dissociated from the actual model's learning process. Nevertheless, there have been some exceptions (e.g., (Zhu et al., 2019), who proposed optimizing both objectives at once by formulating data selection as a robust optimization problem). We consider that active learning data selection policies developed up to now used inherent data properties or model behavior as proxies to understanding the actual model's learning process and decision boundaries evolution. The advent of explainable artificial intelligence has opened a wide range of opportunities. Explainable artificial intelligence is a sub-field of artificial intelligence concerned with enabling humans to understand machine learning models, trust and effectively manage them (Arrieta et al., 2020). While much effort is being invested into developing new explainable models, explainability techniques, and metrics to assess their quality, to our knowledge, insights obtained from explainable artificial intelligence have not been leveraged yet to drive data selection in an active learning context. Nevertheless, such insights enable selecting data with a more direct understanding of the ongoing learning process inside the machine learning model.

Multiple explainability techniques have been devised to provide insights into machine vision models. Among the most popular techniques we find GradCAM (Selvaraju et al., 2017) (it uses the gradient information to understand how strongly the neurons activate), DeepLIFT (Shrikumar et al., 2016) (uses a derivative-based method to propagate activation differences, and determine how changes in the image would affect predictions), or Smooth-Grad (Smilkov et al., 2017) (measures local sensitivity based on small image perturbations). Many ways have been devised to convey explainable artificial intelligence outcomes for machine vision. Authors have considered overlaying the original image with a cloud of points, an activation or heat map, or some outline (Hudon et al., 2021), displaying

only a relevant part of the image (Ribeiro et al., 2016), or highlighting it (Buhrmester et al., 2021).

While Meng et al. (Meng et al., 2020) consider there is a research void regarding using active learning manufacturing domain, we have found some examples where it was successfully applied to the quality inspection use cases. In particular, Van et al. (van Garderen, 2018) explored how active learning could enhance the learning process of a machine learning model measuring the local displacement between layers on a chip. In the same line, Shim et al. (Shim et al., 2020) described a technique developed to select wafer maps, which provide key information to engineers for detecting root causes of failure in the semiconductor manufacturing process. Finally, Dai et al. (Dai et al., 2018) described how active learning was used to automatically enlarge a dataset when training a model to recognize solder joint defects in printed circuit boards.

3 GradCAM$_{\mathbf{avg}}$

So far, Active learning techniques have mainly focused on the characteristics of the data or the models' behavior to find instances that would be helpful to refine the decision boundary. Nevertheless, we consider both approaches as proxies to direct insights into models' rationale, which can now be obtained using explainable artificial intelligence techniques. One such explainable artificial intelligence technique is GradCAM (Selvaraju et al., 2017), which computes the importance map by taking the derivative of the reduction layer output for a given class with respect to a convolutional feature map. The importance map can be rendered as an activation map for visualization purposes.

For this research, we have developed a novel active learning technique named GradCAM$_{\text{avg}}$, which consists of generating GradCAM activation maps for all of the images in the pool and computing the structural similarity index measure (SSIM) between the average GradCAM representation of the images seen so far and the rest of the unlabeled images. The unlabeled images are then selected considering those most dissimilar compared to the average GradCAM image. The intuition behind the algorithm is that explainable artificial intelligence techniques can provide a means to understand how the model perceives each unlabeled image and the ones seen so far. The unlabeled images whose GradCAM activation map is most dissimilar from those seen so far are likely to offer more novel information to the model and result in a steeper learning curve. We detail the procedure in Algorithm 1.

Little work has been done combining explainable artificial intelligence and active learning. E.g., Ghai et al. (2021) used explainable artificial intelligence insights to reduce opaqueness regarding the samples presented to the human oracle labeling the data. To our knowledge, GradCAM$_{\text{avg}}$ is among the first active learning techniques leveraging insights from explainable artificial intelligence to boost models' learning. Concurrently similar approaches have been developed by Zajec et al. (2023). We do not know of other methods leveraging explainable artificial intelligence in the context of active learning.

4 Use Cases

The research we performed was based on two real-world use cases concerning the visual inspection of products manufactured by two European companies (*Kolektor Group d.o.o.* from Slovenia and *Philips Consumer Lifestyle BV* from The Netherlands).

Kolektor Group d.o.o. published a dataset (Kolektor SDD2 a.k.a. KSDD2) of over 3.000 images based on a real-world example (Božič et al., 2021). The dataset was constructed from color images captured in a controlled environment by a visual inspection system and annotated by the company and a group of researchers. The dataset has been considered in several publications related to automated visual inspection. We adopt it as a reference dataset for which the results obtained can be compared later with further research on this topic.

Philips Consumer Lifestyle BV has printing machine setups for several products, which are manually handled and inspected to determine their quality. If a defect is observed, the manufactured product is removed from the manufacturing line to ensure only products conforming to high-quality standards are left. Human inspectors spend several seconds handling, inspecting, and labeling the products. If automating the visual inspection process, it is expected that human inspectors would require several seconds to label each product to create an initial dataset. Therefore, any gains in reducing the needed labeled data and labeling times directly translate into cost savings. *Philips Consumer Lifestyle BV* provided a dataset of over 3,000 annotated images. While fine-grained labels were supplied, we trained the models to recognize only whether a defect exists in the manufactured product.

5 Methodology and Experiments

5.1 Methodology

When training the machine learning models, we performed stratified cross-validation (Zeng and Martinez, 2000). We adopted *k=10* based on recommendations by Kuhn et al. (2013) and considered one fold for testing, one for validation, and the rest was used to simulate a pool of unlabeled data for active learning, from which we sourced the data samples. We used a ResNet-18 network (He et al., 2016) and trained it incrementally by sourcing batches of 32 images from the active learning pool until pool exhaustion.

Algorithm 1 $GradCAM_{avg}$ active learning algorithm. We select batches of images based on how different their GradCAM activation maps are from the average GradCAM activation map obtained from the images labeled so far.

$Set_{UnlabeledImages} \leftarrow \{img_i, img_{i+1}, img_{i+2}, \ldots, img_n\}$
$Set_{LabeledImages} \leftarrow \{img_j, img_{j+1}, img_{j+2}, \ldots, img_m\}$
procedure SELECTINSTANCES($Set_{UnlabeledImages}$, $Set_{LabeledImages}$, $n_{images\,to\,select}$)
    $Set_{GradCAM_{LabeledImages}} \leftarrow \emptyset$
    $Set_{GradCAM_{Unlabeled}} \leftarrow \emptyset$
    for $img \in Set_{LabeledImages}$ **do** $Set_{GradCAM_{LabeledImages}} \cup \{GradCAM(img)\}$
    end for
    for $img \in Set_{UnlabeledImages}$ **do** $Set_{GradCAM_{Unlabeled}} \cup \{GradCAM(img)\}$
    end for
    $GradCAM_{avg} \leftarrow avg(Set_{GradCAM_{LabeledImages}})$
    $Dictionary_{GradCAM_{Unlabeled}\,to\,SSIM} \leftarrow \{\}$
    for $GradCAM_{img} \in Set_{GradCAM_{Unlabeled}}$ **do**
        $Dictionary_{GradCAM_{Unlabeled}\,to\,SSIM} \leftarrow GradCAM_{img}, d_{SSIM}(GradCAM_{avg}, GradCAM_{img})$
    end for
    return $top(Dictionary_{GradCAM_{Unlabeled}\,to\,SSIM}, n_{images\,to\,select})$
end procedure

5.2 Experiments

To understand the inherent performance limitations of the machine learning models on the datasets, we trained a model on all of the data available in the active learning pool. We considered the resulting performance to be the best that could be obtained by a ResNet-18 model trained on that particular data.

We trained the ResNet-18 model considering three active learning strategies: (i) random sampling (baseline), (ii) uncertainty sampling (selects samples with the highest uncertainty, computed as the difference between the highest predicted score and one), and (iii) a novel technique of our own that we named GradCAM$_{avg}$, which we described in Section 3. We examined the models' performance over time for each case and analyzed when the learning performance peaked. Furthermore, we took the subset of images presented to the model until learning peaked in the active incremental learning and trained the model with it in a batch setting. This enabled us to understand differences in performance between models trained in batch or incremental learning settings while reducing the amount of data shown to them based on the active learning criteria.

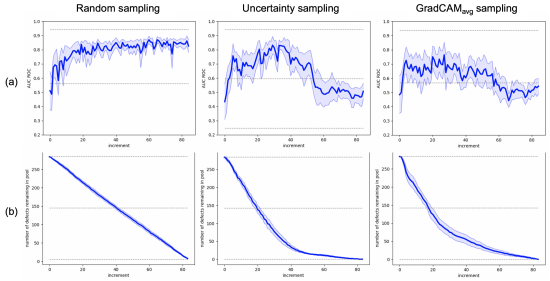


Figure 1: We display plots in two rows: (a) discriminative performance over time and (b) number of images regarding defective products left in the active learning pool over time. The plots correspond to the Kolektor SDD2 dataset.

6 Results and Analysis

6.1 Active learning: batch vs. incremental learning

Among the experiments, we were first interested in assessing how data sampling strategies affected the learning process and discriminative performance of the ResNet-18 model. Regarding discriminative performance (see Table 1), we found that the random sampling led to the highest AUC ROC score (0.8967) for the Kolektor SDD2 dataset, while uncertainty sampling and GradCAM$_{avg}$ lagged with an AUC ROC score of 0.8647 and 0.8502 in an incremental learning setting. Nevertheless, when considering the same sampled images in a batch learning setting, the differences were less pronounced: random and uncertainty sampling achieved almost the same mean AUC ROC (0.9103 and 0.9187), and GradCAM$_{avg}$ slightly lagged (0.9006). For the Philips shavers dataset, we verified the same performance pattern concerning active learning techniques in the incremental learning setting. Nevertheless, uncertainty sampling and GradCAM$_{avg}$ lagged behind random sampling when training the model in a batch setting.

6.2 Active learning: learning curves vs. pool composition

When analyzing the results (see Fig. 1 and Fig. 2), we focused on two plots: (a) AUC ROC plots, display-

AUC ROC (average)		Active Learning strategy (data selection)		
Dataset	**Learning strategy**	**Random**	**Uncertainty**	**GradCAMavg**
Kolektor SDD2	**Incremental Learning**	0.8967±0.0133	0.8647±0.0098	0.8502±0.0310
	Batch Learning	0.9103±0.0161	0.9187±0.0215	0.9006±0.0302
Philips (shavers)	**Incremental Learning**	0.9416±0.0100	0.9124±0.0123	0.9095±0.0162
	Batch Learning	0.9799±0.0088	0.9526±0.0086	0.9395±0.0113

Table 1: AUC ROC for binary classification setting, comparing three active learning techniques and how the same set of images affects learning when considering incremental or batch learning.

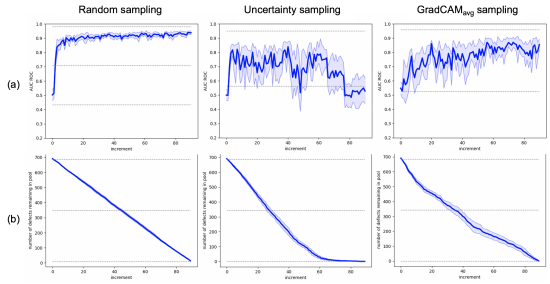


Figure 2: We display plots in two rows: (a) discriminative performance over time and (b) number of images regarding defective products left in the active learning pool over time. The plots correspond to the Philips Consumer Lifestyle BV shavers dataset.

ing how models' discriminative performance evolved and (b) plots showing the number of samples of defective products left in the pool of unlabeled data. We found that the AUC ROC curve usually started decreasing close to the knee point of the plot (b).

The curve in plot (b) strongly reminded the scree plot curves. Based on this observation, we wondered whether a similar rule of thumb as the one used to select principal components in principal component analysis with a scree plot (Satopaa et al., 2011) could be used in this case to determine how long to sample data from the active learning pool. While the information displayed in plot (b) assumes the labels are known, the active learning setting assumes unlabeled data. Nevertheless, some soft labeling strategies could be used to mitigate this issue.

When considering the results presented in Table 2 and Table 3, we found that in all cases, the model with the best discriminative performance in the validation set until the knee point in the plot had better performance than the model at the knee point. Nevertheless, these models displayed inferior performance in all but one case than the best discriminative model obtained by annotating the data samples until pool exhaustion. Furthermore, the best models lead by at least 0.1 AUC ROC points when compared against the best models obtained until the knee point.

Analyzing the Kolektor SDD2 dataset, we found that the best-performing models originated after the knee point for random sampling. This was not the case for the uncertainty sampling and GradCAM$_{avg}$ strategies where, on average, the best models appeared before the knee point, but not always. On the other hand, in the Philips shavers dataset, the best models were observed after the knee point when using random and GradCAM$_{avg}$ sampling. This was not the case for uncertainty sampling, where, on average, the best models appeared before the knee point. Moreover, the pattern of having the best models appear before the knee point seems to correlate to the shape of the plot (b): the best models appear before the knee point when the plots have a part of the curve that resembles debris fallen from a mountain and lying at its base. While uncertainty sampling displayed a consistent behavior across both datasets, early stopping at the knee point significantly degraded the model's performance when trained incrementally.

7 Conclusion

In this research, we describe a novel active learning technique we named GradCAM$_{avg}$ given it aims to select candidate images for labeling from unlabeled data based on their SSIM score when compared against the average of GradCAM activation maps for the images labeled up to that point in time. To our understanding, this method is among the first ones leveraging insights from explainable artificial intelligence techniques to drive data selection. When analyzing AUC ROC plots and plots describing the composition of the unlabeled dataset, we found that active learning strategies that greedily source data instances of defective products tend to achieve best-performing models before the knee point at those plots. While applying such criteria led to sub-optimal results in this research, more effort is required to overcome this limitation and capitalize on this insight to develop a novel early stopping criteria. We envision that such an early stopping criteria could guide active learning data annotation efforts. Future work will attempt to evaluate the GradCAM$_{avg}$ technique on a comprehensive number of datasets and develop a novel early-stopping technique based on the insights of this research.

Number of samples seen by the model		Active Learning strategy (data selection)		
Dataset	**Cut-off criteria**	**Random**	**Uncertainty**	**GradCAMavg**
Kolektor SDD2	**Best Discriminative Performance**	1502±416	1088±160	672±288
	Knee Point in Plot	1184±256	1184±96	800±160
	Best Discriminative Performance until Knee Point	896±320	1024±128	448±192
Philips (shavers)	**Best Discriminative Performance**	2272±384	960±352	2112±416
	Knee Point in Plot	1600±352	1664±128	768±256
	Best Discriminative Performance until Knee Point	1792±384	768±320	544±224

Table 2: Number of samples for binary classification setting, comparing three active learning techniques. We aim to understand whether the best models are learned before or after the knee point. Furthermore, we are interested in understanding when the best model before a knee point originates.

AUC ROC (average)		Active Learning strategy (data selection)		
Dataset	**Cut-off criteria**	**Random**	**Uncertainty**	**GradCAMavg**
Kolektor SDD2	**Best Discriminative Performance**	0.8967±0.0133	0.8647±0.0098	0.8502±0.0310
	Knee Point in Plot	0.7965±0.0348	0.7183±0.0665	0.6646±0.1236
	Best Discriminative Performance until Knee Point	0.8216±0.0298	0.8255±0.0317	0.7137±0.0917
Philips (shavers)	**Best Discriminative Performance**	0.9103±0.0161	0.9187±0.0215	0.9006±0.0302
	Knee Point in Plot	0.9147±0.0195	0.5819±0.1481	0.6445±0.1205
	Best Discriminative Performance until Knee Point	0.9130±0.0182	0.7458±0.0979	0.6942±0.0641

Table 3: AUC ROC for binary classification setting, comparing three active learning techniques. We aim to understand the performance of the best overall model and how it compares against the best model obtained until the knee point.

Acknowledgments

This work was supported by the Slovenian Research Agency and the European Union's Horizon 2020 program project STAR under grant agreement number H2020-956573.

References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115.

Beltrán-González, C., Bustreo, M., and Del Bue, A. (2020). External and internal quality inspection of aerospace components. In *2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, pages 351–355. IEEE.

Božič, J., Tabernik, D., and Skočaj, D. (2021). Mixed supervision for surface-defect detection: from weakly to fully supervised learning. *Computers in Industry*.

Buhrmester, V., Münch, D., and Arens, M. (2021). Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, 3(4):966–989.

Ching, N. T., Ghobakhloo, M., Iranmanesh, M., Maroufkhani, P., and Asadi, S. (2022). Industry 4.0 applications for sustainable manufacturing: A systematic literature review and a roadmap to sustainable development. *Journal of Cleaner Production*, 334:130133.

Chouchene, A., Carvalho, A., Lima, T. M., Charrua-Santos, F., Osorio, G. J., and Barhoumi, W. (2020). Artificial intelligence for product quality inspection toward smart industries: quality control of vehicle non-conformities. In *2020 9th international conference on industrial technology and management (ICITM)*, pages 127–131. IEEE.

Ciregan, D., Meier, U., and Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3642–3649. IEEE.

Dai, W., Mujeeb, A., Erdt, M., and Sourin, A. (2018). Towards automatic optical inspection of soldering defects. In *2018 International Conference on Cyberworlds (CW)*, pages 375–382. IEEE.

Frank, A. G., Dalenogare, L. S., and Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210:15–26.

Ghai, B., Liao, Q. V., Zhang, Y., Bellamy, R., and Mueller, K. (2021). Explainable active learning (xal) toward ai explanations as interfaces for machine teachers. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–28.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hudon, A., Demazure, T., Karran, A., Léger, P.-M., and Sénécal, S. (2021). Explainable artificial intelligence (xai): how the visualization of ai predictions affects user cognitive load and confidence. In *Information Systems and Neuroscience: NeuroIS Retreat 2021*, pages 237–246. Springer.

Javaid, M., Haleem, A., Singh, R. P., Rab, S., and Suman, R. (2022). Exploring impact and features of machine vision for progressive industry 4.0 culture. *Sensors International*, 3:100132.

Justus, D., Brennan, J., Bonner, S., and McGough, A. S. (2018). Predicting the computational cost of deep learning models. In *2018 IEEE international conference on big data (Big Data)*, pages 3873–3882. IEEE.

Kocaguneli, E., Menzies, T., Keung, J., Cok, D., and Madachy, R. (2012). Active learning and effort estimation: Finding the essential content of software effort estimation data. *IEEE Transactions on software engineering*, 39(8):1040–1053.

Konstantinidis, F. K., Mouroutsos, S. G., and Gasteratos, A. (2021). The role of machine vision in industry 4.0: an automotive manufacturing perspective. In *2021 IEEE international conference on imaging systems and techniques (IST)*, pages 1–6. IEEE.

Kuhn, M., Johnson, K., et al. (2013). *Applied predictive modeling*, volume 26. Springer.

Meng, L., McWilliams, B., Jarosinski, W., Park, H.-Y., Jung, Y.-G., Lee, J., and Zhang, J. (2020). Machine learning in additive manufacturing: A review. *Jom*, 72(6):2363–2377.

Newman, T. S. and Jain, A. K. (1995). A survey of automated visual inspection. *Computer vision and image understanding*, 61(2):231–262.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., and Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, 51(5):1–36.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., and Wang, X. (2021). A survey of deep active learning. *ACM computing surveys (CSUR)*, 54(9):1–40.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops*, pages 166–171. IEEE.

See, J. E. (2012). Visual inspection: a review of the literature. *Sandia Report SAND2012-8590, Sandia National Laboratories, Albuquerque, New Mexico*.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.

Shim, J., Kang, S., and Cho, S. (2020). Active learning of convolutional neural network for cost-effective wafer map pattern classification. *IEEE Transactions on Semiconductor Manufacturing*, 33(2):258–266.

Shrikumar, A., Greenside, P., Shcherbina, A., and Kundaje, A. (2016). Not just a black box: Learning important features through propagating activation differences. *arXiv preprint arXiv:1605.01713*.

Smilkov, D., Thorat, N., Kim, B., Viégas, F., and Wattenberg, M. (2017). Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*.

Suleiman, Z., Shaikholla, S., Dikhanbayeva, D., Shehab, E., and Turkyilmaz, A. (2022). Industry 4.0: Clustering of concepts and characteristics. *Cogent Engineering*, 9(1):2034264.

van Garderen, K. (2018). Active learning for overlay prediction in semi-conductor manufacturing.

Wang, Y., Yao, Q., Kwok, J. T., and Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34.

Whang, S. E., Roh, Y., Song, H., and Lee, J.-G. (2023). Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, pages 1–23.

Wu, D. (2018). Pool-based sequential active learning for regression. *IEEE transactions on neural networks and learning systems*, 30(5):1348–1359.

Yang, J., Li, S., Wang, Z., Dong, H., Wang, J., and Tang, S. (2020). Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges. *Materials*, 13(24):5755.

Zajec, P., Rožanec, J. M., Theodoropoulos, S., Fontul, M., Koehorst, E., Fortuna, B., and Mladenić, D. (2023). Few-shot learning for defect detection in manufacturing. submitted.

Zeng, X. and Martinez, T. R. (2000). Distribution-balanced stratified cross-validation for accuracy estimation. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(1):1–12.

Zhu, D., Li, Z., Wang, X., Gong, B., and Yang, T. (2019). A robust zero-sum game framework for pool-based active learning. In *The 22nd international conference on artificial intelligence and statistics*, pages 517–526. PMLR.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76.

Towards Computational Content Analysis of Crises-Related News in Electronic Media

Vedran Orešković
Faculty of Informatics and Digital Technologies
University of Rijeka
Radmile Matejčić 2, 51000 Rijeka, Croatia
vedran.oreskovic@student.uniri.hr

Ana Meštrović, Slobodan Beliga
Faculty of Informatics and Digital Technologies
Center for Artificial Intelligence and Cybersecurity
University of Rijeka
Radmile Matejčić 2, 51000 Rijeka, Croatia
{amestrovic, sbeliga}@inf.uniri.hr

Abstract. *In this study, a computational content analysis of crisis-related articles published in electronic media for two crises, namely COVID-19 and the Russia-Ukraine war, is conducted. A set of methods for content analysis is proposed, which involves the combination of exploratory analysis, main topic filtering, and performing named entity recognition enhanced with network analysis. The main finding of this study is a list of the most frequently mentioned persons, locations, and organizations in the news articles of both crises and how they are connected. The results obtained affirm the suitability of the proposed approach for crisis-related content analysis.*

Keywords. natural language processing, NER, computational content analysis, electronic media news, COVID-19, Russia-Ukraine war, eVaražd in.hr

1 Introduction

Analysis of electronic media content is a powerful tool for gaining insights into public opinion. This type of analysis is particularly important in domains related to global crises such as wars, pandemics, climate change, migration crises, etc. as it can help in crisis management. Crisis-related communication in electronic media leads to a massive amount of news articles that can cause an infodemic (Eysenbach, 2002, Eysenbach, 2009). Recently, we have experienced an infodemic caused by the COVID-19 pandemic (Beliga et al. 2021; Beliga et al. 2022; Eysenbach, 2020). The main characteristic of an infodemic is an overabundance of information, misinformation, and disinformation (Eysenbach, 2002). This is a serious problem because, in the event of a crisis, the need for information increases beyond the usual level and people try to get informed as quickly as possible. Thus, automated computational analysis can be of great help in providing fast and relevant information during a crisis.

Many studies use various natural language processing (NLP) techniques to uncover certain aspects of crisis-related content published in electronic media. To identify public perceptions, opinions, and attitudes about specific crisis-related topics, researchers typically combine topic modelling and sentiment analysis (Boon-Itt & Skunkan, 2020) and/or perform named entity recognition (Imran et al., 2015).

An extensive NLP-based analysis of COVID–19-related content in the Croatian language published in electronic media have already been performed (Beliga et al., 2021; Beliga et al., 2022; Bogović et al., 2021, Ilić & Beliga, 2021; Pandur et al., 2021) as well as in social media (Babić et al., 2021a; Babić et al., 2021b; Meštrović et al., 2022). In (Beliga et al. 2021) the authors performed a longitudinal analysis of the COVID-19–related content on electronic media based on natural language processing, extracted main keywords, terminology, and entities, and analysed trends and changes. In (Beliga et al., 2022) the authors combined NLP and statistics and showed that there was an infodemic present in the Croatian news. Next, in (Ilić & Beliga, 2021) the authors performed sentiment analysis of COVID-19 news articles published in electronic media, while in (Pandur et al., 2021) they analysed COVID-19 news articles by combining topic modelling and sentiment analysis.

The results of the authors' previous studies provide detailed insights into the COVID-19–related content published in electronic media. One possible direction for expanding previous research is to compare the content of different crises. Therefore, in this study, the authors extend their previous research to the new domain of news about the Russia-Ukraine war (RUW). The main objective of this study is to analyse and compare the content of news published in electronic media related to the two crises. The authors perform content analysis in the sense of exploratory analysis and apply methods from the field of NLP, specifically topic filtering and named entity recognition (NER).

First, news articles from the *eVarazdin.hr* electronic media were collected and the *eVarazdin-19-23* dataset was prepared. In the next step, articles related to COVID-19 and articles related to the RUW were filtered, afterwards an exploratory analysis was per-

formed. The authors identify different topics in the dataset and compare the number of articles related to different topics and categories. In the last step, NER was applied to extract important entities from the news articles. Based on the NER results, the most frequent entities were organised into the network connecting entities which appear together in the same article. This way it is possible to further explore relations between entities.

The proposed approach provides detailed insights into the content of both crises and allows us to better understand crises. Results of this study indicate that COVID-19–related topics were much more present in news articles than RUW-related topics. Furthermore, NLP-based methods in combination with other techniques such as network analysis can reveal the specifics of the content of both crises. The proposed methodology is applicable to different crises, and the same approach can be easily ported to another dataset.

The rest of the paper is organised as follows. In the next section, an overview of the related work is given. The third section explains the dataset and methods. In the fourth section, the details of the proposed approach and the main results are described. The last section concludes this study with a discussion and recommendations based on the results.

2 Related Work

Fu et al. (2022) suggest an approach based on natural language processing techniques and machine learning to mitigate **flood hazards**. In order to avoid floods, it is very important to identify flood-prone areas. In the conducted study, they emphasize the importance of collecting textual data from the media in assessing the sensitivity of urban areas to floods. In a case study of the city of Dalian, China, the locations of floods were identified from news media data using a NER model. Then, a method based on frequency or distance was used to improve the quality of the extracted flood locations. Finally, flood conditioning factors that included information about historical flood locations were entered into a Support Vector Machine (SVM) model to assess flood susceptibility. In the obtained flood susceptibility map, the high flood susceptibility areas got a recall of 90% compared with the high flood hazard areas in the planning report (Fu et al., 2022).

Liu et al. (2021) have explored how useful the Reddit social media platform is, to monitor the **COVID-19** pandemic. In addition, they tried to answer the question of how human behavior changes over the course of the COVID-19 pandemic in North Carolina. They collected data for the research from the COVID-19 pandemic-related posts from North Carolina subreddit communities. In the task of named entities recognition, they developed a customized system in which they focused on 5 specific categories: Distancing, Disinfection, Personal Protective Equipment, Symptoms and Testing. In addition to NER, the monitoring of the COVID-19 pandemic was carried out through other NLP tasks: feature engineering (using CBOW, Skip-Gram, Glove, and BERT-based sentence clustering), topic entities discovery (Cosine Similarity and LDA topic modeling) and frequency statistics. They concluded that the comprehensiveness of the mentioned NLP methods shows effective monitoring and discovering of how the public's concerns changed over the course of the pandemic, all based on data collected from Reddit. However, the presented results show that representative social media data can be utilized to surveil the epidemic situation in a specific community (Liu et al., 2021).

Truong et al. (2021) presented a dataset of approximately 10k informative sentences about **COVID-19** patients filtered from news articles published on reputable Vietnamese online news sites. When filtering sentences, care was taken to ensure that they reported confirmed, suspected, recovered, or dead cases, as well as the travel history or location of the cases. The authors defined 10 types of entities associated with COVID-19 patients that may be particularly useful for downstream applications in any type of future epidemic. The created dataset represents the largest Vietnamese dataset with corresponding NER annotations (35k entities). Experiments on new a NER dataset compare strong baselines and find that the input representations and the pre-trained language models all have influences on this COVID-19 related NER task (Truong et al., 2021). Datasets like this certainly have an important role in future COVID-19 research and NLP applications related to the Vietnamese NLP community.

Alam et al. (2020) conducted case studies of **hurricanes** Harvey, Irma, and Maria and produced descriptive summaries using artificial intelligence techniques – NLP and computer vision. The studies were conducted using text and image data from Twitter posted during the three major disasters in 2017. One of the main questions of this study was whether and how computer analysis techniques of natural language processing, such as topic modelling, sentiment analysis, NER, etc. can be used to process text and image data in social media to improve situational awareness. The Stanford NER toolkit, which is based on CRFs, was used in the study. The authors investigated entities based on humanitarian categories, especially focused on donation and volunteering category. This focuses on the donation needs that people are writing about, the things that are needed, and in what places. The authors provide visualizations using word clouds to point out which organizations and celebrity people have the most frequent entity mentions (i.e., donate the most). Ultimately, the study showed numerous visualizations made on the basis of data obtained by NLP techniques (including NER). In addition, the successful combination of textual and image data was proven for a comprehensive analysis of crisis situations. Various types

of useful information can inform crisis managers and responders and facilitate the development of future automated systems for disaster management (Alam et al., 2020).

Eligüzel et al. (2022) propose a deep learning approach (RRN-based) for named entity recognition on tweets during **earthquake disaster** in Nepal. They used two different tools to recognize named entities, first the Natural Language Toolkit (NLTK) and then the General Architecture for Text Engineering (Gate). The obtained entities were then used as input data for RNN models. In addition, on the initially labeled data, they trained different variants of models with the original labeled data, from GloVe word embeddings to different variants of RNN models such as LSTM and BLSTM. They concluded that RNN-based approaches perform well in finding named entities. In emergencies, the results of this work can help reduce the effort required to detect event locations and enable better disaster management (Eligüzel, Çetinkaya & Dereli, 2022).

Garcia & Ladeira (2021) conducted a study that aims to propose an approach for the application of NER in the identification of possible adverse events related to the application of COVID-19 vaccines. NER is used for information extraction from social media (websites related to health topics and similar sources). In addition, the authors also conducted an extensive survey on the application of NER for information extraction, in the domain of **postmarket surveillance** for the supplementation of traditional post-market surveillance. The contribution of this research is in the systematization of the positive and negative sides of NER in the information extraction task to supplement traditional postmarket surveillance systems. More practical use of information extraction and NER was shown in work conducted by Nemes and Kiss (2021), where sentiment analysis on Twitter messages is augmented with information extraction and NER to get an even more comprehensive picture of the sentiments of people and to determine how people have related to **COVID-19** over a given period of the pandemic. By extracting additional information and analyzing named entities on the sentiment results categorized and labeled by RNN, they get a better insight into the sentiment. In other words, one gets a detailed insight into how people write on Twitter, what are the characteristics of their expression, which words are used in positive and which in negative attitudes/tweets, which people, locations and other entities are mentioned, but also the events that can affect tweets. With such an analysis they provide a whole new picture to traditional sentiment analysis (Nemes & Kiss, 2021).

Phopli et al. (2021) investigate the events of **natural disasters** by using information from a social network to report on crisis situations. The detection of natural disaster locations is based on the recognition of named entities (NE) in messages from Twitter. In doing so, they use Complex, or Social Network Analysis (CNA or SNA) techniques. Apropos, they analyze the network of named entities, and compute values for centrality measures for certain named entities at the local level of the network, such as: degree, betweeness, and closeness centrality measure. They conducted the experiment only on tweets written in Thai language in the topic of Super Typhon Hagibis blowing in Japan. Based on the graph theory, mentioned centrality measures and majority voting, they determine first 5 words that are related to the event for particular method, and finally only one most important word which can be used to indicate a location that had a natural disaster.

The authors found motivation for their investigation in this approach. They used NER to detect not only locations that are mentioned in electronic media news related to a particular crisis, but also named entities from other categories. However, they measured the importance of a single entity in the corpus based on graph theory using the centrality of a node in the network, which will be discussed in more detail in the following sections.

Related work on the crisis situation related to the **Russia-Ukraine war** mostly analyzes data collected from Twitter with the aim of sentiment analysis (Wadhwani et al., 2023) or topic detection and tracking (De Santis et al., 2023). To a slightly lesser extent, data collected from other social networks have also been analyzed, such as Facebook (Ngo et al., 2022), where sentiment analysis was performed, then data from the Chinese microblogging website Weibo, where emotion recognition and opinion clustering and mining were performed (Chen et al., 2022), and Reddit (Guerra & Karakuş, 2023), where sentiment analysis was also performed. In addition to the above works, there are many other works with sentiment analysis or topic modeling of data related to the Russia-Ukraine war. However, to the best of the authors' knowledge, they have not found any work that uses NER. With this work, they would like to fill the gap in the analysis of news in Croatian electronic media about the Russia-Ukraine war and use NER to analyse important entities.

As far as we know, this is the first study that provided a dataset of newspaper articles written in the Croatian language, collected from electronic media, for the domain that is related to the Russia-Ukraine war. In addition, it is also the first report on NER and the computational content analysis of the crisis-related to the RUW on Croatian electronic news texts.

3 Methodology

3.1 Dataset

News articles were collected from the *eVarazdin.hr*[1] web portal. *eVarazdin* is an independent Internet portal for the area of Varaždin County, which is read in the

[1] https://evarazdin.hr/

wider area of northern Croatia. According to available statistics from July 2023, about 64,000 users followed this portal on the social network Facebook[2].

eVarazdin was launched on March 4, 2010, and since 2013 it has been operating as a media company within the Varaždin consulting financial investment holding *Fine's Grupa*. With the transfer to the ownership of Fine's, the portal was recapitalized in 2015 in the amount of HRK 4,16 million, which makes it one of the most capitalized local media in the Republic of Croatia.

The data collection process included web scraping techniques to extract relevant data from electronic news portal *eVarazdin*. Python libraries such as *BeautifulSoup* (Richardson, 2023) were used to extract textual content from web pages.

The news on the portal are classified into nine newspaper categories: *regional news, sports, focus, politics, black chronicle, eTV, entertainment, lifestyle*, and *interviews*. The data crawler retrieved textual content from all nine categories, encompassing all website URLs in a time span of 4 years, i.e. all news published in the period from April 5, 2019 to April 12, 2023. For the mentioned period, 21,562 newspaper records were collected.

The collected data was stored in a comma-separated values file format, which provides a simple and widely supported structure for tabular data storage. The dataset consist of 12 columns, each representing a specific attribute of the posts. These columns include: *title, subtitle, text, author name, photographer name, URL, publication date, publication time, news category, tags,* and *number of images in news.* Dataset *eVarazdin-19-23* [3] is publicly available for use. It has been cleaned up of duplicate posts and inconsistencies in text formatting.

3.2 Methods

A short overview of methods that were applied in the crisis-related analysis was given in this section. The authors propose an approach based on three main steps: (i) exploratory analysis, (ii) topic filtering and (iii) NER enriched with network analysis.

In the first step, an exploratory analysis of the dataset in terms of giving a timeline with the number of news articles published during the observed period was performed. The number of crisis-related news articles was compared with the number of other news articles. As well, this analysis enables a comparison between the number of articles related to COVID-19 and RUW. In order to provide a more detailed analysis, the distribution of news articles classified into news categories was additionally performed.

Next, a set of the main topics related to COVID-19 and another set of the main topics related to RUW were identified. Then, news articles were filtered according to selected topics. The goal of this step is to determine the number of articles on every topic and identify the most popular topics within the crisis.

In the last step, the named entity recognition task was performed in combination with network analysis. NER is a natural language processing task that extracts information from text to identify and classify named entities into predefined categories, such as personal names, locations, organizations, dates, numeric values, monetary values, and sometimes very specific entities, depending on the domain or application in which they are used. For example, Liu et al. (2021) in their study on the monitoring of the COVID-19 pandemic on the Reddit network introduce their own categories by which they monitor specific entities related to the pandemic: distancing, disinfection, personal protective equipment, symptoms and testing.

In this research, the implementation of NER was performed using the programming language Python and spaCy (v3.6)[4], a well-known open-source software library for advanced natural language processing. A pretrained pipeline and predefined NER model for the Croatian language is used. In this experiment, a small amount of data is used for training. A small Croatian model, *hr_core_news_sm* was trained for NER on written web text (news, media), that includes vocabulary, syntax and entities. The model uses training corpus hr500k 1.0 (Ljubešić et al., 2016) which contains about 500,000 manually annotated tokens on the levels of tokenisation, sentence segmentation, morphosyntactic tagging, lemmatisation and named entities. Accuracy evaluation in terms of the F1-score for the Croatian NER pipeline is 76.1%.

Entities were extracted and divided into four categories: *persons* (PER), *locations* (LOC), *organisations* (ORG) and *miscellaneous* (MISC). Then, the top 15 most frequent entities in each category were selected and four entity networks for both crises were constructed. All networks are weighted and undirected. Nodes represent entities, and two nodes are linked if these entities appear together in at least one news article. The weight on the link represents the number of different news articles in which both entities appear.

The details of every method applied within the proposed approach are described in the next section.

4 Content Analysis and Results

In this section, details of the content analysis were provided, along with explanations of results obtained by the proposed approach.

[2]https://www.facebook.com/evarazdin.hr

[3]https://github.com/sbeliga/InfoCoV/tree/main/CECIIS2023/eVarazdin-19-23.csv

[4]https://spacy.io/models/hr

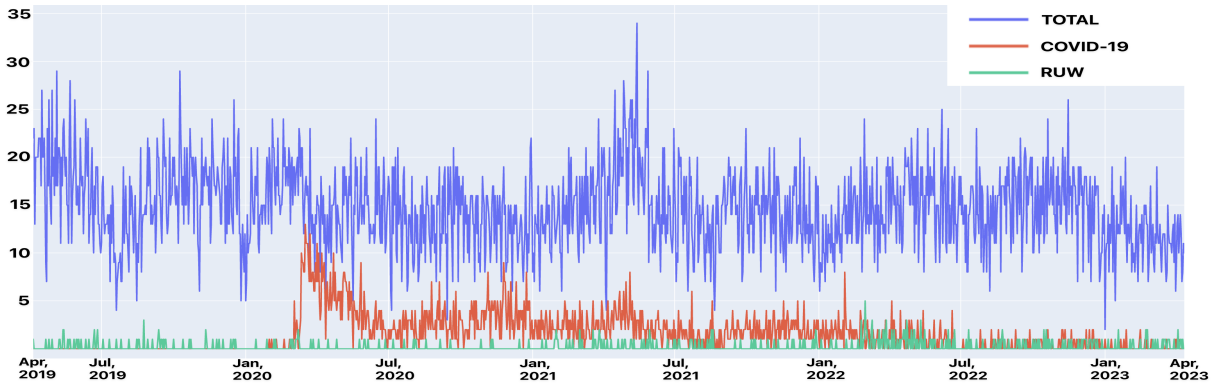


Figure 1: Dynamics of publication of newspaper articles on the *eVarazdin* news portal (blue line) in the time span from April 5, 2019 to April 12, 2023, and announcements related to the COVID-19 (red line) and RUW domain (green line).

4.1 Exploratory Analysis and Topic Filtering

In addition to the observed time period from April 5, 2019 to April 12, 2023, consideration of experiment is also limited to articles on two specific domains: COVID-19 or RUW. Before a deeper analysis of the created dataset *eVarazdin-19-23*, filtering based on keywords has been implemented. Keywords are a set of terms that best describe the subject of a document (Beliga et al., 2015). Filtering was performed using two different thesauruses [5] which were created manually and contain words that succinctly describe a particular topic discussed in the text. One with keywords from the domain of RUW (such as Ukrajina, Rusija, Zelenski, etc.), and the other from the domain of COVID-19 pandemic (such as coronavirus, COVID-19, covid, etc.). Articles that did not meet certain filtering criteria (e.g., news that is not primarily related to the domain of COVID-19 or RUW, but only marginally or incidentally mentions COVID-19 or the war) were excluded from the analysis.

It is important to emphasize that when filtering articles the morphological variations of words were taken into account. Given that the Croatian language is morphologically very rich, news were filtered by considering a truncated form of the (key)word in order to include into extraction criteria all the morphological variations that can arise from the suffixal formation of an individual (key)word.

The articles satisfying the previous condition were extracted along with their associated metadata.

Fig. 1 shows the dynamics of newspaper article publication on the *eVarzdin* portal on the timeline (from April 5, 2019 to April 12, 2023). The red line represents the newspaper publications related to the domain of COVID-19, the green line those related to the RUW domain, while the blue line represents the total number of all publications on the portal. The appearance of the SARS-CoV-2 virus in the world occurred at the end of 2019, and the first newspaper publications on the *eVarazdin* portal mentioned the coronavirus at the very end of January 2020. It was a newspaper article about the suspicion that a patient has COVID-19 and denial of the director of Varaždin Hospital because the patient had the ordinary flu. It is interesting that in the title of this publication, there is a statement "*Coronavirus is disinformation – there is no reason to worry*". The first appearance of the coronavirus in Croatia was recorded on February 25, 2020. The trend of news publications related to this topic on the graph also begins to increase immediately after that, i.e. at the beginning of 2020. After that, the red line begins to grow steeper and reaches its highest peak (March 17, 2020). This is where the real infodemic begins. Press discussions on measures for entrepreneurs, passes for passing through neighboring counties, and the Civil Protection Headquarters begins to prescribe the work regime in the time of the coronavirus crisis. The news reports that public gatherings and manifestations are being canceled due to the bad epidemiological situation, that the local hospital in Varaždin is expanding its capacities for those infected with the coronavirus, and that emergency meetings of the Civil Protection Headquarters are being held to determine measures and guidelines for action. This is followed by a flood of inscriptions about the way catering establishments and service activities work, about the closure of cafes, gyms, hair salons, shopping centers, and banks, but also about people who do not respect the prescribed self-isolation. The announcements are also related to sports topics, primarily about athletes who must report going to countries where COVID-19 is present, the postponement of competitions by UEFA, etc. A large peak on the graph is also visible in mid-March 2020, immediately after the Minister of Health declared the beginning of the COVID-19 epidemic in the Republic of Croatia, and the World Health Organization declared the previous epidemic a pandemic. Electronic media

[5] COVID-19 and Russia-Ukraine thesaurus used for main article filtering: https://github.com/sbeliga/InfoCoV/tree/main/CECIIS2023

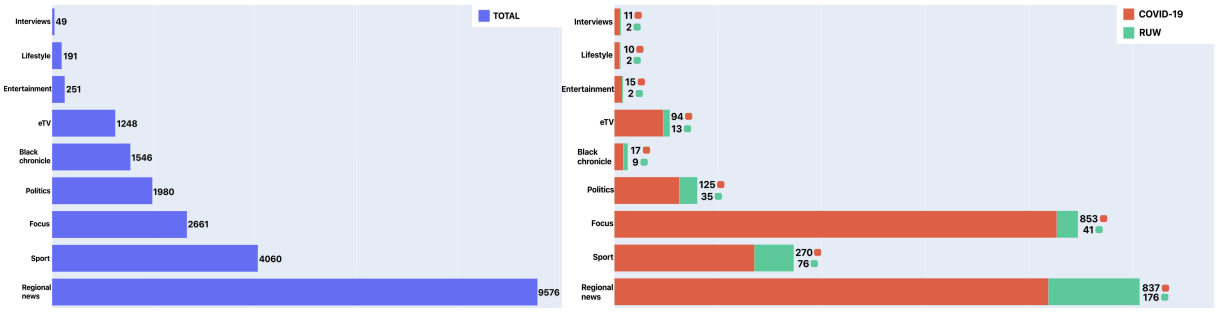


Figure 2: Total number of collected news articles in the time span from April 5, 2019 to April 12, 2023 on *eVarazdin* news portal, classified to corresponding news category (left barplot), and comparison of filtered articles (right barplot) from the COVID-19 (red bars) or RUW domain (green bars).

eVarazdin then intensively reported on the guidelines of the Varaždin General Hospital, Varaždinske Toplice also gave recommendations for the situation with the coronavirus, news were written about the mass closure of sports facilities, announcements were written about measures asking the elderly to avoid going out unnecessarily from their homes, etc.

Later, the steep fall and regrowth of the red line to a new peak occur on several occasions, which follows waves of new coronavirus infections. These results are in accordance with findings in previous research (Beliga et al., 2021). The last big peak was detected at the beginning of February 2022. There is still mass writing about the number of PCR tests and the number of infected, announcements are written about politicians and their responsibility during the pandemic, but also about the lack of blood supplies because people stopped donating their blood during the pandemic. As the virus weakens and the number of infected people decreases, the publication frequency drops slightly. The end of the epidemic in the Republic of Croatia was declared on May 11, 2023. The trend of a small number of publications on the graph is already visible from the very beginning of 2023.

The beginning of news publications about the unrest in Ukraine goes back far before the mass invasion that took place on February 24, 2022. The unrest in Ukraine has been recorded since 2014, and a fairly constant number of posts on the graph is manifested from the very beginning of our defined timeline, which is April 2019. The dynamics of publishing news related to RUW are not as steep and sudden as those related to the coronavirus pandemic. However, some minor peaks can be found. For example, the one at the beginning of January 2022, when the portal reports on the humanitarian organization Red Cross, how the city of Varaždin accepts refugees from the war-affected Ukraine, how the first refugees arrive in Varaždin spa, about accommodation centres, about sports halls where people will sleep and live, about private accommodation offered by local residents. It is also written about the need for Ukrainian language translation.

The scale of the crisis in the domain of COVID-19 and RUW, which manifests itself in the form of the publication of newspaper articles related to that topic, is stronger when it comes to COVID-19 than RUW. The explanation one can find is that Croatia, like the rest of the world, was directly involved in the pandemic. In the case of the Russian-Ukrainian war, it is not directly involved, although it is suffering from an economic crisis and strong inflation, which is partly a consequence of that conflict. Another reason for the stronger intensity of announcements related to the COVID-19 crisis is certainly the two strong earthquakes that occurred on the territory of the Republic of Croatia during the COVID-19 epidemic, which made the fight against the virus more difficult in the areas affected by the earthquake.

The statistics of the collected articles in the observed period are presented in Fig. 2. Overall, (10.35%) of the news items refer to the COVID-19 domain, while only (1.65%) refer to the RUW domain. As expected, among all categories of newspaper articles, the same three categories (*regional news, sports* and *focus*) have the largest number of published articles related to COVID-19, but also to the RUW domain. This is to be expected, as these are the most popular newspaper categories on the portal. However, if we observe them in ratios, the first three places for the COVID-19 domain are occupied by the categories *focus* (32.06%), *interviews* (22.45%) and *regional news* (8.74%). In the RUW domain, the first three places are occupied by the categories *interviews* (4.08%), *regional news* (1.84%) and *politics* (1.87%). It is interesting to note that the categories *politics* and *black chronicles* do not rank higher, as one might expect, considering that we are analyzing crisis situations. The reason could be that some news from COVID-19 or RUW fell into the category *focus* or *regional news*, considering that we are dealing with a local web portal where the type of news can be of a very local nature and therefore be placed in the mentioned categories.

In order to define more specific topics within individual domains, additional topic filtering was done.

Based on additional keywords[6] in the domains about COVID-19 and RUW, the news was filtered into four more sophisticated subtopics of the domain, which gives a more detailed insight into the specificity and importance of each topic. Topic filtering in the domain of COVID-19 was carried out for the topics of *vaccines, measures, infected* and *famous personalities*, and in the domain of RUW for *victims, locations, humanitarian* and *economy*.

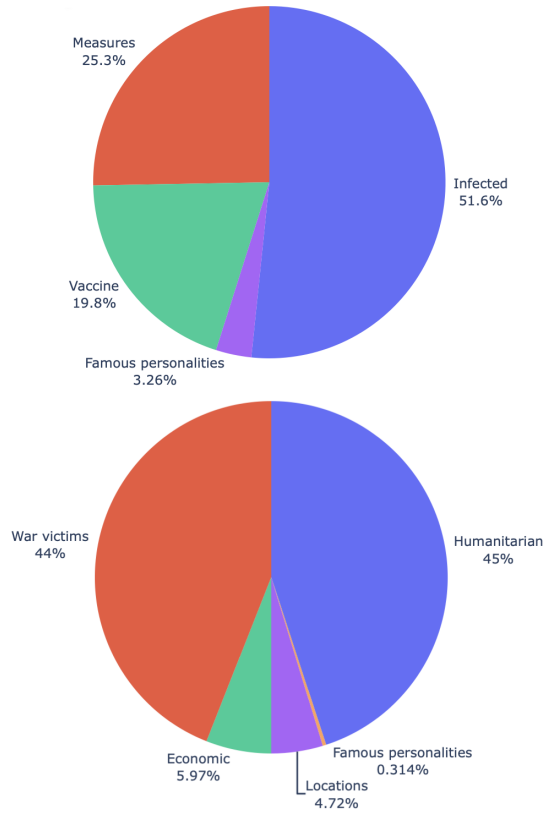


Figure 3: Comparison of the ratios of published news items with regard to the topic filtering for the COVID-19 (top) and RUW domain (bottom).

Additional topic filtering of news articles was performed automatically in such a way that only those news records that met the following criteria were taken into consideration. First, the article should meet the criteria of basic filtering with the previously mentioned filtering dictionary to become a member of the COVID-19 or RUW domain. Second, topic filtering ensures the checking of subtopics within the domain. True positive cases are those articles that meet the filtering criteria of the basic COVID-19 thesaurus, but also of individual thesauruses for a specific COVID-19 topic (e.g. for vaccines or infected subtopic). True negative cases are represented by articles that generally satisfy filtering based on the basic COVID-19 filter, but not the thesaurus of secondary filtering topics for a specific subtopic. Such records are not considered in topic filtering statistics. The same applies to the RUW domain.

The total number of news records is 2229 and 344, of which 2083 and 308 are true positives and 146 and 36 are false positives, in the COVID-19 and RUW domains, respectively. Only true positive news records are considered in further topic filtering analysis (the results of which are shown in Fig. 3).

By analyzing and filtering topics in the domain of COVID-19 (see Fig. 3 - top pie chart), it was found that the most dominant topic in the newspaper articles is related to *infected*, after which equally dominant topics are related to *measures* and *vaccines*. Articles related to the topic of *famous personalities* were published least frequently. Articles related to the topic of *famous personalities* mostly included statements or interviews from famous people such as politicians, people from a specific profession such as epidemiologists, etc. When analyzing the RUW domain (see Fig. 3 - bottom pie chart), the most dominant topic is *humanitarian* and only slightly less dominant is *victims*. About the same number of contributions refer to the topics of *economy* and *locations*, and the topic of *famous personalities* is the least frequent, as is also the case in the domain of COVID-19. It can be expected that during the pandemic, the most relevant newspaper articles are to expect statements or interviews of specialized persons who can speak professionally about the epidemic caused by the coronavirus. However, the analysis shows that the number of publications on such topics was the rarest. The source of the infodemic can be detected and confirmed precisely with the largest number of articles related to the topic *infected*, which was the most talked about and written about during the pandemic. Related to the RUW domain, perhaps one could have expected a greater number of articles written about the economic consequences of the war, but it was still not the most dominant topic. The reason for this may be that a local newspaper portal is analyzed, and not some mainstream portal. It is interesting that most of the coverage was on humanitarian topics such as who and how much donations (money, food, medicines and even weapons) were sent to the victims and the vulnerable.

4.2 NER Enriched with Network Approach

Named entities extracted from newspaper articles are *persons*, *locations*, *organizations* and *miscellaneous*. The approach used to visualize the results is enriched with concepts from the field of complex network analysis (i.e. graph theory). Nodes in a constructed graph represent entities and connections between two nodes in the graph (named entities) exist if they are mentioned in the same newspaper article. If they appear in more than one article, the connection is stronger. In other words, the number of newspaper articles in which the entities appear together is proportional to the weight of

[6]Keyword lists for additional topic filtering: https://github.com/sbeliga/InfoCoV/tree/main/CECIIS2023

the link.

In the visualisations of these NER-based networks, the links with small weights are shown with a thinner red line, while links with larger values of weights are shown with a thicker blue line. The size of a node is expressed with the value of the node's degree centrality. In other words, larger nodes are more important because they have a higher degree centrality.

Due to limited space, only two visualizations for some selected categories of named entities are presented below. The left graph in Fig. 4 illustrates the 15 most significant entities for the *organisations'* category extracted from articles that are in the COVID-19 domain. The entity "City-Municipal Council" stood out as the most significant. In addition, the "Croatian Institute for Public Health", "headquarters" and "general hospital" also stood out as very significant entities in this domain, as well as the larger cities in this region – "Varaždin" and "Ivanec". Here they are recognized as separate entities even though they do not represent organizations. However, they have strong connections with the hospital, the headquarters and the city-municipal council, as they are often mentioned right next to their names, and sometimes as an integral part of the name, for example, "General Hospital Varaždin". In this case, due to the lack of NER tools, they are recognized as separate entities. Two more groups of entities are recognized here as important. One is related to political organizations, parties such as HDZ, SDP and HNS. The second group is related to sports organizations such as HNL (Croatian football league), NK (football club) and football club Dinamo. At the very end, COVID-19 is not an organization, but it is singled out as a very important entity. This is also one of the disadvantages of the used NER tool, but it is reasonable that this entity is important for the domain we are analyzing. The advantage of the combined NER and network approach is that we can also analyze the links between entities. In this way, we can explain the strength of the relationship between individual entities. For example, we see that there are connections between health organizations and headquarters, sports organizations and cities, and, for example, strong mutual connections between political parties.

The right part of Fig. 4 illustrates the extracted entities for the *locations* category from articles that are in the RUW domain. Russia, Ukraine and Croatia are the three central entities with very strong ties, which implies that they are mentioned in the same press releases and that they are also the three most important entities in this domain. Russia is the most frequent entity. According to the visualization, in addition to the capital of the county – the city of Varaždin, other smaller towns and larger cities also appear in the announcements. These are Ludbreg, Ivanec, Novi Marof, Čakovec and Zagreb. It is often written about Europe and Australia, as well as other countries, except for the aforementioned Russia, Croatia and Ukraine. The entity "Stričak" does not represent the location. Due to the imperfection of the NER tool, it is recognized as a location, but it actually refers to a politician and the mayor of Varaždin County. Interestingly, "Municipality" entity is also mentioned, which by itself represents an incomplete location because it does not specify an exact municipality. Nevertheless, the "Municipality" node has many connections to other nodes and still belongs to one of the 15 most significant entities in this analysis.

5 Conclusion

In this study, an approach for the content analysis of crisis-related news articles in the Croatian language published in electronic media is proposed. The authors analyse and compare two domains: COVID-19 and RUW. Exploratory analysis, topic filtering and NER were performed. The proposed approach extends NER with the network analysis by constructing NER-based networks of entities that co-occur in news articles. In this way, analysis was enriched by gaining better insight into the relationships of entities involved in the crisis. Moreover, this overcomes the lack of the NER technique by noticing strong connections between entities that can form a whole, which can not be recognized by NER. This is a particularly important extension in the case of crisis-related content analysis, as the vocabulary of the language expands with new terminology and previously trained NLP-based algorithms do not perform well with the new terms.

From the results presented in the previous section, several conclusions about the content can be drawn. Firstly, the number of news articles about COVID-19 is higher than the number of news articles about RUW in all categories. This can be explained by the fact that the pandemic was present in Croatia while the war was taking place in other countries. Secondly, the most represented category is *regional news* which is to be expected in this type of electronic media. Next, the frequencies of news articles over the observed period reveal peaks that occur during crises. Finally, the network analysis of the main entities helps in the detection of most present persons/locations/organisations in the news articles about the crisis and how they are connected.

Overall, the presented results confirm that the proposed approach provides a good insight into crisis-related content for two different crises. This indicates that the same approach could be applied to the analysis of any other crisis-related content.

This is only preliminary research which the authors intend to expand into a comprehensive methodology for analysing crisis-related communication in general. The next step in their future work will be to extend the proposed approach with new methods and techniques suited for crisis-related communication analysis.

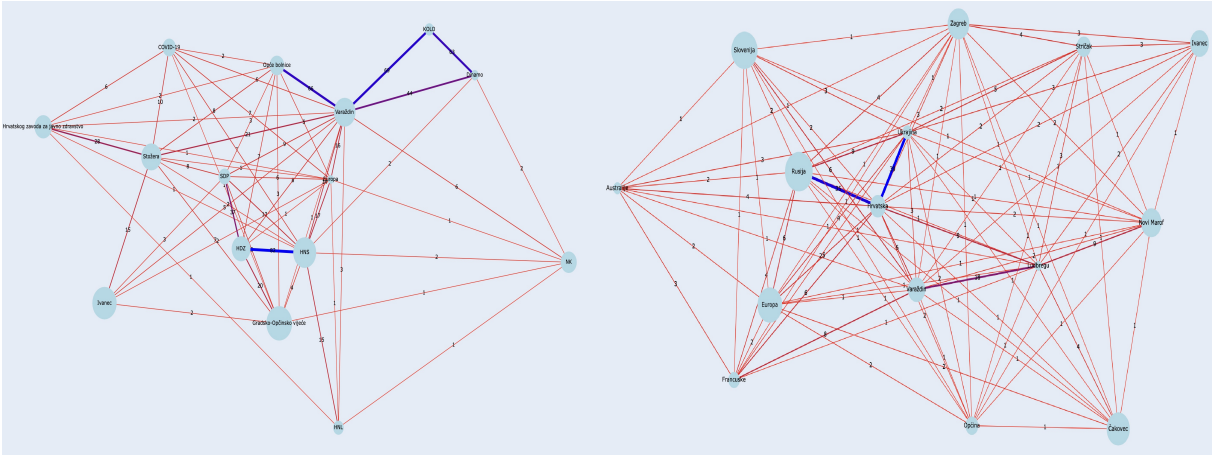


Figure 4: Networks of named entities: the left graph represents entities of class ORGANIZATIONS for press releases from the COVID-19 domain, and the right graph represents entities of class PERSONS for press releases from the RUW domain.

Acknowledgments

This work has been fully supported by the University of Rijeka projects: uniri-mladi-drustv-22-39 and uniri-drustv-18-38.

References

Alam, F., Ofli, F., & Imran, M. (2020). Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria. *Behaviour & Information Technology*, 39(3), 288-318. doi:10.1080/0144929X.2019.1610908

Babić, K., Petrović, M., Beliga, S., Martinčić-Ipšić, S., Matešić, M., & Meštrović, A. (2021a). Characterisation of COVID-19–related tweets in the Croatian language: framework based on the Cro-CoV-cseBERT model. *Applied Sciences*, 11(21), 10442. doi:10.3390/app112110442

Babić, K., Petrović, M., Beliga, S., Martinčić-Ipšić, S., Jarynowski, A., & Meštrović, A. (2021b). Covid-19–related Communication on Twitter: Analysis of the Croatian and Polish Attitudes. In *Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, Volume 3* (pp. 379-390). Singapore: Springer. doi:10.1007/978-981-16-1781-2_35

Beliga, S., Meštrović, A., & Martinčić-Ipšić, S., (2015). An overview of graph-based keyword extraction methods and approaches. *Journal of information and organizational sciences*, 39(1), 1-20.

Beliga, S., Martinčić-Ipšić, S., Matešić, M., Petrijevčanin Vuksanović, I., & Meštrović, A. (2021). Infoveillance of the Croatian Online Media During the COVID-19 Pandemic: One-Year Longitudinal Study Using Natural Language Processing. *JMIR public health and surveillance*, 7(12), e31540. doi:10.2196/31540

Beliga, S., Martinčić-Ipšić, S., Matešić, M., & Meštrović, A. (2022). Natural language processing and statistic: The first six months of the COVID-19 infodemic in Croatia. *The Covid-19 Pandemic as a Challenge for Media and Communication Studies*, (pp. 78-92) Routledge: London, Delhi. doi:10.4324/9781003232049-9

Bogović, P. K., Meštrović, A., Beliga, S., & Martinčić-Ipšić, S. (2021). Topic modelling of Croatian news during COVID-19 pandemic. In *Proceedings of 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 1044-1051). IEEE. doi:10.23919/MIPRO52101.2021.9597125

Boon-Itt, S., & Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. textitJMIR Public Health and Surveillance, 6(4), e21978.

Chen, B., Wang, X., Zhang, W., Chen, T., Sun, C., Wang, Z., Wang, F.-Y. (2022). Public Opinion Dynamics in Cyberspace on Russia-Ukraine War: A Case Analysis With Chinese Weibo, In *IEEE Transactions on Computational Social Systems*, 9(3), pp. 948-958, doi: 10.1109/TCSS.2022.3169332

De Santis, E., Martino, A., Ronci, F., & Rizzi, A. (2023). An Unsupervised Graph-Based Approach for Detecting Relevant Topics: A Case Study on the Italian Twitter Cohort during the Russia-Ukraine Conflict. *Information*, 14(6), 330. doi:10.3390/info14060330

Eligüzel, N., Çetinkaya, C., & Dereli, T. (2022). Application of Named Entity Recognition on Tweets During Earthquake Disaster: a Deep Learning-Based Approach. *Soft Computing* 26(1), 395-421. https://doi.org/10.1007/s00500-021-06370-4

Eysenbach, G. (2020). How to fight an infodemic: the four pillars of infodemic management. *Journal of medical Internet research*, 22(6), e21820.

Eysenbach, G. (2002). Infodemiology: The epidemiology of (mis) information. *The American journal of medicine*, 113(9), 763-765.

Eysenbach, G. (2009). Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of medical Internet research*, 11(1), e1157.

Fu, S., Lyu, H., Wang, Z., Hao, X., & Zhang, C. (2022). Extracting historical flood locations from news media data by the named entity recognition (NER) model to assess urban flood susceptibility. *Journal of Hydrology*, 612(C), 128312, https://doi.org/10.1016/j.jhydrol.2022.128312

Garcia, G., & Ladeira, M. (2021). A proposal for the supplementation of traditional Post-market Surveillance systems based on Named Entity Recognition, In *Proc. of 16th Iberian Conference on Information Systems and Technologies (CISTI)*, Chaves, Portugal, (pp. 1-6), doi:10.23919/CISTI52073.2021.9476560.

Guerra, A. & Karakuş, O. (2023). Sentiment analysis for measuring hope and fear from Reddit posts during the 2022 Russo-Ukrainian conflict. *Front Artif Intell.* 5;6:1163577. doi:10.3389/frai.2023.1163577

Ilić, A., & Beliga, S. (2021). The Polarity of Croatian Online News Related to COVID-19: A First Insight. In *Proceedings of 32nd Central European Conference on Information and Intelligent Systems-CECIIS 2021* (pp. 237-246). Faculty of Organization and Informatics, University of Zagreb.

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR)*, 47(4), 1-38.

Liu, Y., Whitfield, C., Zhang, T., Hauser, A., Reynolds, T., & Anwar, M. (2021) Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. *Health Inf Sci Syst*, 9(25). https://doi.org/10.1007/s13755-021-00158-4

Ljubešić, N., Klubička, F., Agić, Ž., & Jazbec, I.-P. (2016). New Inflectional Lexicons and Training Corpora for Improved Morphosyntactic Annotation of Croatian and Serbian, In *Proceedings of the Tenth International Conference on LLREC 2016*, ELRA, (pp. 4264-4270).

Meštrović, A., Petrović, M., & Beliga, S. (2022). Retweet Prediction Based on Heterogeneous Data Sources: The Combination of Text and Multilayer Network Features. *Applied Sciences*, 12(21), 11216. https://doi.org/10.3390/app122111216

Ngo, V.M., Huynh, T.L.D., Nguyen, P.V. & Nguyen, H.H. (2022). Public sentiment towards economic sanctions in the Russia-Ukraine war. *Scottish Journal of Political Econom*, 69, 564-573. https://doi.org/10.1111/sjpe.12331

Nemes, L., & Kiss, A. (2021). Information Extraction and Named Entity Recognition Supported Social Media Sentiment Analysis during the COVID-19 Pandemic. *Appl. Sci.*, 11(11017). https://doi.org/10.3390/app112211017

Pandur, M. B., Dobša, J., Beliga, S., & Meštrović, A. (2021). Topic modelling and sentiment analysis of COVID-19 related news on Croatian Internet portal. In *Proceeding of Conference on Data Mining and Data Warehouses 2021*, Slovenian KDD Conference on Data Mining and Data Warehouses (SiKDD 2022), Ljubljana, Slovenia.

Phopli, W., Maliyaem, M., Haruechaiyasak, C. & Ketmaneechairat, H. (2021). Microblog Entity Detection for Natural Disaster Management. *Journal of Advances in Information Technology*, 12(4), 351-356. doi:10.12720/jait.12.4.351-356

Richardson, L. (2023). Beautiful soup documentation. Retrieved from https://www.crummy.com/software/BeautifulSoup/ (1 June 2023)

Truong, T. H., Dao, M.H., & Nguyen, D. Q. (2021). COVID-19 Named Entity Recognition for Vietnamese. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp.2146-2153). ACM. https://aclanthology.org/2021.naacl-main.173

Wadhwani, G.K., Varshney, P.K., Gupta, A., & Kumar, S. (2023). Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia-Ukraine War. *SN COMPUT. SCI.* 4, 346. https://doi.org/10.1007/s42979-023-01790-5

Quality in ICT

Miro Zdilar
Towards Automated Detection of Qualitative Spreadsheet Errors in Multi-user Environments

Towards Automated Detection of Qualitative Spreadsheet Errors in Multi-user Environments

Miro Zdilar
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
miro.zdilar@gmail.com

Abstract. *Spreadsheets are one of the most used software systems in business and academia. Since the first introduction of electronic spreadsheets for personal computers in 1979, they have evolved into a powerful computing platform. Their use spans from asset administration to complex scientific analysis and process modelling. However, spreadsheets are associated with high incidence of errors, causing companies and organizations significant reputational and financial losses. The goal of this work-in-progress article is to present a model for an automated detection of qualitative spreadsheet errors in multi-user environments, based on abstract state machines.*

Keywords. Spreadsheets, Spreadsheet Errors, Automated Error Detection, Qualitative Spreadsheet Errors in Multi-User Environments, Spreadsheet Quality Assurance

1 Introduction

Spreadsheets are widely used and can be considered as one of the most successful end-user programming systems. End-user programming systems allow end-users to build and execute powerful computer programs without the use of traditional programming languages and supporting development tools. It has been estimated that the number of end-user programmers outnumber traditional software programmers (Scaffidi et al., 2005). Spreadsheets are used in almost all companies in the US and Europe for financial reporting and strategic planning (Panko & Ordway, 2008). The framework for the identification of spreadsheet usage patterns (Reschenhofer & Matthes, 2015) identified use cases for the following business processes; capacity planning, financial reporting, stakeholder analysis, risk management, performance calculation, data transformation, cash-flows analysis, time-series transformations, and simulations. Despite their great success and importance, spreadsheets are known to be error prone (Rajalingham et al., 2000). The European Spreadsheet Risk Interest Group (EuSpRIG), a non-profit and voluntary organization, maintains a list of horror stories that illustrate problems with uncontrolled usage of spreadsheets (European Spreadsheet Risk Interest Group, 2023).

This work-in-progress article presents a model for automated detection of qualitative spreadsheet errors in multi-user environments. According to taxonomy of spreadsheet errors (Rajalingham et al., 2008), qualitative errors do not immediately produce incorrect values but degrade the quality of the spreadsheet. Spreadsheet also becomes more prone to misinterpretation and more difficult to maintain. The rest of this paper is organized as follows. Section 2 is a short overview of related work with the problem statement in section 3. Afterwards, in section 4, the architecture of the proposed model for automated detection of qualitative spreadsheet errors in multi-user environments is presented. In section 5 we provide initial results of presented model simulations. Finally, in section 6 we discuss results and provide directions for further model evaluation and future research.

2 Related Work

2.1 Taxonomy of Spreadsheet Errors

Omnipresence of spreadsheets and impact of spreadsheet errors triggered a significant interest of the research community. Focus of researchers have been in the area of detection of spreadsheet errors, as well as identification of frameworks and methodologies that should prevent occurrences of spreadsheet errors (Powell et al., 2008). An important aspect of spreadsheet research deals with the development of taxonomies for spreadsheet errors.

Early studies listed types of errors detected without a classification of spreadsheet errors (Brown & Gould, 1987). Galetta et al. (1993) introduced two classes of spreadsheet errors. The authors distinguished between domain errors and device errors. Domain refers to the spreadsheet's application area (e.g., accounting), while

device refers to the spreadsheet technology itself. Authors conducted an experiment with 30 accounting experts and 30 students to seek up to two errors introduced in each of six provided spreadsheets used during the experiment. While accounting experts performed better in detection of domain errors, students demonstrated comparable performance in detection of device errors. For example, a mistake in logic due to a misunderstanding of depreciation is a domain error but entering the wrong reference in the depreciation function SLN is a device error.

In one of the first attempts to offer complete classification of errors, researchers distinguished between quantitative and qualitative errors (Panko & Halverson, 1996). Quantitative errors are related to the current version of a spreadsheet, while qualitative errors refer to risky practices that might lead to an error in later stages of a spreadsheet's lifecycle and degrade the quality of the spreadsheet. Authors divided quantitative errors into three categories:

- Mechanical errors, due to mistakes in typing or pointing
- Logic errors, due to choosing the wrong function or creating the wrong formula
- Omission errors, due to misinterpreting the situation to be modelled

A more comprehensive approach to taxonomy of spreadsheet errors is focused on user-generated errors (Rajalingham et al., 2008). This taxonomy explicitly distinguished between developer and end-user generated spreadsheet errors. End-user errors are further classified into Data Inputter and Interpreter errors.

In recent years, researchers identified the need to relate types and occurrences of spreadsheet errors with quality of spreadsheets. Intuitively, a higher incidence of spreadsheet errors suggests that the overall quality of a spreadsheet is low. O'Beirne (2008) presented an overview of Information Quality and Data Quality within the context of spreadsheets. The author presented a comprehensive list of information quality attributes in the context of spreadsheet programs. In addition, the author presented checks and control procedures for spreadsheet information and quality processes. Cunha at al. (2012) proposed a quality model for spreadsheets with a set of domain specific metrics for spreadsheets, which are used to measure concrete spreadsheet characteristics. The presented quality model for spreadsheets is based on the widely accepted ISO/IEC 9126 international standard for software product quality (International Organization for Standardization, 2001). Authors provided a comprehensive analysis of ISO/IEC 9126 standard and map relevant quality attributes to spreadsheets.

2.2 Spreadsheet Errors Detection

High incidents of spreadsheet errors have led to a series of research and commercial auditing tools. Nixon & O'Hara (2010) provided a structured assessment of several commercial auditing tools. The authors conducted an experiment based on real spreadsheets used for generating turnover reports in a retail company. In total, 17 qualitative and quantitative errors have been added to the reporting spreadsheet used for testing 4 commercial spreadsheet auditing tools and Microsoft Excel built-in auditing functions. Results of the test proved that spreadsheet auditing tools are generally useful as supporting tools for spreadsheet experts. One commercial auditing tools achieved detection rates of over 80%. However, results were largely subjective and limited to only one type of a relatively simple spreadsheet used for financial reporting. Interesting to note is that Microsoft Excel built-in auditing functions only scored 24%.

Another study conducted by Aurigemma & Panko (2010) compared error detection success rate of human expert auditors with two commercial spreadsheet static analysis tools. The corpus of 75 spreadsheets used in the study was created by undergraduate students in an introductory management information system course (Panko, 2010). The overall results indicated that human spreadsheet auditors significantly outperformed automated static analysis tools in the detection of spreadsheets errors generated by humans.

Several commercial and governmental institutions published their own spreadsheet auditing tools and methodologies. Officers of HM Customs and Excise have been performing field audits of taxpayers' spreadsheet applications since 1985. Effectiveness of HM Customs and Excise spreadsheet testing methodology has been evaluated by Butler (2000). This procedure is hybrid and includes manual activities performed by an expert auditor, as well as automated activities performed with commercial software auditing tool. The procedure is based on the identification of high-risk cells and formulas with potentially high financial impact and tax losses. The main contribution of commercial auditing tools is with identification of high risk and complex formulas with many dependencies to linked cells and other data sources. Final assessment on identified high risk cells and formulas are performed by expert human auditors using different testing strategies such as control checks, reviews and reperformance. This methodology is effective for smaller to medium sized spreadsheet models used for tax calculations and reporting.

An automated method to infer data types from a spreadsheet was presented by Erwig & Burnett (2002). The proposed method for inferring types from spreadsheets is based on the concrete notion of units instead of the abstract concept of types. Authors used header information given by spreadsheets to derive units. In continuation to the presented concept around units, Ahmad et al. (2003) developed a type system for statically detecting spreadsheet errors. Author named

the proposed type system for spreadsheets "unit checking". The same researchers also presented a collection of rules for the identification of spreadsheet weaknesses that are likely to be errors. The presented type system for spreadsheets also relies on the concept of the header that defines common unit for groups of cells.

Jannach et al. (2014) conducted comprehensive literature review and proposed two main categories of automated spreadsheet quality assurance approaches:

- "Finding and fixing errors" is about techniques and tools that are mainly designed to help the user detect errors and understand the reasons for the errors.
- "Avoiding Errors" is about techniques and tools that should help developer create error free spreadsheets.

In addition to research related to manual and automated spreadsheet quality assurance, important to note is research related to the automation of spreadsheet testing (Abraham & Erwig, 2008). Modern approaches to spreadsheet development follow good practices and methodologies for software development life cycle and the automation of spreadsheet testing is valuable both for researchers and industry practitioners. To minimise bias and ensure validity of test cases, the authors developed mutation operators for spreadsheets. Authors followed original concepts of mutation testing developed for general purpose programming languages.

Recent research and studies are focused on the application of large language models to spreadsheet environments. A team of researchers from Microsoft Corporation presented the FLAME language model for spreadsheet formulas (Joshi et al., 2023). FLAME uses the Microsoft Excel specific formula tokenizer and other techniques to achieve competitive performance with a substantially smaller model (60 million parameters) and two order of magnitude less training data, compared to other large language models such as Codex. Researchers used a training dataset of 972 million formulas extracted from a corpus of 1,8 million Excel workbooks. FLAME was evaluated on three different tasks for Excel formulas: last-mile repair, autocompletion and syntax reconstruction. Syntax reconstruction is a novel term coined by the authors of FLAME, and the goal of this task is to reconstruct original Excel formula from formulas with removed delimiters. The presented FLAME language model outperforms larger language models, such as Codex-Davinci (175 billion parameters), Codex-Cushman (12 billion parameters), and CodeT5 (220 million parameters), in 6 out of 10 experimental settings.

3 Problem Statement

Our proposed model for automated detection of qualitative spreadsheet errors in multi-user environments will address the following spreadsheet errors:

- Noncompliance and deviations from spreadsheet development guidelines (Esch et al. 2010).
- Unauthorized changes and modifications to spreadsheet programs and data during all stages of a spreadsheet's lifecycle.

Proposed model for automated detection of qualitative spreadsheet errors in multi-user environments, falls into both categories of automated spreadsheet quality assurance approaches developed by Jannach et al. (2014). It allows users to find qualitative spreadsheet errors as well assists spreadsheet developers in creation of error free spreadsheets.

Qualitative spreadsheet errors are difficult to identify and troubleshoot in multi-user environments due to frequent organizational changes in users' authorizations and complex development standards with codified rules for spreadsheet development. In the following we list a few practical examples of qualitative spreadsheet errors according to guidelines for the development and validation of spreadsheets (Esch et al. 2010):

- String literals has been entered for nominal volume in input worksheet. Guidelines for the development require that only numerical values are allowed for nominal volume.
- Inventory number of the asset is not visible on screen and on the print-out. Guidelines for the development require that inventory asset number is visible on screen and on print-out.
- Additional sample numbers have been added to input worksheet and validation range was overwritten. Guidelines for the development require that only limited number of sample numbers are allowed in input worksheet according to predefined range of input values.
- Calculation for sample standard deviation was changed by unauthorized spreadsheet end user. Guidelines for the development require that changes to calculations are performed only by authorized users.

4 Model Architecture

Bellow we provide a conceptual model for automated detection of qualitative spreadsheet errors in multi-user environments based on the application of Abstract State Machines (ASM) (Gurevich, 2000). The following notation is used to describe data flow and interaction between model components and spreadsheet users:

- S_i: State of spreadsheet at time i
- A_{Si}: State of ASM at time i
- B_{U_n}: Behaviours B of user U_n

- $\Delta S_i(B_{U_n})$: Spreadsheet state transitions between state S_i and state S_{i-1}, caused by behaviours B of user U_n
- S_e: Final state of spreadsheet (end of spreadsheet lifecycle)

The core idea for the proposed model for automated detection of spreadsheet errors is based on equivalence between different lifecycle states of a spreadsheet program and the corresponding automatically synthesized ASM. As depicted visually in Figure 1, the state of spreadsheet S corresponds to an equivalent state of ASM A_S. During the lifecycle, a spreadsheet might transition through m transitions to reach state S_i. The corresponding ASM should follow state transitions of spreadsheet S_i to reach state A_{Si}.

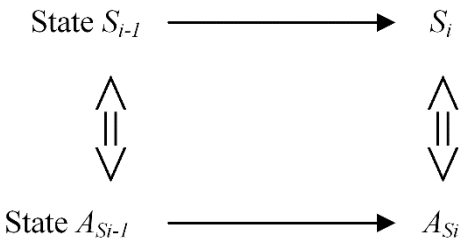


Figure 1. Spreadsheet and ASM state equivalence

Using the same notation, spreadsheet stages during different phases of the lifecycle are modelled by a finite sequence of state transitions as follows:

$$S_0 \xrightarrow{\Delta S_0(B_{U_n})} S_1 \xrightarrow{\Delta S_1(B_{U_n})} S_2 \xrightarrow{\Delta S_2(B_{U_n})} \dots S_e \quad (1)$$

Where S_0 is the initial state of the spreadsheet ("first creation") and S_e is the final state of spreadsheet ("end of lifecycle"). $\Delta S_i(B_{U_n})$ are transitions between spreadsheet states caused by behaviours B of user U_n. When required by the spreadsheet user, at each state of S_i where $i \in \{0,1,2,..,e\}$, the equivalent state A_{Si} for ASM is synthesized in order to determine $\Delta S_{i-1}(B_{U_n})$.

Figure 2 illustrates a high-level component diagram and interaction of users for presented conceptual model.

Data processing within proposed model for automated detection of qualitative spreadsheet errors is performed in three consecutive steps. During first step, spreadsheet submitted by user as input to proposed model is parsed with dedicated parser structured around spreadsheet formula language grammar. We implemented a spreadsheet parser in Python 3 programming language (Van Rossum & Drake, 2009). We utilized the OpenPyXL library for manipulating spreadsheet files and the Tokenizer module for tokenizing spreadsheet formulae (Zumstein, 2021). As a result of successful spreadsheet parsing, a directed graph is generated with hierarchical representation of all resources that constitute the input spreadsheet. During second step, based on graph representation of spreadsheet, abstract state machine is generated. Structure and properties of graph components in synthesized abstract state machine are determined with constitutive elements of spreadsheet programs parsed during the first step of data processing. During the third step of data processing within proposed model, synthesized abstract state machine are explored to detect structures that correspond to spreadsheet errors. We utilized the NetworkX python library (Hagberg et al. 2008) to generate and query a directed graph with the representation of the abstract state machine and associated transition states.

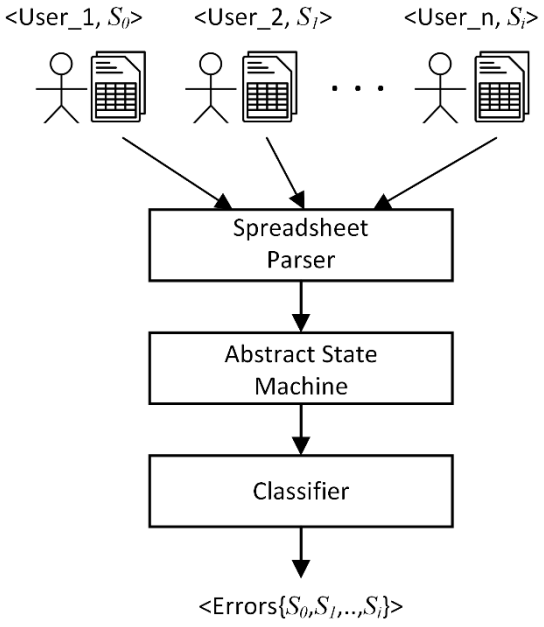


Figure 2. High-level conceptual model

5 Model Evaluation

As this research is work-in-progress, we explore following initial research question:

- RQ: How does the proposed model perform in detection of unauthorized spreadsheet changes performed by spreadsheet users?

To evaluate research question, we randomly assigned different user roles to 15 experienced spreadsheet users. Details of designed user roles and permissions are listed in Table 1.

We created an initial error free spreadsheet template with a sales report for 10 markets and 5 products. Input data was presented in the INPUT worksheet. A model to calculate total sales per product and per market was developed in the MODEL worksheet. Results were presented in the OUTPUT worksheet with references to the calculated model. The TESTING worksheet had a simple testing procedure and empty named ranges to assist the Tester with documenting the performed testing. This initial spreadsheet template was used to generate an

equivalent abstract state machine representation with proposed model for automated detection of qualitative spreadsheet errors.

Table 1. Assigned spreadsheet user roles and permissions

User role	Permissions
Developer	1. Create and modify all resources in MODEL worksheet. 2. Utilizes all data from INPUT worksheets. 3. Create and format results in OUTPUT worksheet.
Tester	1. Create and modify all resources in TESTING worksheet.
Data Inputter	1. Modify cell values in INPUT worksheet.
Owner	1. Modify cell values in INPUT worksheet. 2. Modify named ranges in all worksheets.

Without informing the 15 expert spreadsheet users about their roles, we requested them to document and perform 7 different changes in the given spreadsheet template that corresponds to permissions listed in Table 1. Based on spreadsheets provided by expert participants, we generated corresponding abstract state machines with our proposed model. In order to identify changes performed by users we queried all edges of the synthesized abstract state machines. Edges in graph representation of abstract state machines represents state transitions and, in our model, corresponds to changes and modifications to spreadsheets performed by users. We compared results of graph queries and detected graph changes with assigned users' roles and list of actual changes performed by spreadsheet users.

Our model correctly identified all changes and modifications performed by spreadsheet users. Dictionaries generated as the result of query with NetworkX python package were complex and difficult to analyse, and we utilized a simple script to transform them into tabular format. In addition, we manually reviewed all query result in tabular format with participating spreadsheet experts. Even though format of generated queries is difficult to interpret by human reviewers, our model and corresponding abstract state machines correctly detected all changes introduced by spreadsheet users.

6 Conclusion and Future Research

In this paper we have presented initial evaluations of our model for automated detection of qualitative spreadsheet errors. We developed the presented model with the goal to address specific classes of spreadsheet qualitative errors that are difficult to detect in multi-user environments. As this research is work-in-progress, our initial evaluations are limited to a smaller population of spreadsheets and users. However, presented results are helpful and provide valuable insights to model capabilities.

In our future research, we will focus on performance improvement of spreadsheet parsing and data structures for abstract state machines representation. We will further evaluate proposed model in more comprehensive case studies and large multi-user environments. We will also explore opportunities to implement machine learning algorithms for the detection of spreadsheet qualitative errors in multi-user environments.

Acknowledgments

The author would like to thank Prof. dr. sc. Markus Schatten from Artificial Intelligence Laboratory, for valuable comments and suggestions.

References

Abraham, R., & Erwig, M. (2008). Mutation operators for spreadsheets. *IEEE Transactions on Software Engineering, 35*(1), 94-108.

Ahmad, Y., Antoniu, T., Goldwater, S., & Krishnamurthi, S. (2003). A type system for statically detecting spreadsheet errors. *In 18th IEEE International Conference on Automated Software Engineering, 2003. Proceedings.* (pp. 174-183). IEEE.

Aurigemma, S., & Panko, R. R. (2010). The detection of human spreadsheet errors by humans versus inspection (auditing) software. *arXiv preprint arXiv:1009.2785*.

Brown, P.S. and Gould, J.D., (1987). An experimental study of people creating spreadsheets. *ACM Transactions on Information Systems (TOIS), 5*(3), pp.258-272.

Butler, R. J. (2000). Is this spreadsheet a tax evader? How HM Customs and Excise test spreadsheet applications. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences* (pp. 6-pp). IEEE.

Cunha, J., Fernandes, J. P., Peixoto, C., & Saraiva, J. (2012). A quality model for spreadsheets. In *2012 Eighth International Conference on the Quality of Information and Communications Technology* (pp. 231-236). IEEE.

Erwig, M., & Burnett, M. (2002). Adding apples and oranges. In *PADL* (Vol. 2, pp. 173-191).

Esch, P. M., Moor, C., Schmid, B., Albertini, S., Hassler, S., Donzé, G., & Saxer, H. P. (2010).

Good Laboratory Practice (GLP)–Guidelines for the Development and Validation of Spreadsheets. *The Quality Assurance Journal, 13*(3-4), 41-56.

European Spreadsheet Risk Interest Group. (2023). EuSpRIG Horror Stories. https://eusprig.org/research-info/horror-stories/ (May 30, 2023)

Galletta, D. F., Abraham, D., El Louadi, M., Lekse, W., Pollalis, Y. A., & Sampler, J. L. (1993). An empirical study of spreadsheet error-finding performance. *Accounting, Management and Information Technologies*, *3*(2), 79-95.

Gurevich, Y. (2000). Sequential abstract state machines capture sequential algorithms, In *ACM Transactions on Computational Logic* (Vol. 1, No. 1, pp.77–111). ACM.

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX* (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

International Organization for Standardization. (2001). *Software engineering-product quality-part 1: Quality model* (ISO Standard No. ISO/IEC 9126-1).

Jannach, D., Schmitz, T., Hofer, B., & Wotawa, F. (2014). Avoiding, finding and fixing spreadsheet errors–a survey of automated approaches for spreadsheet qa. Journal of Systems and Software, 94, 129-150.

Joshi, H., Ebenezer, A., Cambronero, J., Gulwani, S., Kanade, A., Le, V., ... & Verbruggen, G. (2023). FLAME: A small language model for spreadsheet formulas. *arXiv preprint arXiv:2301.13779*.

Nixon, D., & O'Hara, M. (2010). Spreadsheet auditing software. *arXiv preprint arXiv:1001.4293*.

O'Beirne, P. (2008). Information and data quality in spreadsheets. *arXiv preprint arXiv:0809.3609*.

Panko, R. R. (2000). Two corpuses of spreadsheet errors. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences* (pp. 8-pp). IEEE.

Panko, R. R., & Halverson, R. P. (1996). Spreadsheets on trial: A survey of research on spreadsheet risks. In *Proceedings of HICSS-29: 29th Hawaii International Conference on System Sciences* (Vol. 2, pp. 326-335). IEEE.

Panko, R. R., & Ordway, N. (2008). Sarbanes-oxley: What about all the spreadsheets*?*. *arXiv preprint arXiv:0804.0797*.

Powell, S. G., Baker, K. R., & Lawson, B. (2008). A critical review of the literature on spreadsheet errors. *Decision Support Systems, 46*(1), 128-138.

Rajalingham, K., Chadwick, D. R., & Knight, B. (2008). Classification of spreadsheet errors. *arXiv preprint arXiv:0805.4224*.

Rajalingham, K., Chadwick, D., Knight, B., & Edwards, D. (2000). Quality control in spreadsheets: a software engineering-based approach to spreadsheet development. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (pp. 10-pp). IEEE*.

Reschenhofer, T., & Matthes, F. (2015). A Framework for the Identification of Spreadsheet Usage Patterns. In *ECIS*.

Scaffidi C., Shaw M., Myers B. A. (2005). Estimating the numbers of end users and end users programmers, In *Proc. of VL/HCC '05, pp. 207-214*.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace.

Zumstein, F. (2021). *Python for Excel: A Modern Environment for Automation and Data Analysis* (pp. 155-179). O'Reilly Media.

Strategic Planning and Decision Making

Hannia Gonzalez-Urango, Enrique Mu, Anna Florek-Paszkowska, Milagros Pereyra-Rojas
Validation and use of a framework to assess challenges to virtual education in the context of emergency remote teaching: Peru and Spain

Sam Sabah, Imran Hossain, David Weiss, Alexander Tillmann
Towards Increasing Active Citizen Involvement in Urban Planning through Mixed Reality Technologies

Mehtap Dursun, Rana Duygu Alkurt
Evaluating the Net Zero Policy Performances of European Countries Through DEA

Nikola Zornić, Aleksandar Marković, Nebojša Avramović
A Bibliometric Review of Agent-Based Models Application in Economics

Validation and use of a framework to assess challenges to virtual education in the context of emergency remote teaching: Peru and Spain

Hannia Gonzalez-Urango
University of Pittsburgh
Latin American Studies Association
4338 Bigelow Blvd, Pittsburgh, PA 15213
Universitat Politècnica de València
INGENIO (CSIC-UPV)
Camino de Vera, 46022, Valencia, Spain
hkgonzal@upvnet.upv.es

Enrique Mu
Carlow University
Department of Business Management, Accounting and Ethics
3333 Fifth Avenue, Pittsburgh, PA 15213. USA
emu@carlow.edu

Anna Florek-Paszkowska
CENTRUM Católica Graduate Business School
Pontificia Universidad Católica del Perú (PUCP)
Jirón Daniel Alomía Robles 125, Surco, Lima-Perú
aflorekpaszkowska@pucp.edu.pe

Milagros Pereyra-Rojas
University of Pittsburgh
Latin American Studies Association
4338 Bigelow Blvd, Pittsburgh, PA 15213
milagros@pitt.edu

Abstract. *In this study, the validity and use of a recently developed evaluation Framework to Assess Challenges To Virtual Education (FACVE) were tested and refined using a comparative perspective between students from Peru and Spain. The findings provide a limited endorsement for the validity of the assessment framework while also highlighting interesting similarities and differences between the students from both countries in terms of the challenges faced in the context of virtual education. This study demonstrates that the FACVE is a valid assessment instrument and can be used in any country or institutional context by selecting the relevant dimensions and sub-dimensions of interest.*

Keywords. Virtual education challenges, FACVE, emergency remote teaching, e-learning readiness.

1 Introduction

The outbreak of COVID-19 and the resulting pandemic presented a significant obstacle to education at all levels. A staggering 168 million schoolchildren across 14 nations were affected by school closures (Unicef, 2021). Higher education, like other levels of education, required an adaptation and reconfiguration of its traditional teaching methodology. Teaching suffered a forced transformation process to an online model, which involved the acquisition of certain competencies and characteristics specific to online education (Area Moreira et al., 2020), as well as some challenges arising from the transition.

This forced adaptation from a face-to-face model to a distance model has been called emergency distance education by some authors (Adell, 2020) or Emergency Remote Teaching (ERT), which, in contrast with long-planned online experiences, is an abrupt and temporary shift of instructional delivery to an alternate delivery mode due to crisis circumstances (Charles Hodges et al., 2020; Whittle et al., 2020). Both faculty and students were forced to change their traditional practices having to respond to the same assessment models (Ligardo-Herrera et al., 2022).

The transformations and adjustments experienced in university education have yielded valuable insights for virtual education. This paper delves into the analysis and solution of the impediments to virtual education in the ERT framework. The focus is particularly on individuals who have traditionally pursued face-to-face education and were compelled to shift to virtual learning in two distinct countries.

2 Theoretical Background

An evaluation framework was recently proposed to assess challenges to virtual education in the case of ERT (Mu et al., 2022). It was developed in the context of a pandemic in a specific developing country (Peru) and from a student perspective. What makes this framework unique is that it was developed from the

students' perceptions of challenges faced in the ERT context using a mixed qualitative and quantitative research approach. Various frameworks regarding virtual education have been discovered in the literature while FACVE is the first to deal with ERT contexts (Elahi et al., 2023; Mukasheva et al., 2023; Tiemann & Annaggar, 2023; van den Beemt et al., 2023).

In this study, the validity and use of this assessment framework in other contexts were explored and refined using a comparative perspective. The virtual education challenge assessment framework (FACVE) is structured in a way that it can be used in any country or regional context simply by selecting the relevant dimensions and sub-dimensions of interest. The present case study was developed by surveying students in Peru and Spain using relevant criteria. Those indicators directly related to the pandemic were not considered, given that the data collection for the present study took place a considerable time after the pandemic was over. The goal was first, to confirm the validity of the proposed evaluation framework and second, to use this framework to compare the original results from institutions in Peru with those in Spain; two countries with different levels of socio-economic development. Hence, the aim of the present study was to test and refine a framework to assess the challenges of virtual education in the context of ERT. This study provides the first confirmation of the validity and convenience of using the proposed *Framework for the Assessment of Challenges in Virtual Education (FACVE)* as well as highlights the unexpected differences in virtual education challenges based on different economic contexts. This research is significant as it is the first to confirm the validity of the FACVE and to be tested in a real educational context, adding to the growing body of the existing literature on virtual education (Bearman et al., 2023; Mou, 2023; Wong, 2023).

3 Methodology

The study was developed using a quantitative approach. Students from Peru and Spain were surveyed using the recently developed FACVE (Mu et al., 2022).

The target institution in Peru is a private leading higher education institution. The participants were all graduate business students taking courses at the time of the study. At the beginning of the pandemic, the University was closed to facilitate the reorganization to remote work. Previously, virtual teaching was not used or was used very little (work was always performed in person on the University campus). The university closed while their faculty was trained in the use of technology to teach in virtual environments. Furthermore, some more technically savvy instructors were used as additional trainers and advisors in virtual teaching for the least technology-oriented teachers.

The data in Spain was collected from two institutions. Both institutions are public universities, one oriented toward social sciences and the other toward engineering. As a result of the health emergency caused by COVID-19, in March 2020, teaching at both institutions switched from traditional to virtual instruction through specialized platforms such as Microsoft Teams. No additional instructions were provided to the faculty.

Finally, for practical terms, the data collection in both countries was conducted during January 2021 to March 2022, once the lockdown period was over and the pandemic controlled.

The main aspects related to the methodology are detailed as follows.

3.1 Data collection

The original FACVE proposed seven dimensions (Table 1). These dimensions were screened in order to select only those relevant for the comparison of the data in the two different contexts of Peru and Spain.

Two criteria were considered when screening the dimensions. The first was to eliminate all the dimensions that were directly related to the pandemic such as "Personal and Psychological Issues" (C3). This dimension includes sub-dimensions such as "Health Concerns" that had indicators that were specifically linked to the pandemic context (e.g. "I am concerned about family/friends who have become ill"), which was over at the time of the current data collection. The second criterion was the ability of the variables to allow a fair comparison of the results for both countries. For this reason, dimension C6 "Financial Issues" (e.g. "I am worried about my financial situation") was discarded given that the financial issues of the students were expected to be very different between the two countries, given that Peru is a developing country while Spain is a developed country. Finally, dimension C7 "University Administration and Costs" was discarded given that the indicators were not applicable to the public education context in Spain, unlike in the original framework that was developed mainly with students from a private institution. Table 1 summarizes the dimensions selected for the comparison study.

Table 1. FACVE Dimensions

Dimensions	**Selected for the comparison**
C1. Perceived quality of instruction and learning	Yes
C2. Poor Internet connectivity and lack of proper equipment	Yes
C3. Personal and psychological issues	No
C4. Lack of appropriate home infrastructure	Yes

Dimensions	Selected for the comparison
C5. Learning platform and access to resources	Yes
C6. Financial issues related to students and families	No
C7. General concerns related to the university administration and costs and others	No

The same questionnaire developed in the FACVE (Mu et al., 2022) was used to collect the data (Table 2). A link to the web-based questionnaire was sent via email to the invited participants who responded anonymously during the period January 2021 to March 2022.

Table 2. Description of the questionnaire

Dimensions	Number of questions	Questions
C1. Challenges to the quality of virtual instruction	18	Q13.1 - Q13.7
		Q14.1 - Q14.6
		Q15.1 - Q15.5
C2. Connectivity & equipment	5	Q10.1 - Q10.5
C4. Home infrastructure & study environment	4	Q12.1 - Q12.4
C5. Learning platform and access to resources	8	Q11.1 - Q11.8

In Spain, 400 students students from four different programs were invited to participate in the study with a response rate of 28%. After cleaning and debugging the responses received, the sample size was 94. The participants in Spain are undergraduate and graduate students, mainly between 20-34 years of age. The proportion of participants was slightly higher for the engineering area (55%) than for the social sciences (45%). In addition, 73% had never received any type of distance or virtual training, which indicates that the situation was totally unprecedented for most of the participants.

In Peru, 570 postgraduate MBA students from one of the top private Peruvian universities were invited to participate in this survey, and 165 students responded, yielding a survey response rate of 29%. The participants are mainly 25-35 years old. The proportion of participants was higher for men (64%) than for women (36%).

3.2 Data Validation and Analysis

Before analyzing and comparing the data both databases were tested following well established statistical practices. Exploratory Factor Analysis (EFA) was used in order to analyze the convergence of the individual indicators into their thematic categories or sub-dimensions. Loading factors of 0.4 were used as a lower threshold (Field, 2009). Dimension C1 was not considered in the EFA analysis for the Spanish dataset, due to the sample size (18 questions while our sample was only of 94 students). Therefore, the same three sub-dimensions obtained from the Peruvian dataset remain. Table 3 shows the disaggregation of challenges into the convergence dimensions and sub-dimensions.

Table 3. Disaggregation of dimensions, sub-dimensions and challenges.

Dimensions	Sub-dimensions	Challenges	
C1. Challenges to the quality of virtual instruction	C1.1. Teaching quality	C1.1.1	Teachers are not trained to teach a virtual class in a didactic way
		C1.1.2	Students do not have knowledge of how to study in a virtual class
		C1.1.3	Teachers teach fewer hours than they should
		C1.1.4	Teachers are not motivated to teach classes online
	C1.2. Interaction	C1.2.1	Interaction between classmates is very little
		C1.2.2	Interaction with the teacher is very little
		C1.2.3	It is very difficult to do group tasks
		C1.2.4	No spaces to interact with classmates
		C1.2.5	It is not possible to form study groups
		C1.2.6	Interaction in a virtual class is less than in a face-to-face class
	C.1.3 Assessment	C1.3.1	Academic overload for students in virtual classes
		C1.3.2	Exams are not suitable for online classes
		C1.3.3	Lack of good feedback on the assignments
		C1.3.4	Lack of flexibility of teachers in terms of deadlines
		C1.3.5	Quality of assignments and exams is lower in virtual classes
C2. Connectivity & Equipment		C2.1	I do not have (or have limited) access to the internet where I live
		C2.2	My internet speed is not adequate for my classes

Dimensions	Sub-dimensions	Challenges	
		C2.3	I do not have access (or it is rather limited) to a computer at home
		C2.4	There are many technical problems while accessing classes or study material
		C2.5	My educational institution does not have the appropriate computer equipment (e.g., servers) for virtual teaching
C4. Home infrastructure & Study environment		C4.1	My home does not have adequate physical space for my virtual classes
		C4.2	Activities of other people at home produce a lot of noise and interruptions
		C4.3	I do not have adequate furniture at home (e.g. desk, chair) for my virtual classes
		C4.4	The physical infrastructure that one has at home for virtual classes is less than in the educational institution
C5. Learning platform and access to resources	C5.1 Learning platform	C5.1.1	The educational platform in use is not suitable for virtual instruction
		C5.1.2	Teachers do not know how to use the platform
		C5.1.3	Students do not know how to use the platform
		C5.1.4	There is no information about the use of the platform
	C5.2 Access to resources	C5.2.1	Lack of access to library books is a severe limitation
		C5.2.2	Lack of access to laboratories is a problem
		C5.2.3	It is necessary to have access to more study material (e.g., PPTs) in addition to the recordings of the class
		C5.2.4	Access to teaching resources is less in virtual instruction

Later, the means for the dimensions and sub-dimensions were calculated. In the case of Peru, the scale was from 1 to 5, while for Spain it was from 1 to 7. Hence, for later analysis and comparison the values obtained in both countries were normalized (from 0 to 1 scale). Therefore, the intensity of the challenges has a range from 0 to 1; where 0 corresponds to a non-existent challenge and 1 corresponds to an extreme challenge concern

4 Results

Figure 1 shows the overall results for the dimensions and sub-dimensions. The questions in the survey required to indicate to what extent a participant agreed that the listed situation constituted a challenge. The normalized scale from 0 to 1 represents the extent to which each of the challenges is present, ranging from "not at all" (0) to "definitely yes" (1). In principle, given that the "not at all" constitutes the total absence of the challenge, any value above 0 should be cause of concern and be investigated. However, this is neither feasible nor recommended in practice. For this reason, the 0.2 value (green line in Figure 1) is used in this study to consider a sub-dimension as a potential challenge concern, given that it constitutes 20% of the total possible score. These are sub-dimensions that represent, in this study terminology, **minor concern problems**. On the other hand, and any score above 0.5 (red line in Figure 1) are considered **major concern problems** (red line in Figure 1). Still, some organizations may prefer to use different thresholds given their particular circumstances but taking into account that the ideal situation is the total absence of the virtual education challenge.

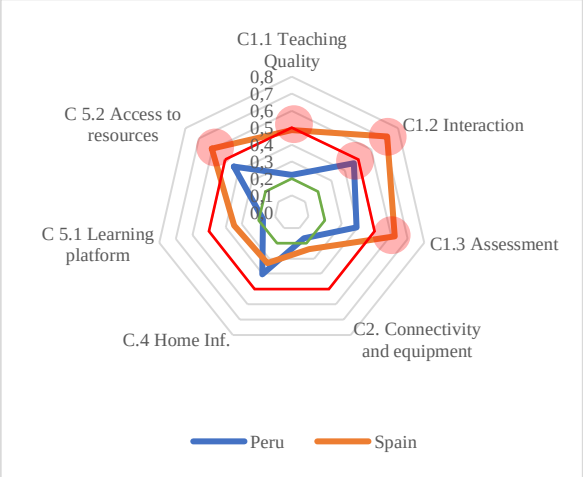


Figure 1. General comparison. Green line: minor concern threshold; Red line: major concern threshold

In the case of Spain, four of the sub-dimensions were considered major concern problems: Teaching quality (C1.1), Interaction (C1.2), Assessment (C1.3) and Access to resources (C5.2). Interaction was the main challenge faced while Connectivity and equipment were the least problematic. In the case of Peru, Interaction (C1.2) represented the major concern problem, while Teaching quality (C1.1), Connectivity

and equipment (C2), and Learning platform (C5.1) were of the least concern.

Figure 1 also clearly shows the differences between the two groups. As can be seen, the intensity of the challenges faced by the Spanish students was much higher in almost all the sub-dimensions in comparison with the Peruvian students, except for the case of Home infrastructure (C4.). This makes sense because of the difference in resources and infrastructure between the contexts.

Challenges to the Quality of virtual instruction (C1) in Spain were much more a major concern than in Peru. This means that Spanish students perceived a lower quality of education compared to Peruvian students. One possible explanation may be the fact that the Peruvian university trained its faculty members to teach virtual classes during a two-week period prior to restarting classes and continued with ongoing monitoring of their faculty and students' needs in the virtual environment.

Connectivity and equipment (C2) was a slightly more concerning problem in Spain than in Peru. This is surprising, given the differences in the contexts; however, both values are below 0.25. Therefore, these were the least of the challenges. Again, we learned that the Peruvian higher education institution supplied laptops, routers and internet access to faculty and students who were in need.

Moreover, Learning platform (C5.1) was another minor concern for both groups. This may be due to the fact that in both contexts specialized e-learning platforms were contracted to support the teaching and learning process (Microsft Teams or Adobe Connect).

Finally, Access to resources (C5.2) was another major difference between the two groups, being more dramatic in the case of the students in Spain. This may be caused by the fact that while all students in the Peruvian sample were business students, the Spanish sample had many engineering students as participants and their resource needs may have been far more sophisticated (e.g., labs).

From Figure 1, we can conclude that the main similarity between the two groups was the concern about Interaction (C1.2), while the main differences were identified in Teaching quality (C1.1), Assessment (C1.3) and Access to resources (C5.2). These four sub-dimensions are discussed in detail in the next section.

4.1 Disaggregation by Sub-dimensions

The detailed results for the sub-dimensions Interaction (C1.2), Teaching quality (C1.1), Assessment (C1.3) and Access to resources (C5.2) are given in the following figures (2-5). In all four subdimensions, Peruvian students perceived to have faced less or equal challenges than the Spanish students.

Students in Spain showed a greater concern for the of dimension Quality of virtual education (C1). In terms of Teaching quality, Peruvian students did not perceive any of these challenges as important; while for Spanish students, the most important challenges were those related to teachers' training and motivation (C1.1.1 and C1.1.4), followed by student preparation for virtual lessons (C1.1.2). The time spent in lessons (C1.1.3) was not perceived as a challenge by either group.

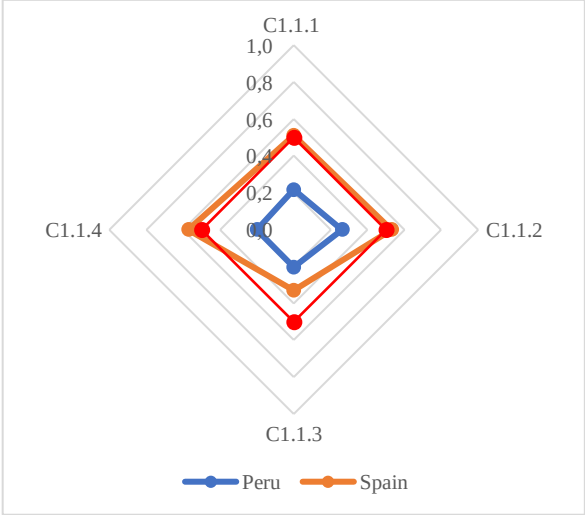


Figure 2. Teaching quality. Red line: major concern threshold

All challenges related to interaction scored close to or greater than 0.5. The inability to interact as much as in face-to-face classes and the lack of opportunities to interact (C1.2.4 and C1.2.6) were the biggest challenges for both groups.

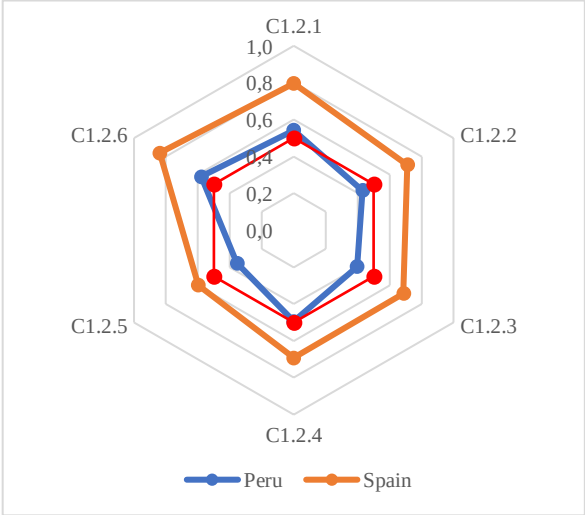


Figure 3. Interaction. Red line: major concern threshold

The intensity of the challenges faced by the Spanish students regarding Assessment was still higher. However, for both groups, the main challenges faced were academic overload for students and the lack of good feedback on the assignments (C1.3.1, and

C1.3.3). Spanish students also expressed major concern about the kinds of exams and the lack of flexibility of teachers in terms of deadlines (C1.3.2, C1.3.4 and C1.3.5).

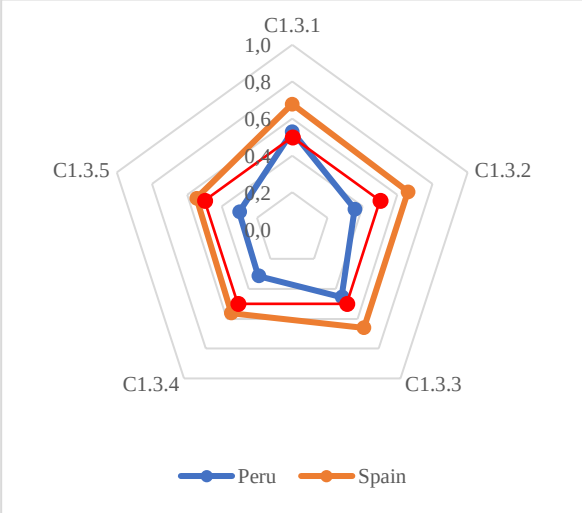


Figure 4. Assessment. Red line: major concern threshold

Although Spanish students maintain a higher valuation of the challenges related to access to resources, in this subdimension, the degree of estimation of the concerns of both groups is closer. They both expressed a limitation in the lack of access to a library and teaching resources (C5.2.1, C5.2.4). Spanish students also consider access to laboratories important (C5.2.2). This may be due to the engineering profile of Spanish students.

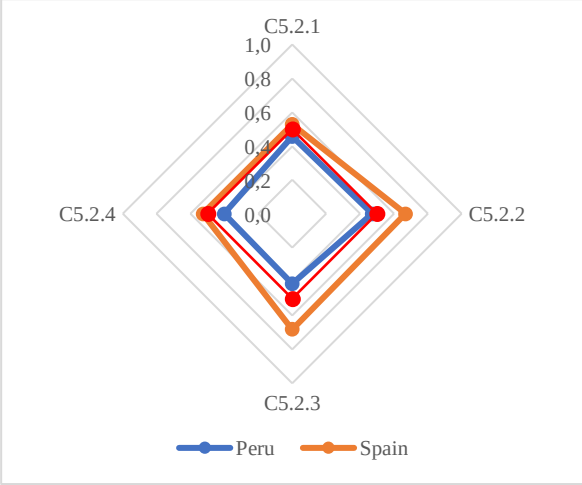


Figure 5. Access to resources. Red line: major concern threshold

All the previous figures confirm that the Spanish students faced more challenges during the period of ERT. It is possible that the preparation time that the teachers had in Peru allowed them to better adjust the content of the classes and the type of evaluation as well as better prepare material. This was reflected in the evaluation of these concerns.

Moreover, we analyzed the dimension of Home infrastructure and study environment (C4) because it was the only one in which Peruvian students expressed having faced greater challenges than Spanish students. Figure 6 shows that interruptions caused by other people at home and the lesser conditions of the physical infrastructure at home in comparison to the educational institution (C4.2 and C4.4) were the points of greatest agreement between the two groups.

The lack of adequate physical space and furniture at home (C4.1 and C4.3) represented a major challenge for Peruvian students which makes sense when considering the contexts of these students. Peruvian students tend to share spaces with the family, while Spanish students tend to share spaces with other students which provides more suitable conditions for students.

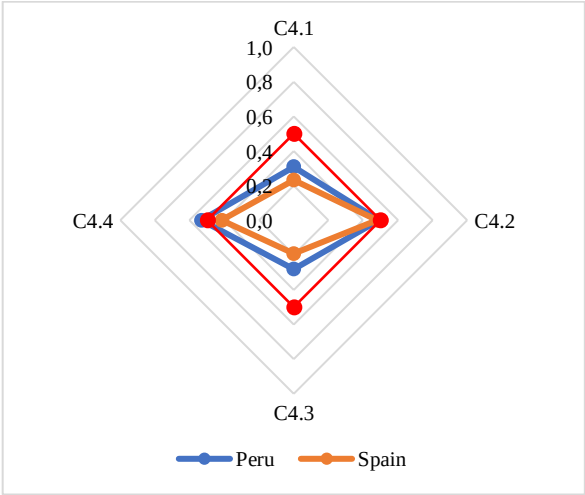


Figure 6. Home infrastructure and study environment. Red line: major concern threshold

5 Conclusions

In conclusion, this study shows the statistical validity and versatility of the FACVE framework to assess the extent of challenges to virtual education in an ERT context. Also, the comparative analysis of the existing challenges leads us to suggest some possible ways to minimize these challenges as follows:

1. Training prior to virtual teaching. The Peruvian institution stopped providing classes for several weeks while training their faculty. The comparison institution trained their faculty on the go and students suffered because of this and complained about the overall quality of the virtual instruction.
2. Assist students with home infrastructure and study environment (e.g. noisy, lack of own desk). Universities could provide (or rent) to students in need some simple material such as headphones

(to minimize noises), portable desk (so the student does not need to connect virtually from bed) and similar. This was a more serious issue in Peru than in Spain although it was present in both countries.
3. Update the activities and evaluation systems to the virtual environment. Taking into account the differences in time available, opportunities for interaction between students, etc.
4. Providing some additional opportunities for interaction among students and between students and teachers. For instance: Blogs, Chat rooms, discussion forums, etc. Where they can more freely exchange questions, comments, materials, etc.
5. In general, having extra time for planning may provide a positive difference in results even when compared with institutions in better socio-economic situations.

One of the limitations of our study was the size of the sample which did not allow for statistically testing the challenge dimension (C1) with the greatest number of indicators. On the other hand, the next step in this study is to assess the different challenges in terms of their assessed relative importance by the students. Finally, this comparative study does not attempt to illustrate country differences in the way they address ERT, but simply highlight how a quick and effective higher-education management response may help to overcome socio-economic differences in an ERT situation.

Acknowledgments

This work was partially supported by the European Union - Next generation EU [Ministerio de Universidades de España; Margarita Salas grant]

References

Adell, J. (2020). *La Educación que viene: Un modelo híbrido y más tecnológico*.

Area Moreira, M., Bethencourt Aguilar, A., & Martín Gómez, S. (2020). De la enseñanza semipresencial a la enseñanza online en tiempos de Covid-19: Visiones del alumnado. *Campus Virtuales: Revista Científica Iberoamericana de Tecnología Educativa*, *9*, 35–50.

Bearman, M., Nieminen, J. H., & Ajjawi, R. (2023). Designing assessment in a digital world: An organising framework. *Assessment & Evaluation in Higher Education*, *48*(3), 291–304. https://doi.org/10.1080/02602938.2022.2069674

Charles Hodges, Stephanie Moore, Barb Lockee, Torrey Trust, & Aaron Bond. (2020). The Difference Between Emergency Remote Teaching and Online Learning | EDUCAUSE. *EDUCAUSE Review*.

Elahi, S., baan, Kanaani, F., & Shayan, A. (2023). Designing a Framework for Effective Factors on Virtual Studentsâ€ ™Tendency to the Electronic Learning and its Assessment. *Quarterly Journal of Research and Planning in Higher Education*, *17*(2), 59–80.

Field, A. (2009). Discovering statistics using IBM SPSS statistics. In *Statistics* (Third, Vol. 58). Sage.

Ligardo-Herrera, I., Corona-Sobrino, C., & Gonzalez-Urango, H. (2022). La docencia universitaria en tiempos de pandemia: Perspectiva desde el personal docente y el alumnado. In Á. Pérez García, N. Reyes Ruiz de Peralta, & F. de M. Sánchez Aguirre (Eds.), *La COVID-19 llega a las aulas* (pp. 247–265). Thomson Reuters Aranzadi.

Mou, T.-Y. (2023). Online learning in the time of the COVID-19 crisis: Implications for the self-regulated learning of university design students. *Active Learning in Higher Education*, *24*(2), 185–205. https://doi.org/10.1177/14697874211051226

Mu, E., Florek-Paszkowska, A., & Pereyra-Rojas, M. (2022). Development of a Framework to Assess Challenges to Virtual Education in an Emergency Remote Teaching Environment: A Developing Country Student Perspective;The Case of Peru. *Education Sciences 2022, Vol. 12, Page 704*, *12*(10), 704. https://doi.org/10.3390/EDUCSCI12100704

Mukasheva, M., Kornilov, I., Beisembayev, G., Soroko, N., Sarsimbayeva, S., & Omirzakova, A. (2023). Contextual structure as an approach to the study of virtual reality learning environment. *Cogent Education*, *10*(1), 2165788. https://doi.org/10.1080/2331186X.2023.2165788

Tiemann, R., & Annaggar, A. (2023). A framework for the theory-driven design of digital learning environments (FDDLEs) using the example of problem-solving in chemistry education. *Interactive Learning Environments*, *31*(2), 1199–1212. https://doi.org/10.1080/10494820.2020.1826981

Unicef. (2021). *COVID-19 and School Closures One Year of Education*.

van den Beemt, A., van de Watering, G., & Bots, M. (2023). Conceptualising variety in challenge-based learning in higher education: The CBL-compass. *European Journal of Engineering Education*, *48*(1), 24–41. https://doi.org/10.1080/03043797.2022.2078181

Whittle, C., Tiwari, S., Yan, S., & Williams, J. (2020). Emergency remote teaching environment:

A conceptual framework for responsive online teaching in crises. *Information and Learning Science*, *121*(5–6), 301–309. https://doi.org/10.1108/ILS-04-2020-0099/FULL/PDF

Wong, R. (2023). When no one can go to school: Does online learning meet students' basic learning needs? *Interactive Learning Environments*, *31*(1), 434–450. https://doi.org/10.1080/10494820.2020.1789672

Towards Increasing Active Citizen Involvement in Urban Planning through Mixed Reality Technologies

Sam Sabah, Imran Hossain, David Weiss, Alexander Tillmann
Goethe University Frankfurt
studiumdigitale
Eschersheimer Landstr. 155-157, 60323 Frankfurt am Main
{sabah, hossain, weiss, tillmann}@sd.uni-frankfurt.de

Abstract. *Augmented Reality (AR) and Virtual Reality (VR) have the potential to revolutionize urban planning by providing immersive and interactive experiences for visualizing urban spaces. However, citizen participation is hindered by the perceived costs of learning and the technical skills required for AR and VR. Physical models offer a tangible representation of plans, enabling citizens and students to visualize their ideas. Combining physical objects with AR and VR is challenging but holds promise for citizen participation. This paper explores the integration of physical models with AR/VR to create a synergistic effect that fosters citizen engagement in urban planning. We introduce the concept of the 'City Planner,' a mixed-reality application currently being developed, which combines AR/VR, urban planning, and physical objects to transcend traditional participation barriers. The proposed concept aims to make urban planning accessible, intuitive, and inclusive for citizens and students, regardless of their technical expertise.*

Keywords. Social citizenship; Urban Planning; Citizens Participation; Extended Reality (XR); Augmented Reality (AR); Virtual Reality (VR), Mixed Reality (MR); Object detection; Image recognition

1 Introduction

The integration of Extended Reality (XR), which encompasses Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) technologies, has the potential to revolutionize urban planning processes by visualizing the unseen.

XR models of cities can be created to provide a flexible and functional representation of the urban environment, making it easier to visualize the unseen (Thompson et al., 2006). However, ensuring the active participation of civil society in urban planning projects remains a challenge. Citizens often show "rational ignorance" toward urban planning events, and the cost of learning how to participate in such projects often outweighs their perceived potential benefit from taking part (Krek, 2005). XR technology can improve civil society participation in urban planning by creating interactive and immersive experiences that are easily understandable. This, in turn, can increase engagement and participation in planning processes, allowing community needs and opinions to be considered. By involving civil society in planning processes, people can develop a stronger sense of belonging and connection to their community and city. They have the opportunity to participate in decisions that affect their immediate surroundings and daily lives, making them better able to engage in their community (L., 2016; Siembab, 2019).

In school education, developing skills for participation and co-design of urban spaces is crucial (A., 2016; D., 2015). The educational concept of Spatial Citizenship in Geography involves children and adolescents participating in local urban planning processes to promote their socio-spatial perception and understanding of urban spaces. Geomedia, including digital tools like scribble maps and sketching programs, are used to facilitate this process (D., 2015; Kanwischer et al., 2018). Visualizing ideas for public space redesign remains abstract and difficult to imagine how one's plans would look and feel in reality. Digital Geomedia and maps also pose accessibility issues. A solution that provides better visualization and ease of operation is highly desirable for participatory planning processes involving students and citizens. Tangible interfaces can be more inviting and conducive to collaborative interaction (Thompson et al., 2006), therefore combining the allure of something tangible with the advantages of XR technology holds the promise to raise the level of citizen engagement in urban planning.

2 State of the art

A substantial body of academic research emphasizes the significance of employing XR technology in urban planning. Using an AR platform can enhance public familiarity with urban design indicators and effectively promote public participation in evaluating urban design (Wang and Lin, 2023). Similarly, the utilization of Virtual Reality (VR) technology in participatory

urban planning also offers numerous advantages, including higher engagement and more vivid memory of the viewed content, which can lead to better-informed decision-making (Van Leeuwen et al., 2018). Within the scope of a participatory design project, a positive social impact on the neighborhood can be observed, reflecting the needs and wishes of the inhabitants and increasing their interest in local urban development (Van Leeuwen et al., 2018). In order for VR to effectively support urban planning, factors such as 3D visualization, simulating real-world scenes, user-friendliness, interactivity, inspiring enthusiasm, and inspiring creative thinking should be considered (Jiang et al., 2023).

VR has significant potential in urban design, offering real-time testing, effective communication, and improved designs. However, challenges include computational costs, limited accessibility, technical shortcomings, and the need for further research. Overcoming these challenges will enable VR to enhance the urban design and increase public participation (El Araby, 2002). The use of 3D data, therefore, seems to be beneficial for urban planning, and XR technologies can help predict the look and impact of new architectural objects on the surrounding environment (Cirulis and Brigmanis, 2013). The use of XR technologies in urban planning can increase citizen participation, particularly among young people, through its ability to provide spatial and contextual cues that enhance creativity and interactivity. The use of haptic 3D tools enabled collaboration in teams and provided an option for user engagement (Saßmannshausen et al., 2021).

For architecture students, including new interactive and gamified visualization systems can have a positive impact on the development of urban projects as an educational tool. These systems can improve students' academic and competence development, which is linked to their motivation (Fonseca et al., 2017). Integrating Service Learning (SL) into geography lessons effectively encourages interest in participating in urban planning processes among children and adolescents. By incorporating SL activities into the curriculum, students gain new perspectives and multi-perspectives on the daily actions of children and adolescents in the context of urban planning (Schulze et al., 2015).

XR technology offers promising improvements to urban planning processes by increasing comprehensibility, interactivity, and accessibility and enabling earlier citizen participation in the planning process. However, further systematic development and analysis of XR applications in urban planning are required (Wolf et al., 2020). The utilization of such an approach can lead to more informed and collaborative decision-making (Seichter, 2007).

3 Research objective

The state of the art shows that there is a growing body of scientific evidence that supports the benefits of combining tangible elements and XR technology in urban planning to increase citizen participation.

A City Planner that combines planning with tangible and XR technology is being developed to explore the efficacy of actively involving citizens in urban planning and helping students develop spatial citizenship skills. The application creates a Mixed Reality environment that utilizes physical objects to plan the construction of a virtual space. The aim is to gain insights into the potential benefits of XR technology in urban planning and increase citizen participation in the planning process. It is essential to clarify that, at this stage, the City Planner is purely a conceptual idea under development and not a fully developed or commercially available product.

3.1 Methodology

Utilizing XR technology with 3D rendering in urban planning has been found to offer numerous benefits, including higher engagement and more vivid memory compared to traditional 2D presentation technologies (Van Leeuwen et al., 2018). By immersing users in a realistic virtual environment, VR enhances their spatial understanding and enables them to interact with urban spaces in a more immersive and intuitive manner. However, despite the potential advantages, the development of XR applications for urban planning still presents significant obstacles for many individuals (Ashtari et al., 2020). Challenges such as technical complexity, high costs, and the requirement of specialized skills hinder widespread adoption and limit the accessibility of these technologies.

To address these challenges, there is a growing need to create more compelling and meaningful user experiences in XR that can help overcome the barriers to adoption (Ashtari et al., 2020). The focus is shifting towards developing user-friendly tools and platforms that simplify the process of creating XR applications and empower users to harness the full potential of these technologies.

In line with this objective, our proposed approach is the City Planner, which aims to streamline the development process of XR experiences in urban planning. By automating various technical aspects and providing pre-designed templates and tools, the City Planner aims to alleviate the burden on users and facilitate the creation of immersive and interactive XR experiences. This approach empowers urban planners, designers, and citizens with limited technical expertise to actively participate in the urban planning process and make informed decisions.

The City Planner serves as a bridge between the complexities of XR development and the need for accessible and user-friendly tools. It empowers users to create engaging and realistic virtual simulations of urban spaces, allowing stakeholders to visualize and evaluate proposed designs, make more informed decisions, and actively contribute to the planning process. By pro-

viding an intuitive and streamlined user experience, the City Planner aims to democratize the use of XR in urban planning and foster greater public engagement and collaboration.

3.2 Design and Development

City Planner consists of six key components that work together. These components interact with each other to create a tool that allows different stakeholders to explore various planning scenarios, understand the potential impacts of different strategies, and make informed decisions.

Figure (1) illustrates the different component of The City Planner.

1. Gameboard Component: This component is a portable surface that represents a piece of real-world land where the Physical Objects can be placed to create a virtual model of the city or land to be planned.
2. Physical Objects Component: These are elementary abstract tangible 3D objects that represent real-world objects, such as trees, buildings, roads, cars, and other urban infrastructure elements. These customizable 3D objects allow users to recreate and experiment with different layouts, designs, and arrangements of various urban elements.
3. Stakeholders Component: This aspect of City Planner includes all individuals and groups involved in the urban planning process, such as citizens, decision-makers, and other interested parties. City Planner can be integrated into the classroom as a part of Service Learning, making students a part of the Stakeholders Component. By incorporating the perspectives and opinions of a diverse range of stakeholders, the planning process becomes more inclusive and informed.
4. Monitoring System: This component is responsible for observing and directing the changes made by users on the Gameboard and sending this information to the Authoring component. It ensures that all modifications and arrangements made on the Gameboard are accurately tracked and recorded in real time.
5. Authoring Component: This Component of City Planner constructs the XR (extended reality) experience. It processes the information collected by the Monitoring System and translates it into a VR, MR, or AR experience that users can engage with.
6. Viewing device: This can be an XR-capable device, usually a VR/MR Headset or an AR-capable smartphone, that allows users to view the city planning experience. This device enables users to immerse themselves in a virtual model of the city using VR technology or use it for Realworld Mapping via an AR device by overlaying AR virtual objects on a real-world location. This enables the users to explore various planning scenarios and understand the potential impacts of different urban planning strategies.

These six components work together to create a powerful tool that simplifies urban planning. To ensure the effectiveness of the Object Detection module in observing and monitoring any changes made, the 3D objects produced must be distinguishable from one another. This is necessary to facilitate the process of monitoring and tracking any changes or modifications made to the objects on the Gameboard.

The Gameboard itself is intelligently segmented into distinct sections, each of which can be readily identified by the Monitoring System.

This identification is achieved by utilizing Computer Vision (CV) and Artificial Intelligence (AI) techniques through the implementation of two separate AI detection models. The first model is an object recognition model, which serves to identify the specific type of object placed on the Gameboard, while the second model, an object detection model, is responsible for detecting and determining the position of the object relative to the Gameboard.

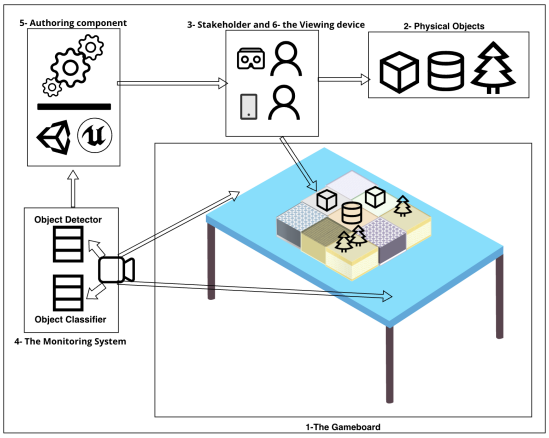


Figure 1: The City Planner Components.

When a stakeholder adds an object to the Gameboard, the Monitoring System detects the change made and recognizes both the type and location of the added object on the Gameboard. This information is then transmitted to the Authoring Tool. Based on the provided information, the Authoring Tool uses UEmbed, a browser-based tool that enables users to map out an area from a top-down perspective and add selective elements like buildings, trees, and characters with various properties. The created map can be exported as a JSON File and imported into a pre-prepared Unreal Engine project with scripts that read the data from the JSON File and build the environment. These functionalities can be used to recreate the world with the data provided by the Monitoring System and pushed to the customized UEmbed Unreal Project.

3.3 Reflection and Improvements

Automated virtual representation creation simplifies the urban planning process, allowing stakeholders to explore the immersive 3D environment and make informed decisions without the burden of manual development. City planners utilize computer vision for object recognition. However, the recognition performance is highly dependent on image quality, which is influenced by factors such as camera sensor, lens quality, illumination, and other camera settings (Kaltenbrunner and Bencina, 2007). Improvements in these aspects can enhance the effectiveness of computer vision in the City Planner.

When considering the redesign of an urban square, it is important to have a clear process in place to ensure that the final design meets the needs of its users. To this end, users can take several steps to guide their decision-making.

1. Identify and preserve surrounding structures or elements: To maintain the character and authenticity of the urban space, it is essential to consider the existing buildings and elements in the area. By preserving these structures, the redesign can seamlessly integrate with the surrounding environment.
2. Arrange new elements and features: Determine the most suitable location and arrangement of new buildings, trees, and other urban elements within the square(Gameboard). This involves carefully considering factors such as functionality, aesthetics, and the preferences of the community. This may include creating open spaces, green zones, and pedestrian-friendly walkways that cater to the needs and preferences of the community.
3. Evaluate different scenarios: Exploring various design options and layouts allows stakeholders to assess the potential impact of each design on the existing urban landscape and the daily lives of its users. This evaluation should consider factors such as sustainability, traffic flow, accessibility, and social dynamics within the community.
4. Gather feedback and incorporate suggestions: Engage with stakeholders, including citizens, urban planners, and decision-makers, to gather feedback on the proposed designs. By incorporating the suggestions and ideas of different stakeholders, the final design can better represent the needs and expectations of all users.
5. Finalize and implement the design: exploring the effect of the design in an XR-capable device to evaluate important factors such as density, relationships between elements, and lines of sight.

4 Summary and Future Work

Urban planning has faced persistent challenges in engaging citizens effectively, limiting decision-making and community involvement. However, recent research highlights the transformative impact of XR technology on urban planning processes. By integrating virtual and augmented reality experiences, citizens can immerse themselves in proposed developments, enabling a deeper understanding of potential changes and their consequences. Additionally, tangible interfaces like gameboards and 3D objects foster collaborative interactions, allowing stakeholders to actively engage, discuss, and work towards consensus, ultimately leading to more inclusive and successful urban planning outcomes.

Using VR technology with 3D rendering in urban planning can provide higher engagement and more vivid memory than 2D presentation technologies (Van Leeuwen et al., 2018).

Despite the promising advantages of XR technology in urban planning, the development of XR applications remains challenging for many. Creating an effective XR application tailored for urban planning purposes involves various difficulties and obstacles that need to be addressed. The City Planner aims to simplify citizen participation in urban planning by automating the development process. This application uses tangible and object detection models to create an immersive 3D environment, which allows users to explore and make informed decisions about the future of their city. Urban planning could be transformed by XR technology integration in terms of decision-making and citizen involvement. Augmenting the XR experience by adding 3D sound can provide an extra layer of information for a more informed decision-making process, but further development and analysis of XR applications are needed.

The integration of XR technology in urban planning and education shows promise to enhance citizen participation, decision-making processes, and community involvement. XR applications offer promising advancements in terms of accessibility and understanding, but additional development and analysis are required to fully realize their benefits.

References

A., S. A. Ł. (2016). Kultusminister konferenz, & bundesministerium für wirtschaftliche zusammenarbeit und entwicklung. *Orientierungsrahmen für den Lernbereich Globale Entwicklung im Rahmen einer Bildung für nachhaltige entwicklung. 2. (aktualisierte und erweiterte) Auflage. Bonn.*

Ashtari, N., Bunt, A., McGrenere, J., Nebeling, M., & Chilana, P. K. (2020). Creating augmented and virtual reality applications: Current practices, challenges, and opportunities. *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–13.

Cirulis, A., & Brigmanis, K. B. (2013). 3d outdoor augmented reality for architecture and urban planning. *Procedia Computer Science*, *25*, 71–79.

D., S. U. G. I. K. (2015). Spatial citizenship education and digital geomedia: Composing competences for teacher. *Journal of Geography in Higher Education*, *39*(03), 369–385.

El Araby, M. (2002). Possibilities and constraints of using virtual reality in urban design. *Proceedings of the 7Th International CORP Symposium, Vienna, Austria*, 457–463.

Fonseca, D., Villagrasa, S., Navarro, I., Redondo, E., Valls, F., Llorca, J., Gómez-Zevallos, M., Ferrer, Á., & Calvo, X. (2017). Student motivation assessment using and learning virtual and gamified urban environments. *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*, 1–7.

Jiang, H., Geertman, S., Zhang, H., & Zhou, S. (2023). Factors influencing the performance of virtual reality in urban planning: Evidence from a view corridor virtual reality project, beijing. *Environment and Planning B: Urban Analytics and City Science*, *50*(3), 814–830.

Kaltenbrunner, M., & Bencina, R. (2007). Reactivision: A computer-vision framework for table-based tangible interaction. *Proceedings of the 1st international conference on Tangible and embedded interaction*, 69–74.

Kanwischer et al. (2018). Globales lernen in der geographischen lehrerinnen- und lehrerbildung durch service learning: Ein fallbeispiel im kontext digitaler geomdeien und räumlicher sozialisation. *In Brendel N. Schrüfer G. Schwarz I. (Hrsg.) Globales Lernen im digitalen Zeitalter. Waxmann. Münster*, *39*(03), 147–170.

Krek, A. (2005). Rational ignorance of the citizens in public participatory planning. *10th symposium on Information-and communication technologies (ICT) in urban planning and spatial development and impacts of ICT on physical space, CORP*, *5*, 420.

L., W. D. J.
bibinitperiod Q. (2016). Participation in urban planning and its impact on the health and well-being of communities. journal of planning literature. *Journal of Planning Literature*, *31*(4), 399–411.

Saßmannshausen, S. M., Radtke, J., Bohn, N., Hussein, H., Randall, D., & Pipek, V. (2021). Citizen-centered design in urban planning: How augmented reality can be used in citizen participation processes. *Designing Interactive Systems Conference 2021*, 250–265.

Schulze, U., Gryl, I., & Kanwischer, D. (2015). Spatial citizenship education and digital geomedia: Composing competences for teacher education and training. *Journal of Geography in Higher Education*, *39*(3), 369–385.

Seichter, H. (2007). Augmented reality and tangible interfaces in collaborative urban design. *Computer-Aided Architectural Design Futures (CAADFutures) 2007: Proceedings of the 12th International CAADFutures Conference*, 3–16.

Siembab, A. Ł. A. (2019). 3d urban visualization for participatory spatial planning: Review of tools and user studies. *Journal of Planning Literature*, *26*(04), 41–61.

Thompson, E. M., Horne, M., & Fleming, D. (2006). Virtual reality urban modelling-an overview.

Van Leeuwen, J. P., Hermans, K., Jylhä, A., Quanjer, A. J., & Nijman, H. (2018). Effectiveness of virtual reality in participatory urban planning: A case study. *Proceedings of the 4th Media Architecture Biennale Conference*, 128–136.

Wang, Y., & Lin, Y.-S. (2023). Public participation in urban design with augmented reality technology based on indicator evaluation. *Frontiers in Virtual Reality*, *4*, 1071355.

Wolf, M., Söbke, H., & Wehking, F. (2020). Mixed reality media-enabled public participation in urban planning: A literature review. *Augmented reality and virtual reality: changing realities in a dynamic world*, 125–138.

Evaluating the Net Zero Policy Performances of European Countries Through DEA

Mehtap Dursun
Galatasaray University
Decision Analysis Research and Application Center
Çırağan Cad. No:36 Ortaköy, Beşiktaş
İstanbul, Türkiye
mdursun@gsu.edu.tr

Rana Duygu Alkurt
Galatasaray University
Industrial Engineering Department
Çırağan Cad. No:36 Ortaköy, Beşiktaş
İstanbul, Türkiye
RANADUYGU.ALKURT@ogr.gsu.edu.tr

Abstract. *Addressing climate change and its environmental impact is one of the greatest challenges of modern life. Most of today's sustainable strategies are essentially related to the intent of reducing overall carbon dioxide emissions. The net zero goal lays the foundation for a sustainable future. This study aims to measure the performance of European Countries based on carbon emissions. To measure performance, Pure Environmental Performance Index and Mixed Environmental Performance Index Data Envelopment Analysis (DEA) methods are used. Decision-making units (DMUs), inputs and outputs are determined. Input is identified as Primary Energy Consumption. Outputs are selected as Gross Domestic Product (GDP), Carbon dioxide (CO_2) emission, Nitrous Oxide (N_2O) emission. According to performance results, Albania, Lithonia, Malta, and Montenegro are classified as efficient countries via the two models. The results shows that the European Countries should take the required measures of the Paris Conference till 2050 to reduce the damage to the environment.*

Keywords. Climate change, Carbon dioxide emission, Data envelopment analysis, Net zero policy, Nitrous Oxide emission, Performance management.

1 Introduction

The net zero goal lays the foundation for a sustainable future. It is in question to keep global warming under control since at least carbon dioxide emissions are balanced with the net zero target. In order to avoid severe climate impacts, global greenhouse gas emissions should decrease by half by 2030 and reach zero by 2050.

In order to reach the net zero target, international negotiations were started in the late 1980s (Vlassopoulos, 2012). Carbon dioxide, methane etc. In order to reduce greenhouse gases, the Kyoto Protocol was signed between 40 countries in 1997. As a result of the inadequacy of the Kyoto protocol, the Paris Conference was held in 2015. After the conference, the Paris Agreement was signed but entered into force in 2016 (International Legal Materials, 2016).

In recent years, researchers have been conducting carbon emission analyses specific to regions. Zhou et. al. (2008) divided the world into 8 regions such as OECD, Africa. They made a carbon, energy density comparison of 8 regions using the Data Envelopment Analysis (DEA) method. On the other hand, some researchers have focused on specific countries. Guo et al. (2011) conducted a study on 29 provinces in China. They aimed to estimate CO_2 emissions reduction in provinces using multiple DEA methods. Millot et al. (2020) evaluated Sweden and France energy consumption. According to the evaluations, it has been revealed that it is important to turn to renewable energy sources. Although renewable energy sources are beneficial to environment, they require high investment and have many uncertainties.

Also, there are studies that make carbon footprint analysis in the literature [Kuo et al. (2014), Mayers et al. (2014), Belkhir and Elmeligi (2018), Millot et al. (2020)]. After the Paris Agreement signed in 2015, studies on the net zero emission approach were carried out [Wu and Skye (2021), Miller et al. (2021), Chaudry et al. (2022), Tudor and Sova (2022)].

The motivation of this study is to measure the environmental performance of European Countries and to investigate how well the European Countries can comply with the aims of the Paris Agreement. For this purpose Primary Energy Consumptions, Gross Domestic Product, Carbon dioxide emissions and Nitrous Oxide emission are identified as performance evaluation criteria via the literature review. To measure performance, the Data Envelopment Analysis (DEA) method is used. As the classical DEA models do not consider the undesirable outputs, this study employed Pure Environmental Performance Index and Mixed Environmental Performance Index DEA methods, which also considers the undesirable outputs, are employed. The contributions of this study are as follows. This study provides a roadmap for European Countries to reach the Paris Agreement's requirement.

The countries that are identified as inefficient should take the necessary actions in terms of carbon dioxide emissions and nitrous oxide emissions till 2050. Moreover, the studies in the literature only measure the performances by taking in the account carbon dioxide emissions. This study considers both the carbon dioxide and nitrous oxide emissions, which are important environmental pollutant gases. The rest of the study is organized as follows. Section 2 explains the DEA method, Section 3 illustrates the case study. Conclusions are provided in the last Section.

2 Data Envelopment Analysis

It is a nonparametric technique to measure the relative efficiency of performance measure within a set of homogeneous decision units (DMUs) with inputs, desirable outputs and undesirable outputs (Martin-Gambo and Iribarren, 2021). The first DEA model, which named as CCR (Charles, Cooper and Rhodes), developed by Charnes et al. (1978). There are two CCR DEA methods according to the change in objective functions. These are input-oriented and output-oriented which have desirable inputs and desirable outputs. The input-oriented CCR-DEA model evaluate the relative efficiency of DMUs. This evaluation is done by maximizing the ratio of the total weighted output to the total weighted input, This model constraint is output-to-input ratio of each DMU should be less than or equal to unity.

The input-oriented CCR-DEA model can be represented as follow;

$$\max E_{j_0} = \frac{\sum_{r=1}^{s} u_r y_{rj_0}}{\sum_{i=1}^{m} v_i x_{ij_0}}$$

subject to

(1)

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \quad \forall j,$$
$$u_r, v_i \geq \varepsilon, \quad \forall r, i,$$

DEA efficiency measures integrated with the concept of environmental DEA technology have gained popularity in environmental performance measurement in recent years. On the orher hand, Outputs or inputs can be undesirable like CO_2 Emission or GHG Emission. The economic justification for using undesirable output variables as inputs in DEA models. Inputs and undesirable outputs cost a DMU money, therefore DMUs often try to minimize both sorts of variables. To deal with undesirable outputs, new mathematical models have been developed by changing the constraints. For example, Tyteca (1996) developed model whose name is Pure Environmental Performance Index. This model assume that the environmental DEA technology exhibits constant returns to scale. In the scope of the theory of productive efficiency, three categories of factor are taken into account as inputs, desirable production outputs and pollutants in the form of ‘‘undesirable’’ outputs. Non-parametric efficiency measures easily and usefully lend themselves to the derivation of environmental performance indicators. They are the duals of indicators that can be obtained in the traditional framework of DEA.

Min Θ

Subject to

$$\begin{aligned} &\textstyle\sum_{k=1}^{K} z_k x_{nk} \leq x_{n0} && \forall n, \\ &\textstyle\sum_{k=1}^{K} z_k y_{mk} \geq y_{m0} && \forall m, \\ &\textstyle\sum_{k=1}^{K} z_k u_{jk} = \Theta u_{j0} && \forall j, \\ &\textstyle\sum_{k=1}^{K} z_k \leq 1 \\ &Z_k \geq 0 \,,\ k = 1, .. \,, K. \end{aligned} \qquad (2)$$

In above linear model is consist of an input vector x_k whose nth component x_{nk} is the amount of input n consumed by DMU_k , output vector y_k whose mth component y_{mk} is the amount of desirable output m yield by DMU_k, an output vector u_{jk} whose jth component u_{jk} is the amount of undesirable output j yield by DMU_k .

Contrary to a pure EPI, an environmental performance measure considering the simultaneous adjustments of desirable and undesirable outputs can be called a mixed EPI. Zhou et al. (2008) discuss the environmental DEA technologies that exhibit non-increasing returns to scale and variant returns to scale. They proposed the pure measures under different situations and a mixed measure under the variant returns to scale environmental DEA technology for measuring environmental performance.

Zhou et al. (2008) introduced Mixed Environmental Performance Index (MEI) as

min λ

Subject to

$$\begin{aligned} &\textstyle\sum_{k=1}^{K} z_k x_{nk} \leq \beta\, x_{n0} && \forall n, \\ &\textstyle\sum_{k=1}^{K} z_k y_{mk} \geq y_{m0} && \forall m, \\ &\textstyle\sum_{k=1}^{K} z_k u_{jk} = \lambda\, u_{j0} && \forall j, \\ &\textstyle\sum_{k=1}^{K} z_k = \beta \\ &z_k \geq 0 \,,\ k = 1, .. \,, K. \end{aligned} \qquad (3)$$

Where an input vector x_k whose nth component x_{nk} is the amount of input n consumed by DMU_k , output vector y_k whose mth component y_{mk} is the amount of desirable output m yield by DMU_k, an output vector u_{jk} whose jth component u_{jk} is the amount of undesirable output j yield by DMU_k .

3 Measuring the Environmental Performances of European Countries

The case study is performed in European Countries. The Data are obtained from https://ourworldindata.org/. Input and outputs are identified by reviewing the literature. Input is identified as Primary Energy Consumption. Outputs are selected as Gross Domestic Product (GDP), Carbon dioxide (CO_2) emission, Nitrous Oxide (N_2O) emission. The data are provided in Appendix.

To measure the performances, Pure Environmental Performance Index mathematical model and Mixed Environmental Performance Index mathematical model are coded in General Algebraic Modeling System v.42.5.0 (GAMS). The models are run by putting normalized values. Models results are shown in Table 1.

Table 1. Performance measurements

Country	Results of Model (2)	Results of Model (3)
Albania	1.000	1.000
Austria	0.034	0.046
Belgium	0.021	0.026
Bulgaria	0.028	0.051
Cyprus	0.327	0.380
Czechia	0.017	0.018
Denmark	0.148	0.180
Estonia	0.163	0.187
Finland	0.065	0.071
France	0.008	0.009
Germany	0.003	0.004
Greece	0.020	0.029
Hungary	0.037	0.048
Iceland	0.845	1.000
Ireland	0.559	1.000
İtaly	0.005	0.005
Lithonia	1.000	1.000
Luxemburg	0.464	1.000
Malta	1.000	1.000
Montenegro	1.000	1.000
Nether	0.015	0.021
North Macedonia	0.129	0.265
Norway	0.070	0.100
Poland	0.005	0.006
Portugal	0.032	0.039
Roma	0.022	0.030
Russia	0.001	0.001
Serbia	0.020	0.047
Slovakia	0.040	0.054
Slovenia	0.116	0.13
Spain	0.008	0.008
Sweden	0.083	0.095
Switzerland	0.084	0.134
Türkiye	0.003	0.005
United kingdom	0.006	0.006

According to Table 1, the results of the two models are close to each other. Albania, Lithonia, Malta, and Montenegro are classified as efficient countries via the two models. Mixed environmental index model also determined Iceland, Ireland, and Luxemburg as efficient countries. Although Albania's GDP value is lowest in country group, its low ranking in emissions and energy consumption has made it one of the countries with the highest performance. Although Malta's GDP value is not the highest, its high GDP value compared to other countries and its low undesirable output values make it one of the countries with the highest performance. Turkey and Germany are the countries with the worst performances in both models as a result of their high N_2O and CO_2 emission values compared to other countries.

4 Conclusions

This study aims to measure the environmental performance of European Countries and to investigate how well the European Countries can comply with the aims of the Paris Agreement. For this purpose Primary Energy Consumptions, Gross Domestic Product, Carbon dioxide emissions and Nitrous Oxide emission are identified as performance evaluation criteria via the literature review. To measure the performance, pure environmenta index DEA method and mixed environmental index DEA method are used. According to performance results, Albania, Lithonia, Malta, and Montenegro are classified as efficient countries via the two models. Mixed Environmental Performance Index mathematical model also provides Iceland, Ireland, and Luxemburg as efficient. European countries should continue to take precautions for a sustainable world. Future researchers may focus on performing this analysis for 2050. Moreover other DEA methods, which have high discrimination power can be used. Necessary measures can be taken according to these measurements.

References

Belkhir, L., & Elmeligi, A. (2018) ICT global emissions footprint: Trends to 2040 & recommendations. *Age and Ageing*, 51.

Charnes, A., Cooper, W. W., & Rhodes, E. L. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, 2 (6), 429-444.

Chaudry, M., Jayasuriya, L., Blainey, S., Lovric, M., Hall, J., Russell, T., & Wu, J. (2022). The implications of ambitious decarbonisation of heat and road transport for Britain's net zero carbon energy systems. *Applied Energy*, 305.

Guo, X., Zhu, L., Fan, Y., & Xie, B. (2011). Evaluation of potential reductions in carbon emissions in Chinese provinces based on environmental DEA. Energy Policy, 2352–2360.

International Legal Materials, https://www.cambridge.org/core/journals/international-legal-materials/all-issues, 2016.

Kuo, T. C., Chen, G. H., Wang, M. L., & Ho, M. W. (2014). Carbon footprint inventory route planning and selection of hot spot suppliers. *Int. J. Production Economics*, 150, 125-139.

Martín-Gambo, M., & Iribarren, D. (2021). Coupled life cycle thinking and data envelopment analysis for quantitative sustainability improvement. *Methods in Sustainability Science*.

Mayers, K., Koomey, J., Hall, R., Bauer, M., France, C., & Webb, A. (2014). The Carbon Footprint of Games Distribution. *Research and Analysis*, 19 (3), 402-415.

Miller, S., Habert, G., Myers, R., & Harvey, J. (2021). Achieving net zero greenhouse gas emissions in the cement industryvia value chain mitigation strategies. *One Earth.*

Millot, A., Krook-Riekkola, A., & Maïzi, N. (2020). Guiding the future energy transition to net-zero emissions: lessons from exploring the differences between France and Sweden. Energy Policy, 139.

Tudor, C., & Sova, R. (2022). EU Net-Zero policy achievement assessment in selected members through automated forecasting algorithms. Int. J. Geo-Inf, 11 (4), 232.

Tyteca, D. (1996). On the Measurement of the Environmental Performance of Firms-A Literature Review and a Productive Efficiency Perspective. *Journal of Environmental Management*, 46.

Vlassopoulos, C. (2012). Competing definition of climate change and the post-Kyoto negotiations. *International Journal of Climate Change Strategies and Management*, 4 (1), 104-118.

Wu, W., & Skye, H. (2021). Residential net-zero energy buildings: Review and perspective. *Renewable and Sustainable Energy Reviews*, 142.

Zhou, P., Ang, B., & Poh, K. (2008). Measuring environmental performance under different. Energy Economics, 1-14.

Appendix. Input and output data of countries

Country	Primary Energy Consumption (twh)	GDP ($/capita)	CO_2 emission (tonnes)	N_2O emission (tonnes of CO_2-equivalents)
Albania	32.457	13656.59	4947485	1100000
Austria	415.964	55833.56	67936184	3850000
Belgium	740.942	51944.17	99432616	4580000
Bulgaria	207.051	23265.88	42255764	4160000
Cyprus	32.379	41521.92	6875935	370000
Czechia	471.76	40981.05	101012960	5320000
Denmark	188.006	57161.69	30955444	5170000
Estonia	60.212	36401.11	12380190	1400000
Finland	312.568	48582.59	42381840	5010000
France	2686.761	46017.77	316386780	38279999
Germany	3625.339	53929.64	707149950	34119999
Greece	331.338	29698.03	65756230	4270000
Hungary	272.143	32553.7	49234644	5700000
İceland	60.856	56935.77	3546263	370000
İreland	184.218	86650.01	37325664	9250000
İtaly	1790.808	42708.04	339233200	15350000
Lithonia	69	37166.21	13923306	3650000
Luxemburg	46.973	116518.3	9751728	380000
Malta	40.376	43950.62	1649193	50000
Montenegro	14	21533.93	2476864	150000
Nether	975.717	56784.04	153032800	8120000
North Macedonia	31	16773.04	7994848	540000
Norway	493.904	64385.01	42784000	3400000
Poland	1176.408	33185.16	318487680	21680000
Portugal	286.524	34945.66	47618828	2990000
Roma	382.183	29875.06	77030616	8940000
Russia	8304.408	27210.55	1692363400	65120003
Serbia	199.87	18306.79	44344496	3960000
Slovakia	182	31927.59	33776188	2020000
Slovenia	79	38946.68	14048142	810000
Spain	1554.292	40802.49	251825150	20070000
Sweden	622.722	52850.57	40982492	5640000
Switzerland	326.727	70944.35	36733064	2420000
Türkiye	1807.463	28197.25	401719740	34750000
United kingdom	2146.401	47368.59	364753280	28370001

A Bibliometric Review of Agent-Based Models Application in Economics

Nikola Zornić, Aleksandar Marković, Nebojša Avramović
University of Belgrade
Faculty of Organizational Sciences
Jove Ilića 154, 11000 Belgrade, Serbia
{nikola.zornic, aleksandar.markovic}@fon.bg.ac.rs, n.d.avramovic@gmail.com

Abstract. *Agent-based modeling and simulation (ABMS) is a computational modeling technique that has been used in various fields of science with growing interest. In economics, ABMS has been used to model a wide range of systems, from financial markets to transportation systems. The goal of this paper is to explore the usage of agent-based modeling and simulation in economics, using bibliometric analysis techniques on data from Web of Science platform. The authors collected data on ABMS usage in economics for over five thousand papers from Clarivate Analytics Web of Science databases in terms of publishing frequency, journal bibliometric data, field of application and productivity by country for the period 2000-2022. Journals with highest productivity within this topic are The Journal of Artificial Societies and Social Simulation, Physica A-Statistical Mechanics and its Applications, and PLOS ONE.*

Keywords. agent-based, modeling, simulation, economics, bibliometric analysis.

1 Introduction

The sheer volume of data available today is staggering, leading to the rapid development of statistical learning techniques. These techniques can be used to describe phenomena, draw conclusions, and create decision-making models. In order to yield the most of the data, more sophisticated methods of data analysis, like machine learning and statistical learning, are increasingly employed. Machine learning models can be used in combination with agent-based models to gain insights into data. This combination is particularly powerful because it allows for the modeling of complex systems with large amounts of data. The use of machine learning models in combination with agent-based models has been applied in various fields such as finance, transportation, and healthcare (Zornić & Marković, 2022).

Agent-based models (ABM) are a type of computational model that simulate the behavior of complex systems by breaking them down into individual agents that interact with each other and their environment. Each agent is characterized by specific attributes and follows a set of rules that govern its behavior. Agents can be programmed to have different levels of autonomy and decision-making capabilities, which allows for the modeling of a wide range of systems, from simple to highly complex. Researchers are using ABMS to analyze policy and resource usage changes in agriculture (Berger, 2001), distributed energy resource management for intelligent microgrids (Kumar Nunna & Doolla, 2013), exploring the impact of shared autonomous vehicles on urban parking demand (Zhang et al., 2015), analyse telecommunication company churn (Zornić & Marković, 2022), etc.

In 2008, Jean-Philippe Bouchaud (Bouchaud, 2008) argued that the financial crisis demanded a change in attitude in economics and financial engineering, emphasizing data and plausible but not exact arguments over axioms that were taken as unquestionable. Farmer and Foley, also advocated for the need of agent-based modeling in economy (Farmer & Foley, 2009). In "The End of Theory: Financial Crises, the Failure of Economics, and the Sweep of Human Interaction" Rick Bookstaber outlines four issues of human experience that conventional economics is not equipped to adequately address (Bookstaber, 2017).

The first issue is computational irreducibility. Computational irreducibility is a concept in computer science that refers to the inability to reduce the behavior of some systems to a mathematical description that would provide a way to predict its future behavior without having to simulate and observe the system in action. This poses a challenge to economists who need to be able to predict the outcomes of economic systems using analytical methods.

The second issue is emergence. Emergence occurs when the collective result of individual actions is qualitatively different from the actions of each individual. This means that one cannot predict the outcome of the entire system based on the behavior of

its members, as the overall system has properties that its components do not have.

The third issue is non-ergodicity. Non-ergodicity is a concept in statistical mechanics where a single trajectory does not reflect an isolated system as a whole, and the past does not necessarily determine the future. This is in contrast to ergodic processes, which follow the same probabilities today as they did in the past and will in the future.

The fourth issue is radical uncertainty. Radical uncertainty is defined as a surprise outcome or event that cannot be put into a probability distribution because it is outside of what is expected to occur. It is distinguished from risk, which can be calculated, by the fact that it is impossible to know all the information needed to calculate the odds. Through ABMs, it is possible to determine properties and patterns that could not be predicted by observing the individual agents due to the complexity of their interactions. According to Wooldridge and Jennings (Wooldridge & Jennings, 1995), a computational agent is an entity that is independent, socially capable, reactive, and proactive in its interactions with the environment in order to meet its objectives.

Economists use simulation methods to study economic dynamics due to the challenges posed by computational irreducibility, emergence, non-ergodicity, and radical uncertainty. Agent-based modeling and simulation (ABMS) is a computational modeling technique that has been increasingly used in economics due to its ability to create detailed, realistic models of a system. ABMS allows researchers to explore the behavior of complex systems and understand their emergent properties.

Bibliometric analysis is a method of exploring research activities and findings through data from research publications. Multiple bibliometric studies regarding ABMS can be found in literature. Some of them are focused mostly on ABMS (Niazi & Hussain, 2011), and others are focused on ABMS usage in specific field, like modeling COVID-19 pandemic (Tang et al., 2022), marketing (Romero et al., 2023), organization management (Wu et al., 2023), construction (Khodabandelu & Park, 2021), and economics (Zehra & Urooj, 2022).

Although Zehra and Urooj (2022) already conducted the bibliometric analysis of agent-based modeling in economics, they used SCOPUS database and collected information about 1,568 documents. The goal of this paper is to explore the usage of agent-based modeling and simulation in economics, using bibliometric analysis techniques using the data from Web of Science platform.

After outlining the need for agent-based modeling and simulation in economics, the paper is organized as follows. A methodology used for data collection and analysis is presented in section 2. A bibliometric review of ABMS application in economics is presented in section 3, including journals and citation analyses, and research area and keywords trends analyses. Finally, conclusions and some visions for the future research are provided in section 4.

2 Methodology

Data is collected from the Web of Science platform (Clarivate Analytics, 2022) (all sources), using the following search configuration (Zornić et al., 2023):

(TS = (agent-based OR individual-based OR multi-agent OR multiagent OR ABM) AND TS = (model* OR simulat*))*
AND TS = (econom)*
Indexes = Web of Science Core Collection
Timespan = All years

Data is collected for 5,426 scientific papers. For the sake of analysis 5,347 papers published in period 2000-2022 were selected. Data contains information such as authors, titles, source titles, publication year, keywords, categories, times cited.

Data is analyzed using descriptive statistics and aggregated results are presented. Additionally, citation network analysis is conducted using Gephi software.

Gephi is an open-source software for network visualization and analysis. It is mainly used for visualizing, manipulating, and exploring networks and graphs from raw edge and node graph data. It uses a 3D render engine to display large graphs in real-time and to speed up the data exploration (Bastian et al., 2009).

3 ABMS and Economics in Literature

Bibliometric analysis geared towards a review of literature productivity and an observation of the trends in agent-based modeling and simulation application in economics is presented in this section of the paper.

Figure 1 presents the trend in number of papers published yearly, starting with 19 in 2000, rising to 466 in 2020, and dropping to 431 in 2022. Popularity of specific scientific domain is known to be based on its number of citations (Niazi & Hussain, 2011). This can be observed in Figure 2. It should be noted that number of citations in the last two years is expected to be lower, as those papers are just published.

We also identified most popular papers, based on number of citations: “Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change” is taking the first place with 891 citations (Kremen et al., 2007); “An economic approach to social capital” is in second place with 685 citations (Glaeser et al., 2002); “Review on modeling and simulation of interdependent critical infrastructure systems” is in the third place with 648 citations (Ouyang, 2014).

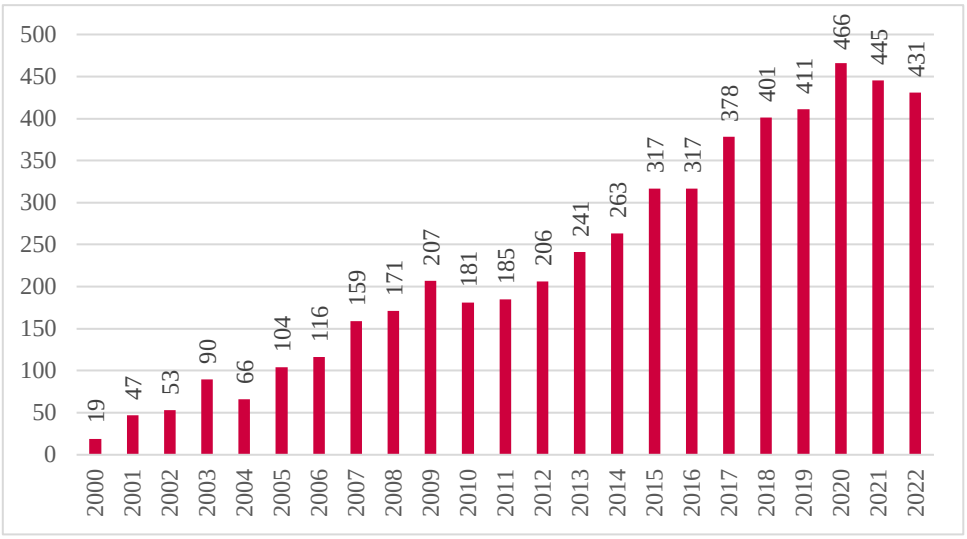


Figure 1. Papers published per year

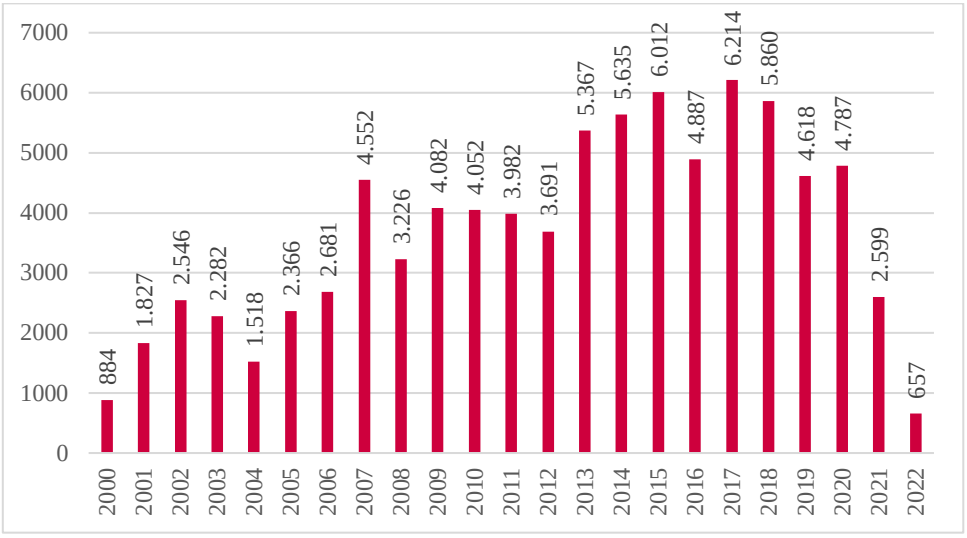


Figure 2. Number of citations per year

3.1 Journals and Citations Analyses

Those 5,347 papers are published in 2430 sources, including journals and conference proceedings. Table 1 presents 18 journals with at least 20 papers with ABMS and economics in topic.

The top contributor is *The Journal of Artificial Societies and Social Simulation* with 120 papers published, followed by *Physica A: Statistical Mechanics and its Applications* and *PLOS ONE*. Since its first issue in 1998, *Journal of Artificial Societies and Social Simulation* has been one of the leading references for readers interested in social simulation and the application of computer simulation in the social sciences (Marković & Zornić, 2016). In addition to journals that are expected to be on the list, journals specialized in Simulation and Modeling, and Economics (*Journal of Artificial Societies and Social Simulation, Journal of Economic Interaction and Coordination, Computational Economics, Journal of Economic Dynamics & Control, Journal of Evolutionary Economics, Journal of Economic Behavior & Organization*, etc.), there are journals specialized in other scientific fields, like: Ecology (*Ecological Modelling* and *Environmental Modelling & Software*), and other multidisciplinary jorunals. Table 1 also reports impact factor (IF), IF without journal self-citations, and percentage of change in IF after the self-citations have been excluded. Data about journals and IF is collected from Clarivate Analytics Journal Citation Reports (Clarivate Analytics, 2023). We can see that the journal with the highest IF is *Applied Energy* (11.20), while the journal with the highest decrease in IF after the self-citations were excluded is *Energies* (-21.88%). We can also see that all listed journals have IF higher than 1. The lowest IF has *Journal of Economic Interaction and Coordination* – 1.10, dropping to 0.90 when self-citations are excluded.

Table 1. Journals accounting for at least 30 papers with ABMS and economics in topic

Journal	Papers	Journal IF (2022)	IF Without Journal Self Cites (2022)	% IF Change Without Self Cites
JASSS-THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION	120	4.20	3.80	-9.52%
PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS	75	3.30	3.00	-9.09%
PLOS ONE	71	3.70	3.50	-5.41%
JOURNAL OF ECONOMIC INTERACTION AND COORDINATION	59	1.10	0.90	-18.18%
COMPUTATIONAL ECONOMICS	58	2.00	1.70	-15.00%
SUSTAINABILITY	54	3.90	3.10	-20.51%
JOURNAL OF ECONOMIC DYNAMICS & CONTROL	53	1.90	1.70	-10.53%
JOURNAL OF EVOLUTIONARY ECONOMICS	53	1.80	1.70	-5.56%
JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION	51	2.20	2.00	-9.09%
JOURNAL OF CLEANER PRODUCTION	49	11.10	10.10	-9.01%
APPLIED ENERGY	46	11.20	10.00	-10.71%
ECOLOGICAL MODELLING	44	3.10	2.90	-6.45%
ENERGIES	43	3.20	2.50	-21.88%
ECOLOGICAL ECONOMICS	42	7.00	6.60	-5.71%
ENVIRONMENTAL MODELLING & SOFTWARE	41	4.90	4.50	-8.16%
IEEE ACCESS	34	3.90	3.60	-7.69%
JOURNAL OF STATISTICAL MECHANICS-THEORY AND EXPERIMENT	31	2.40	2.10	-12.50%
ENERGY POLICY	30	9.00	8.50	-5.56%

Additional analysis of citations is conducted using network visualization and exploration software Gephi (Bastian et al., 2009). Total citation count for the selected journals has been collected from the Clarivate Analytics Journal Citation Reports (Clarivate Analytics, 2023) for the period until 2022. The obtained results are displayed in **Error! Reference source not found.**, where node size is determined by the number of papers with ABMS and Economics in topic. For example, it is noticeable that *The Journal of Artificial Societies and Social Simulation, Physica A-Statistical Mechanics and Its Applications*, and *PLOS ONE* have the highest number of papers. The weight of the line is defined as the sum of percentages of citations acquired from the other observed journal. For example, *Ecological Economics* got 2572 citations from *Sustainability* and has 36,381 citations in total, which means *Ecological Economics* got 7.07% of citations from *Sustainability*. In the other direction, *Sustainability* got 43 citations from *Ecological Economics* and has 187,953 citations in total, which means *Sustainability* got 0.11% of citations from *Ecological Economics*.

The weight of line is determined by the sum of these percentages (7.07% + 0.11%). The size of the arrow is defined separately for each side in a similar manner. Similarly, we can analyze the remaining connections between journals.

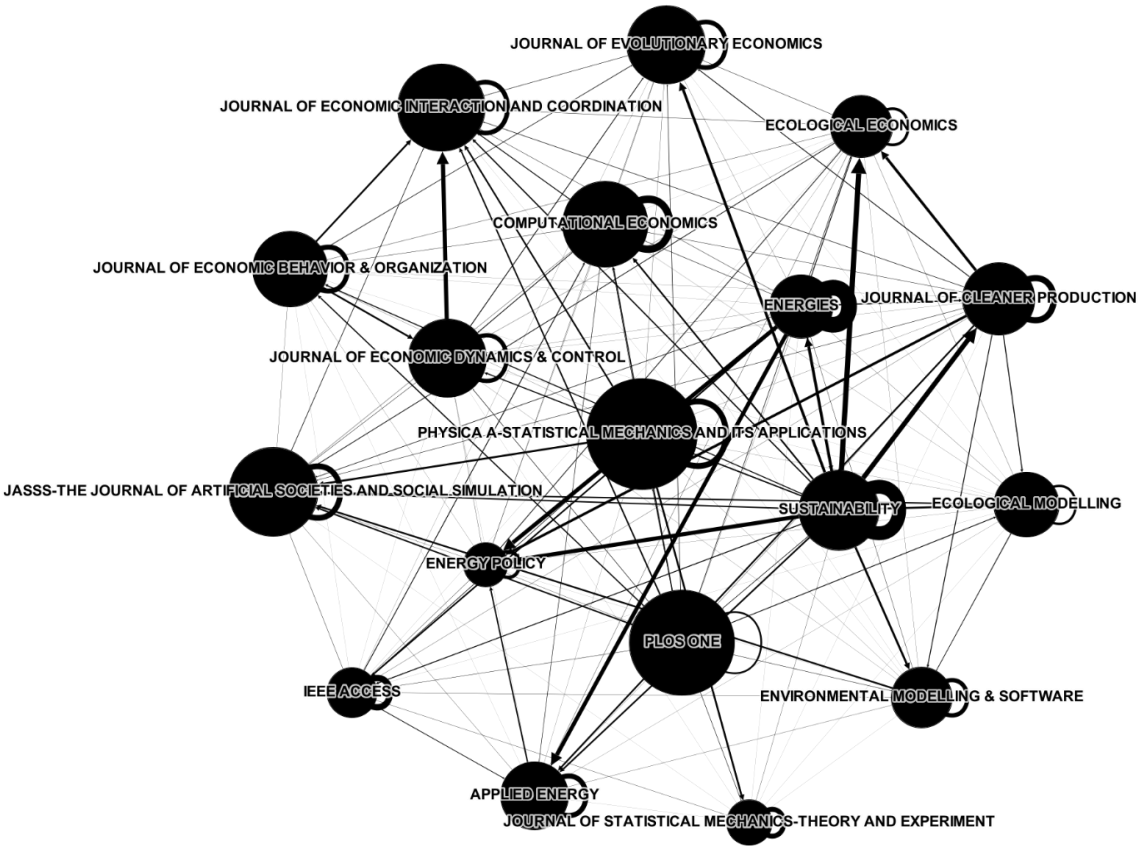


Figure 3. Network of citations among selected journals

3.2 Categories and Countries Analyses

Analyzing the categories in which ABMS are applied we can provide one more evidence that ABMS are used for modeling various economic problems. Data have been retrieved from Web of Science (Clarivate Analytics, 2022). Table 2 presents the most popular Web of Science categories for the papers with ABMS and economics in topic (with at least 100 papers). As expected, the first two are directly related to economics and ABMS. Engineering Electrical Electronic and Environmental Sciences are following, together with different categories of Computer Sciences.

Table 2. Most popular Web of Science categories for papers with ABMS and economics in topic

Web of Science Categories	Count
Economics	881
Computer Science Artificial Intelligence	693
Engineering Electrical Electronic	612
Environmental Sciences	475
Computer Science Interdisciplinary Applications	461
Computer Science Theory Methods	461
Computer Science Information Systems	429
Energy Fuels	343
Environmental Studies	309
Automation Control Systems	250
Operations Research Management Science	249
Mathematics Interdisciplinary Applications	246
Management	240
Ecology	227
Multidisciplinary Sciences	224
Green Sustainable Science Technology	216
Social Sciences Interdisciplinary	208
Engineering Environmental	158
Computer Science Software Engineering	148
Water Resources	146
Business	139
Telecommunications	130
Physics Multidisciplinary	129
Mathematics Applied	127
Engineering Industrial	115
Engineering Civil	110
Transportation Science Technology	107
Geography	106

We also analyzed the Web of Science KeyWords Plus (Table 3). KeyWords Plus is a feature in Clarivate databases that uses an algorithm to identify words or

phrases that frequently appear in the titles of an article's references, but not in the title of the article itself. Disregarding the expected ones (such as *model, dynamics, systems, simulation*, etc.), other keywords are describing the field of application (such as *Management, Economics, Market*, etc.).

Table 3. Most popular Web of Science KeyWords Plus for papers with ABMS and economics in topic

Web of Science Keywords Plus	Count
MODEL	498
DYNAMICS	348
MANAGEMENT	295
SYSTEMS	291
SIMULATION	281
BEHAVIOR	209
ECONOMICS	191
MODELS	189
IMPACT	177
OPTIMIZATION	172
GROWTH	143
EVOLUTION	135
DESIGN	125
SYSTEM	121
FRAMEWORK	118
MARKET	116
POLICY	113
CLIMATE-CHANGE	108
NETWORKS	108
RISK	106
DECISION-MAKING	105
LAND-USE	100

In Table 4 the most productive countries using ABMS in economics are observed (with at least 100 papers). USA and People's Republic China are leading the way accounting for over 40% of total papers published. Five European countries are following (England, Germany, Italy, France, Netherlands) with another 40% of total papers published. Here should be noted that one paper can be accounted for two or more countries (based on author affiliation).

Table 4. Countries with at least 50 papers with ABMS and economics in topic

Countries/Regions	Count
USA	1227
PEOPLES R CHINA	825
ENGLAND	527
GERMANY	500
ITALY	402
FRANCE	367
NETHERLANDS	296
AUSTRALIA	251
CANADA	212
SPAIN	192
JAPAN	180
RUSSIA	149
AUSTRIA	124
BRAZIL	122
SWITZERLAND	116
INDIA	95
PORTUGAL	89
SCOTLAND	83
SOUTH KOREA	82
SWEDEN	79
IRAN	74
TAIWAN	72
POLAND	71
DENMARK	66
CZECH REPUBLIC	65
SINGAPORE	61
BELGIUM	58
SOUTH AFRICA	50

4 Conclusion

In this paper, agent-based modeling and simulation application in economics is reviewed. The comprehensive bibliometric data has been collected and analyzed, including publishing trend, bibliometric analysis of most productive journals, research category, keywords and productivity by country. Bibliometric journal analysis includes IF, IF without journal self-citations, and citation network analysis.

This paper presents bibliometric analysis conducted using information from Web of Science platform about 5,347 scientific papers. Similar study can be found in literature, where authors presented bibliometric analysis of 1,568 documents from SCOPUS database.

We can conclude just at a glance that ABMS usage in economics is growing with the course of time, from 19 papers published in 2000, this number increased to 438 in 2022. Journals with highest productivity within this topic are The Journal of Artificial Societies and Social Simulation, Physica A-Statistical Mechanics and its Applictions, and PLOS ONE, while the most productive countries are USA, and People's Republic China, followed by European countries England, Germany, Italy, France, and Netherlands.

This paper presents a basis for the future research of most influential authors in the field of using agent-based modeling and simulation in economics, as well as systematic review of most influential papers.

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. Proceedings of the International AAAI Conference on Web and Social Media, 3(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

Berger, T. (2001). Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. Agricultural Economics, 25(2–3), 245–260. https://doi.org/10.1111/j.1574-0862.2001.tb00205.x

Bookstaber, R. (2017). The End of Theory: Financial Crises, the Failure of Economics, and the Sweep of Human Interaction. Princeton University Press.

Bouchaud, J.-P. (2008). Economics needs a scientific revolution. Nature, 455(7217), 1181–1181. https://doi.org/10.1038/4551181a

Clarivate Analytics. (2022). Web of Science. https://apps.webofknowledge.com

Clarivate Analytics. (2023, June 30). Journal Citation Reports.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. Nature, 460(7256), 685–686. https://doi.org/10.1038/460685a

Glaeser, E. L., Laibson, D., & Sacerdote, B. (2002). An Economic Approach to Social Capital. The Economic Journal, 112(483), F437–F458. https://doi.org/10.1111/1468-0297.00078

Khodabandelu, A., & Park, J. (2021). Agent-based modeling and simulation in construction. Automation in Construction, 131, 103882. https://doi.org/10.1016/j.autcon.2021.103882

Kremen, C., Williams, N. M., Aizen, M. A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., Packer, L., Potts, S. G., Roulston, T., Steffan-Dewenter, I., Vázquez, D. P., Winfree, R., Adams, L., Crone, E. E., Greenleaf, S. S., Keitt, T. H., Klein, A.-M., Regetz, J., & Ricketts, T. H. (2007). Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. Ecology Letters, 10(4), 299–314. https://doi.org/10.1111/j.1461-0248.2007.01018.x

Kumar Nunna, H. S. V. S., & Doolla, S. (2013). Multiagent-Based Distributed-Energy-Resource Management for Intelligent Microgrids. IEEE Transactions on Industrial Electronics, 60(4), 1678–1687. https://doi.org/10.1109/TIE.2012.2193857

Marković, A., & Zornić, N. (2016). Trends in the Application of Agent-Based Modeling and Simulation. Central European Conference on Information and Intelligent Systems, 65–70.

Niazi, M., & Hussain, A. (2011). Agent-based computing from multi-agent systems to agent-based models: a visual survey. Scientometrics, 89(2), 479–499. https://doi.org/10.1007/s11192-011-0468-9

Ouyang, M. (2014). Review on modeling and simulation of interdependent critical infrastructure systems. Reliability Engineering & System Safety, 121, 43–60. https://doi.org/10.1016/j.ress.2013.06.040

Romero, E., Chica, M., Damas, S., & Rand, W. (2023). Two decades of agent-based modeling in marketing: a bibliometric analysis. Progress in Artificial Intelligence. https://doi.org/10.1007/s13748-023-00303-y

Tang, J., Vinayavekhin, S., Weeramongkolkul, M., Suksanon, C., Pattarapremcharoen, K., Thiwathittayanuphap, S., & Leelawat, N. (2022). Agent-Based Simulation and Modeling of COVID-19 Pandemic: A Bibliometric Analysis. Journal of Disaster Research, 17(1), 93–102. https://doi.org/10.20965/jdr.2022.p0093

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: theory and practice. The Knowledge Engineering Review, 10(02), 115–152. https://doi.org/10.1017/S0269888900008122

Wu, J., Ohya, T., & Sekiguchi, T. (2023). Applications of agent-based modeling and simulation in organization management: a quarter-century review through bibliometric mapping (1998–2022). Computational and Mathematical Organization Theory. https://doi.org/10.1007/s10588-023-09381-8

Zehra, A., & Urooj, A. (2022). A Bibliometric Analysis of the Developments and Research Frontiers of Agent-Based Modelling in Economics. Economies, 10(7), 171. https://doi.org/10.3390/economies10070171

Zhang, W., Guhathakurta, S., Fang, J., & Zhang, G. (2015). Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. Sustainable Cities and Society, 19, 34–45. https://doi.org/10.1016/j.scs.2015.07.006

Zornić, N., & Marković, A. (2022). A methodological framework for the integration of machine learning algorithms into agent-based simulation models. JUCS - Journal of Universal Computer Science, 28(5), 540–562. https://doi.org/10.3897/jucs.73924

Zornić, N., Marković, A., & Avramović, N. (2023). Agent-Based Models in Economics: A Review of Applications. Work in Progress Paper Presented at Challenges of Europe 2023 Conference.

Systems for Smart Mobility

Markus Schatten, Neven Vrček
A Method for Intruder UAV Pilot Localization based on Neural Networks using Intercepting Drone Constellations

A Method for Intruder UAV Pilot Localization based on Neural Networks using Intercepting Drone Constellations

Markus Schatten, Neven Vrček
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
markus.schatten, neven.vrcek@foi.unizg.hr

Abstract. *In this paper we introduce a method for the localization of intruder drone pilot using an intercepting 3-drone constellation. Using datasets collected during field experiments we have analyzed various machine learning (ML) methods including linear regression, decision tree, random forest, gradient boosting, support vector, and neural networks in order to implement a predictive model that will be able to predict the distance of each intercepting drone to the intruders drone pilot based on various parameters, most prominently Received Signal Strength Indicator (RSSI). Using this predicted distance and Global Positioning System (GPS) coordinates of the intercepting drones we are able to trilaterate the coordinates of the drone pilot. The analysis has shown that the neural network model gives the best results with an mean squared error of 0.12185.*

Keywords. pilot localization, drone constellations, intruder interception, neural network, unmanned aerial vehicles

1 Introduction

Unmanned Aerial Vehicles (UAVs), commonly known as drones, have seen a significant increase in popularity and usage in recent years (Idrissi et al., 2022). They are employed in a wide range of applications, from aerial photography and videography to delivery services and agricultural monitoring (Schatten, 2015). However, the proliferation of drones has also raised safety and security concerns. One of the most pressing issues is the unauthorized or malicious use of drones in sensitive areas such as airports, military bases, and public spaces. The ability to locate the pilot of an intruding drone is crucial in mitigating these risks and responding effectively to potential threats.

In this context, this paper presents a novel method for intruder drone pilot localization using an intercepting 3-drone constellation. Our approach leverages machine learning techniques to predict the distance of each intercepting drone to the intruder's drone pilot based on various parameters, most prominently Received Signal Strength Indicator (RSSI). By combining these predicted distances with the GPS coordinates of the intercepting drones, we are able to trilaterate the coordinates of the drone pilot.

This work is part of the ORKAN project - Unmanned Aerial Vehicle policy ecosystem[1], a comprehensive initiative aimed at developing advanced methods and technologies for drone detection, tracking, and neutralization. Our previous work (Magdalenić et al., 2022; Tomaš and Posarić, 2022) has made significant contributions to this area. In (Tomaš and Posarić, 2022), we proposed a novel approach to WiFi access point localization in an urban multi-sensor environment. This approach significantly influenced our current research allowing us to localize access points on UAVs. In (Magdalenić et al., 2022), we further improved this approach by integrating a a specially designed hardware device to allow for capturing (intruder) drone signals by using a drone constellation.

In our analysis, we compared several machine learning methods, including linear regression, decision tree, random forest, gradient boosting, support vector, and neural networks (multilayer perceptron). Our results show that the neural network model provides the best performance, with a mean squared error of 0.12185. This finding underscores the potential of machine learning, and neural networks in particular, in enhancing the capabilities of drone interception systems and improving safety and security in drone-affected areas.

The remainder of this paper is organized as follows. Section 2 provides a detailed description of the proposed method. Section 3 presents the results of our analysis and comparison of different machine learning models. Finally, Section 4 concludes the paper and outlines directions for future work.

2 Related Work

Apart from our research, to the best of our knowledge, there have been very few studies which try to develop methods for drone pilot localization. For example,

[1] https://orkan.foi.hr

(Zhen, 2019) propose a direction of arrival (DOA) estimation algorithm that uses support vector regression and integrates DOAs acquired with array sensors on a reconnaissance aircraft to estimate the location of illegal unmanned aerial vehicle operators. The algorithm performs well in circumstances where signals impinge on the sensor array with a small-angle interval, as well as in conditions of small samples and low signal-to-noise ratio.

On the other hand, (Mashhadi et al., 2020) presents a method for autonomously locating drone operators by tracking the drone's path in the sky using a deep neural network. The model, trained on a realistic simulation environment, achieved 73% accuracy in predicting the operator's location, offering a promising alternative to traditional methods that use RF sensors, which have certain limitations.

Basak and Scheers, 2018 discusses the development of a passive radio system that can detect and localize mini remotely piloted aircraft systems (RPAS) and their operators in three dimensions. The system uses Goodness-of-Fit (GoF) based spectrum sensing for signal detection and the MUSIC algorithm for direction of arrival (DoA) estimation, offering a low-cost, highly sensitive solution capable of detecting multiple drones and operators simultaneously across a wide frequency range.

Our approach to drone pilot localization involves the use of a three-drone constellation that follows an intruder drone and collects signal strength and GPS data. Using this approach we are able to achieve a more accurate result in pilot localization based on field experiment data using machine learning models.

3 Methodology

As already mentioned six machine learning methods have been tested for performance. Bellow is a short overview of each of the methods.

3.1 Linear Regression

Linear regression is a foundational statistical method used to model the relationship between a dependent variable y and one or more independent variables X. The primary goal is to find the best fit straight line that accurately predict the output values within a range. The linear equation is given by:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

Where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients of the independent variables, and ϵ is the error term. The coefficients are estimated using methods like the least squares.

3.2 Decision Tree

A decision tree is a flowchart-like structure where each internal node represents a feature (or attribute), each branch represents a decision rule, and each leaf node represents an outcome. The primary goal of using a decision tree is to create a training model that can predict the class or value of the target variable by learning simple decision rules inferred from prior data.

The decision to split at each node is based on some criterion, such as maximizing the information gain or minimizing the Gini impurity. The tree is constructed in a top-down recursive manner.

3.3 Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees to produce a more accurate and stable prediction. It's essentially a collection of decision trees trained with subsets of the dataset. The main principle behind the ensemble model is that a group of weak learners come together to form a strong learner. In the case of random forests, the model creates an "ensemble" of decision trees from a random subset of the data and aggregates them to produce a final result.

3.4 Gradient Boosting

Gradient Boosting is an iterative ensemble method that adjusts for the errors of the previous models in the ensemble. It builds the model in a stage-wise fashion, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. At each stage, the model introduces a new tree that corrects the errors of the previously built ensemble. This is done by fitting the new tree to the negative gradient of the loss function, effectively moving closer to the optimal model.

3.5 Support Vector Machines

Support Vector Machines (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. It operates by finding the hyperplane that best divides a dataset into classes. The best hyperplane is the one that represents the largest separation, or margin, between the two classes. The equation of the hyperplane is given by:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

Where $\mathbf{w}$ is the normal vector to the hyperplane and b is the bias. The objective of the SVM algorithm is to find the optimal values for $\mathbf{w}$ and b.

3.6 Neural Networks - Multilayer Perceptron

Neural Networks, specifically the Multilayer Perceptron (MLP), consist of layers of interconnected nodes

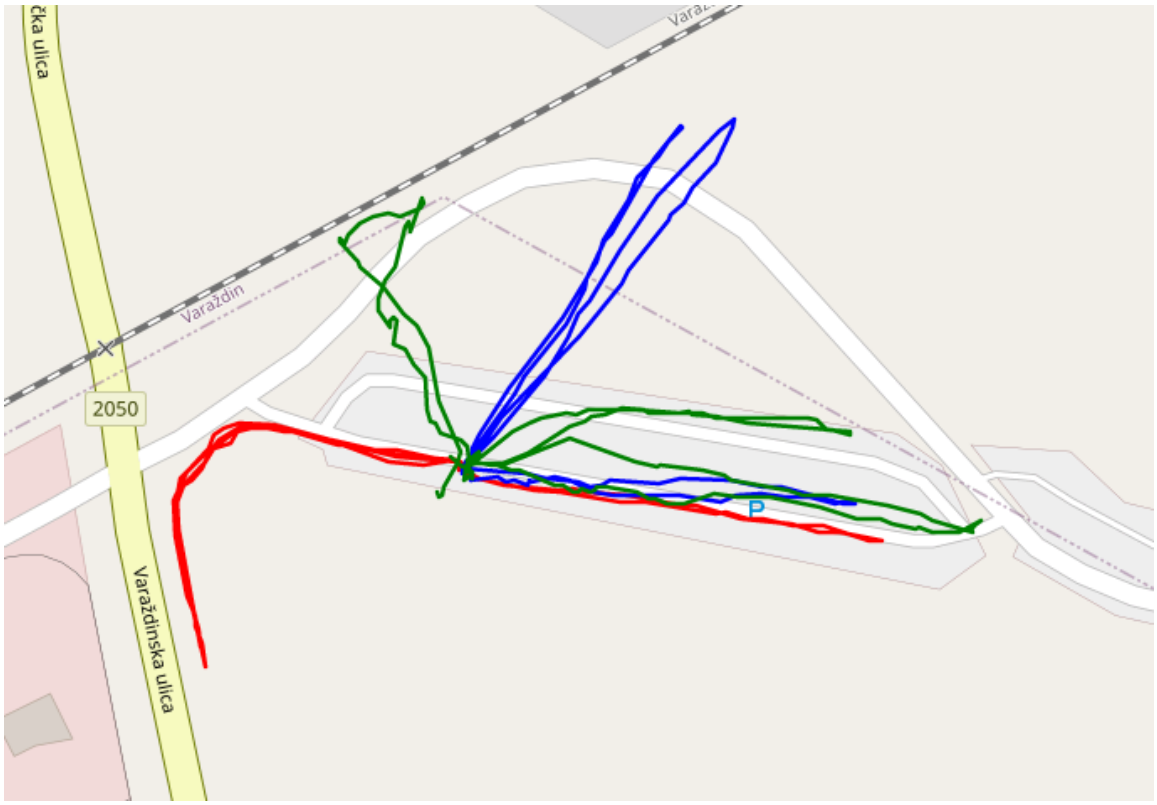


Figure 1: Visualization of the GPS data collected during one of field experiments (each color representing one drone of the 3-drone constellation)

or "neurons". Each connection between neurons has a weight, and each neuron has an activation function. The primary goal of a neural network is to receive inputs, process them in hidden layers using weights that are adjusted during learning, and produce an output. The output y of a neuron is given by:

$$y = f\left(\sum_i w_i x_i + b\right)$$

Where f is the activation function, w_i are the weights, x_i are the inputs, and b is the bias. The network "learns" by adjusting the weights to minimize the error in its predictions.

4 Experiments

The 3-drone constellation is designed such that each drone in the constellation is equipped with a specially designed hardware device that allows for capturing the intruder drone's signals.

The localization process can be divided into two main steps: distance prediction and trilateration.

4.1 Distance Prediction

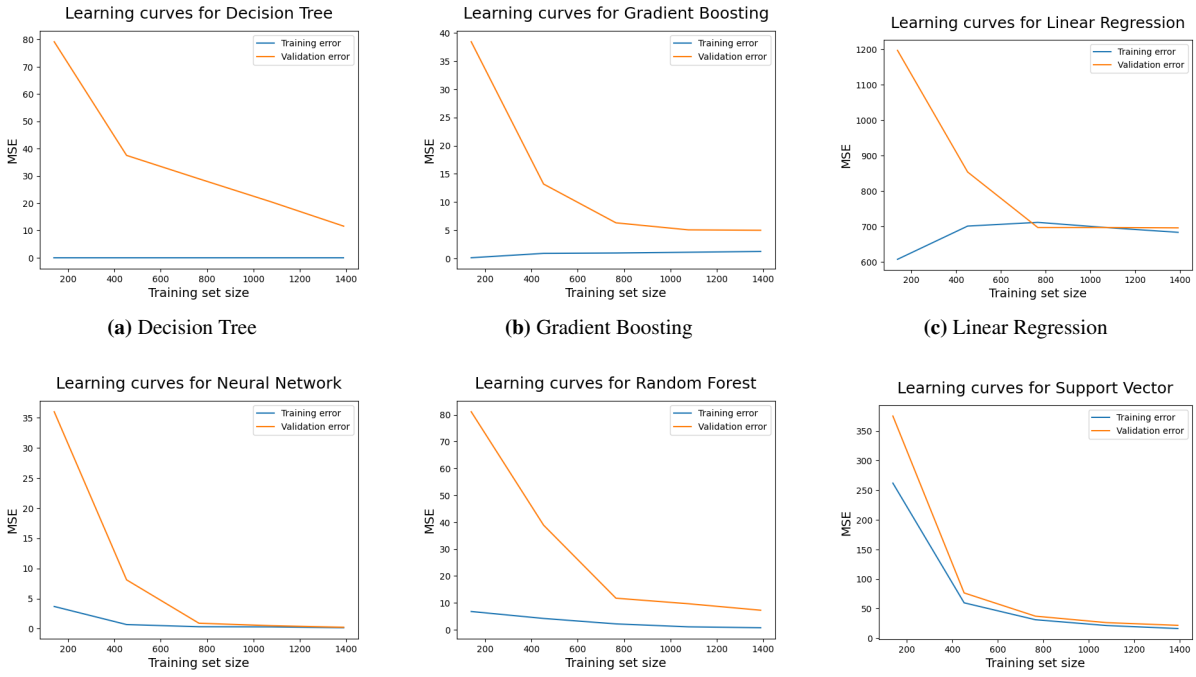
The first step in the localization process is to predict the distance between the pilot and each drone in the constellation. For this purpose, we use a machine learning model that takes as input the collected signal strength and GPS data and outputs a predicted distance. In our case, we found that a neural network model provides the best performance.

The neural network model is trained on a dataset collected during field experiments (see figure 1), with the signal strength and GPS data as the input features and the actual distance as the target variable.

The dataset consists of signal strength and GPS data collected by a three-drone constellation following an intruder drone (a total of 2174 datapoints). The dataset includes the following features: time, number of satellites, longitude, latitude, altitude, Horizontal Dilution of Precision (HDOP), maximum RSSI, average RSSI, and distance. The target variable is the distance between each drone and the pilot.

The machine learning models were implemented and trained using Python in Google Colab, an online platform that allows for the execution of Python code in the cloud. The Python libraries used in the implementation include pandas for data manipulation, numpy for numerical computations, and scikit-learn for machine learning.

In order to determine the best performing machine learning model, we have trained six ML models. For each machine learning model, the dataset was first split into a training set and a test set using a 80-20 split. The models were then trained on the training set and evaluated on the test set.



(a) Decision Tree **(b)** Gradient Boosting **(c)** Linear Regression

(d) Neural Network **(e)** Random Forest **(f)** Support Vector

Figure 2: Training curve visualizations of the machine learning models.

Model	Best Parameters	MSE
Linear Regression	N/A	634.05166
Decision Tree	{'max_depth': 20, 'min_samples_split': 2}	6.93033
Random Forest	{'max_depth': None, 'min_samples_split': 2, 'n_estimators': 200}	2.11822
Gradient Boosting	{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 200}	3.23053
Support Vector	{'C': 10.0, 'kernel': 'rbf'}	11.13768
Neural Network	{'activation': 'relu', 'hidden_layers': (50, 50), 'learning_rate_init': 0.01}	0.12185

Table 1: Model performance comparison with best parameters

The linear regression model was implemented using the LinearRegression class in scikit-learn. The decision tree model was implemented using the DecisionTreeRegressor class, and the random forest model was implemented using the RandomForestRegressor class. The gradient boosting model was implemented using the GradientBoostingRegressor class, and the support vector model was implemented using the SVR class. The neural network model was implemented using the MLPRegressor class.

The hyperparameters of each model were tuned using grid search with cross-validation. The performance of each model was evaluated using the mean squared error (MSE) between the predicted distances and the actual distances. The individual training curves of each model are shown on figure 2.

4.2 Trilateration

The second step in the localization process is to estimate the GPS coordinates of the pilot based on the predicted distances. This is done using a trilateration approach, which is a method used to determine the location of a point by measuring distances to it from known points at fixed distances apart.

Given the GPS coordinates $(\text{lat}_i, \text{lon}_i)$ and the predicted distance d_i for each drone i in the constellation, we first calculate the bearing θ_{ij} between each pair of drones i and j using the formula:

$$\theta_{ij} = \text{atan2}(\text{lon}_j - \text{lon}_i, \text{lat}_j - \text{lat}_i) \tag{1}$$

We then calculate the destination point $(\text{lat}_{ij}, \text{lon}_{ij})$ for each drone i in the direction of each other drone j at a distance d_i using the formula:

$$(\text{lat}_{ij}, \text{lon}_{ij}) = \text{destination_point}(\text{lat}_i, \text{lon}_i, \theta_{ij}, d_i) \tag{2}$$

Finally, we calculate the centroid of the destination points as the estimated GPS coordinates of the pilot:

$$(\text{lat}_{\text{pilot}}, \text{lon}_{\text{pilot}}) = \frac{1}{3}\sum_{i=1}^{3}(\text{lat}_{ij}, \text{lon}_{ij}) \tag{3}$$

Figure 3: Detected pilot location

This process is repeated for each time step, and the final estimate of the pilot's coordinates is obtained by taking the mean of all the estimates.

5 Results

The performance of the ML models varied significantly (see table 1). The neural network model achieved the best performance with an MSE of 0.121846, followed by the random forest model with an MSE of 2.118221. The gradient boosting model achieved an MSE of 3.230534, and the decision tree model achieved an MSE of 6.930330. The support vector model and the linear regression model achieved higher MSEs of 11.137681 and 634.051665, respectively.

These results indicate that the neural network model provides the most accurate predictions of the distance between each drone and the pilot. This model can therefore be used to effectively trilaterate the coordinates of the drone pilot. Figure 3 shows a visualization of the detected pilots coordinates based on the field experiment data.

6 Conclusion

In this paper, we presented a novel method for localizing the pilot of an intruding drone using an intercepting 3-drone constellation. The method leverages machine learning techniques to predict the distance of each intercepting drone to the intruder's drone pilot based on various parameters, most prominently RSSI. By combining these predicted distances with the GPS coordinates of the intercepting drones, we are able to trilaterate the coordinates of the drone pilot.

Our analysis showed that the neural network model provides the best performance, with a mean squared error of 1.994414. This finding underscores the potential of machine learning, and random forests in particular, in enhancing the capabilities of drone interception systems and improving safety and security in drone-affected areas.

However, it is important to note that the proposed method is limited to scenarios where the intruder drone is operated using WiFi. This is a constraint that could limit the applicability of the method in some contexts. Nevertheless, the method offers significant benefits in terms of its ability to accurately and efficiently locate the pilot of an intruding drone, which is crucial for mitigating risks and responding effectively to potential threats.

The proposed method has a wide range of potential applications, including in airports, military bases, and public spaces, where unauthorized or malicious use of drones is a major concern. It could also be used in other contexts where the ability to locate the operator of a drone is important, such as in search and rescue operations or in the monitoring of wildlife.

In conclusion, the proposed method represents a significant contribution to the field of drone pilot localization. It offers a practical and effective solution to a pressing problem, and has the potential to significantly enhance safety and security in a wide range of contexts.

Acknowledgement

This work has been fully supported by the Croatian Science Foundation under the project IP-2019-04-4864.

References

Basak, S., & Scheers, B. (2018). Passive radio system for real-time drone detection and doa estimation. *2018 International Conference on Military Communications and Information Systems (ICMCIS)*, 1–6.

Idrissi, M., Salami, M., & Annaz, F. (2022). A review of quadrotor unmanned aerial vehicles: Applications, architectural design and control algorithms. *Journal of Intelligent & Robotic Systems*, *104*(2), 22.

Magdalenić, I., Vrček, N., Tomaš, B., & Guðmundsson, K. S. (2022). Localization of drone operator using unmanned aerial vehicles: Preliminary research. *Central European Conference on Information and Intelligent Systems*, 335–339.

Mashhadi, E., Oren, Y., & Weiss, G. (2020). Can the operator of a drone be located by following the drone's path? *Cyber Security Cryptography and Machine Learning: Fourth International Symposium, CSCML 2020, Be'er Sheva, Israel, July 2–3, 2020, Proceedings 4*, 85–93.

Schatten, M. (2015). *Multi-agent based traffic control of autonomous unmanned aerial vehicles* (tech. rep.). Artificial Intelligence Laboratory, University of Zagreb.

Tomaš, B., & Posarić, L. (2022). Wifi access point localization in urban multi-sensor environment. *Central European Conference on Information and Intelligent Systems*, 443–448.

Zhen, J. (2019). Localization of unmanned aerial vehicle operators based on reconnaissance plane with multiple array sensors. *IEEE Access*, *7*, 105354–105362.

Trends in Software Engineering

Marija Kuštelega, Renata Mekovec
A Systematic Analysis of Requirements Elicitation Problems and Challenges

Donata Szombathelyi, Tomislav Jakopec, Josipa Selthofer
The impact of using ChatGPT on employees of Croatian IT Companies

Ahmed Shareef, Boris Tomaš, Neven Vrček
Semantic Interoperability of Digital Twins in Smart Cities

Nadica Hrgarek Lechner
Using Card Sorting to Classify Elements of the Security Risk Management Program for Medical Devices

Dijana Peras, Zlatko Stapić, Mislav Matijević
AI techniques and tools in Agile Software Development: Preliminary research

Aránzazu San Juan Llano, Thi Mai Trang Nguyen, Daniel Rodríguez, Matthieu Chardy
POEMA: a Personal Cloud for Inclusive Education

Luis de Marcos, Daniel Rodríguez García, José-María Gutiérrez-Martínez, Adrián Domínguez-Díaz, Sergio Caro Álvaro
Recommendation and Prediction in a Microservice Web Application for Cyber Ranges

William Steingartner, Valerie Novitzká
Semantic methods in our research and teaching

A Systematic Analysis of Requirements Elicitation Problems and Challenges

Marija Kuštelega, Renata Mekovec
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{mkustele1, renata.mekovec}@foi.hr

Abstract. *Adequate research and planning for the elicitation process are thus essential to reduce the possibility of major mistakes hampering a project. The elicitation examines how firmly stakeholders are engaged and encouraged to collaborate, as well as how competing opinions are motivated to come to an agreement. Because it is responsible for bringing project needs into focus, elicitation is the cornerstone of each project. One of the most typical causes of project failures and abandonment, which have a negative impact on the bottom line, is elicitation errors. Some of the most problematic aspects are a lack of user contact, poor specifications, and changes in needs. The goal of this research is to conduct a literature review to identify the most prevalent problems and challenges faced during the request elicitation process, as well as the most commonly used demand elicitation methodologies.*

Keywords. requirements elicitation, challenges, requirements elicitation methods

1 Introduction

The Standish Group study analysis reveals concerning data about the success of IT projects in the development of software solutions (Clancy, 1995). They surveyed IT executive managers to learn why projects succeed and what problems they face. The sample included 365 respondents from large, medium, and small businesses in important industrial areas such as banking, manufacturing, and retail, among others. Despite the fact that the survey was conducted in the 1990s, the tendency toward ad hoc, spontaneous software development is still noticeable. Only 16.2% of projects were finished on schedule and within the allocated budget. The IT management's approach highlights several essential characteristics that are critical for the project's success, such as user participation, executive support, clear explanations of user requirements, and proper planning. Lack of user involvement, incomplete requirements, and changes in requirements, on the other hand, are mentioned as some of the most difficult factors. The majority of the issues that arise throughout this process are connected to insufficiently well-executed early phases of software development, such as requirements elicitation.

Requirements elicitation can be placed in the discovery phase of the software development life cycle (Apriorit.com, 2022). The discovery phase is the first stage of the software development life cycle, during which the primary focus is on research with the goal of getting to know the consumers, creating clear objectives, and determining specifications for requirements (Khalimonchuk, 2022).

The practice of gathering user requirements with the goal of identifying and understanding the true needs of users is referred to as requirements elicitation (Apriorit.com, 2022). Incomplete and inaccurate requirements have been identified as one of the primary causes of project failure (Alflen & Prado, 2021). According to a Pew Research study, 4% to 7% of online survey respondents provide false responses, making it even more difficult to extract true customer requirements (Conrad, 2020). As a result, uncovering obstacles and challenges is critical in order to adopt suitable requirement elicitation techniques.

The aim of this work is to undertake a literature review to identify the most prevalent problems and challenges that arise throughout the request elicitation process, as well as the most commonly used requirement elicitation techniques.

The following research questions (RQs) were developed for this purpose:

RQ1: What problems and challenges arise during the requirement elicitation process?

RQ1.1: How are common problems and challenges categorized and prioritized?

RQ1.2: What are some possible solutions to the difficulties and challenges discovered?

RQ2: What are the common strategies for eliciting requirements in software development?

The remaining part of the paper is organized as follows: Section 2 covers the research methodology adopted; Section 3 outlines the literature analysis that was performed; Section 4 presents the results with

regard to the research questions and discusses findings; and Section 6 closes the research article and addresses future work.

2 Research methodology

When reviewing the literature linked to the requirements elicitation, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach was applied in three key stages: search, selection, and extraction.

2.1 Search strategy

The search phrase "requirements elicitation" was used to find research publications relevant to the topic of the study. Because other search phrases, such as "specification requirements" or "identification requirements" returned an enormous amount of papers that were not as relevant to answering the research questions provided, it was decided to use only the selected search phrase.

Scopus, WoS, and IEEE databases were searched for articles between January 1, 2018, and November 15, 2022 that satisfied the following selection criteria:

- open access to data
- and a search string in the title of the article.

All articles that satisfied all of the mentioned criteria were selected and downloaded to a local database.

2.2 Selection process

The further selection process was carried out in three phases:

1) Initially, all duplicates were removed from the local database.
2) All article titles and abstracts were read, and those that did not meet one or more exclusion criteria were removed.
3) All articles were read completely, and in order to be considered further, they had to meet both defined inclusion criteria.

Exclusion criteria: does not refer to the request collection procedure, there is no systematic literature review provided, and it is not in English.

Inclusion criteria: it has a clearly stated goal of the paper and identified problems or challenges in the request elicitation process.

It was indicated as an exclusion requirement that the papers must include an obligatory systematic evaluation of the literature; therefore, our literature analysis would become even more important due to the possibility of evaluating a large quantity of previously summarized data from prior research.

2.3 Data extraction

After the papers were chosen, each one was carefully examined to determine whether it addressed some of the research questions that had been posed. Following a summary of the findings, key conclusions from each study were derived. The key observations were presented in tabular and graphic form so that the results could be more easily studied.

3 Literature review procedure

A systematic literature review was carried out using the methodology described in the previous section. Figure 1 depicts the applied research methodology to make it easier to understand.

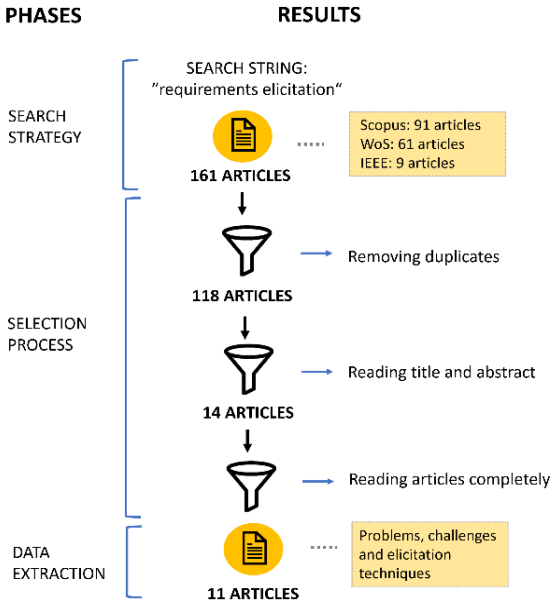


Figure 1. Article selection procedure

Using the PRISMA approach, a total of 161 articles matching the given search strategy were located: 91 in the Scopus database, 61 in the WoS database, and 9 in the IEEE database.

Because the same items appeared in many databases, duplicates as well as articles that could not be accessible despite the selected filter had to be removed. After deleting duplicates from the IEEE and WoS databases, 66 Scopus articles remained, 43 from the WoS database and all nine from the IEEE database.

In subsequent phases, articles were chosen based on the defined exclusion criteria, leaving 14 articles. After reading the articles, it was discovered that three of them did not have documented problems and challenges in the request elicitation process, and one of them did not even have a defined aim for the paper. As a result, a total of 11 articles were included in the subsequent data

extraction and article analysis procedures; the summary results and key observations will be provided in the next section.

4 Results and discussions

This section discusses the findings and responds to the stated research questions. To respond to RQ1, Anwar et al. (2022) state that requirements elicitation challenges with a frequency percentage greater than 50% are considered highly significant, and these are: poor communication, difficulty with identifying relevant stakeholders, ambiguity among stakeholders, lack of awareness of the need, unclear information, expert experience and technical knowledge, insufficient level of request details, communication problems, tacit knowledge, and the possibility of knowledge exchange. Problems with mobile application development are frequently associated with imprecise requirements, limited user interaction, poor planning, and disregarded needs (Dar et al., 2018). Other challenges involve: constant changes in requirements, neglected non-functional requirements, inappropriate architecture, minimal documentation, inaccurate cost and deadline estimates, and difficulties determining the priorities of the requirements themselves (Ferreira Martins et al., 2019).

Regarding RQ1.1, our literature review reveals four of the most prevalent challenges during request elicitation:

- **incorrect requirements** - ambiguous or incomplete requirements, changes in requirements, tacit knowledge
- **lack of communication** - lack of stakeholder involvement, failure in communication
- **choice of elicitation technique** - improper selection, limited knowledge of techniques
- **identification of key stakeholders** - incorrect identification, unknown names of all involved

Table 1 shows the statistics for the most common problems and challenges, with an "x" indicating whether the stated category of problems is mentioned in the selected articles.

The frequency of appearances in publications criterion was used to categorize and prioritize the significance of requirement elicitation problems and challenges.

According to Table 1, the majority of problems in the request gathering procedure are caused by incorrect requirements, which appeared in nine of the eleven examined articles. Then there was a lack of communication in six articles, problems with selecting the proper request elicitation technique in five, and inadequate identification of important stakeholders in five articles.

These findings, which are similar to the earlier Standish Group study (Clancy, 1995), show that the difficulty of identifying the true needs of users continues to present the biggest challenge in the modern world. Because actual client requirements are not fully explored from the beginning, it is more likely that most service development attempts will fail.

Furthermore, the communication issue appears to be the second most important challenge, while other problems are equally represented in the examined articles. This is supported by the fact that in the last two years, problems related to unclear requirements and communication issues with stakeholders have received increased research attention, which can be attributed partially to the impact of the coronavirus disease 2019 (COVID-19), as the term pandemic appeared in the titles of two articles that addressed the aforementioned

Table 1. Problem and challenge categories in request elicitation

	Incorrect requirements	Lack of communication	Choice of elicitation technique	Identification of key stakeholders
Anwar et al. (2022)	x	x	x	*x*
Gillani, Niaz, & Ullah (2022)	x	x		
Khan et al. (2022)	x			x
Alflen & Prado (2021)	x			
Amin, Shahzad, & Shoaib (2021)	x	x	x	
García-López, Segura-Morales, & Loza-Aguirre (2020)		x	x	
Lewellen (2020)				x
Aldave et al. (2019)	x		x	
Ferreira Martins et al. (2019)	x		x	
Dar et al. (2018)	x	x		x
Pacheco, García, & Reyes (2018)	x	x		x

issues. The analysis's findings will be discussed in relation to the established classification of issues and difficulties in the following section.

4.1 Incorrect requirements

Users frequently provide **incomplete** or **unclear requests**, making obtaining a detailed understanding of what consumers genuinely want one of the most difficult tasks (Aldave et al., 2019). Because there are several stakeholders, their perceptions, interests, and expectations of the final software product differ as a result of their experiences and positions in the specific domain of application (Anwar et al., 2022). Of course, due to the large number of stakeholders involved, meeting all of their requirements is complicated.

Since a large number of critical stakeholders are involved in achieving the same goal, **conflict** is inevitable. This is especially visible in the articulation of needs, which is why requirements may be unclear and ambiguous (Khan et al., 2022). Ferreira Martins et al. (2019) also emphasize that reaching consensus is difficult when the system includes more than one group of stakeholders because each is focused on different aspects of the system. Another issue that may occur is the stakeholders' scepticism that the development team truly understood their needs. As a result, Khan et al. (2022) propose limiting the number of valid requests by selecting just those key stakeholders for the software development process before the request elicitation begins.

While most interactions with consumers occur through dialogue, it is often difficult to extract true requirements from them (Aldave et al., 2019). According to Pacheco et al. (2018), there is no one-of-a-kind method or strategy for quickly identifying user requirements. As a result, each project and stakeholder must be approached on an individual basis.

One issue that arises is **tacit knowledge**. Tacit or hidden knowledge is knowledge that cannot be communicated, which can lead to unclear and inadequate demand elicitation (Anwar et al., 2022). When eliciting requirements, tacit knowledge manifests itself in the form of requests that users did not explicitly specify or simply because they failed to emphasize the essential features of the system based on the assumption that these are widely known features that do not need to be emphasized further (Amin et al., 2021). In order to resolve this issue, it is suggested that the requirements engineer "elicit" these needs in some way, allowing for the creation of a requirements specification document from the hazy concepts (Anwar et al., 2022).

Another issue with requirements, according to Amin et al. (2021), is inconsistent requirements or frequent **changes in requirements**. Even though a variety of approaches and technologies may be used to close the gap between requirements and design, list the ambiguity of requirements related to architecture is identified as one of the problems with software engineering (Gillani et al., 2022). Although an individual's knowledge is composed of 90% tacit and 10% explicit knowledge, the most typical difficulty is that stakeholders or customers have a hazy understanding of the required or desired functionality (Anwar et al., 2022).

4.2 Lack of communication

Poor communication and a lack of user interaction are other common causes of difficulties (Aldave et al., 2019; Amin et al., 2021). Ferreira et al. (2019) identify several key challenges and problems that arise in agile methodologies, emphasizing that a lack of stakeholder involvement in the process of determining software requirements is the fundamental cause of many of the problems that arise as a result. For example, establishing the actual requirements may be difficult due to a lack of communication.

Clients were unable to communicate directly with system analysts throughout the COVID-19 pandemic, and Amin et al. (2021) think that this lack of communication with stakeholders was one of the key causes of project failure.

Also, most experts agree that several communication issues occurred throughout the COVID-19 pandemic. According to Anwar et al. (2022), a lack of space for undisturbed work from home reduces employee efficiency and attention, making it impossible to share tacit knowledge. Moreover, most customers have difficulty communicating their requirements, which complicates the requirement collection process (Alflen & Prado, 2021).

Collaborative sessions are mentioned as one of the proposals for a better understanding of user requirements, in which all relevant participants, from developers to business leaders, could participate in the "designing" of requirements in agile development projects (Aldave et al., 2019). However, in order for the communication between the requirements engineer and the stakeholders to be effective, the requirements engineer's experience and expertise in the issue domain are also required (Garcia-López et al., 2020).

4.3 Choice of elicitation technique

Regarding RQ2, five articles with a systematic review of the literature were examined to discover which techniques were referenced the most commonly by the authors.

The literature analysis found fifteen popular elicitation techniques: interviews, surveys, brainstorming, prototyping, observation, scenarios, ethnography, joint application development (JAD), introspection, use cases, social analysis, workshops, focus groups, user stories, and similar systems technique. The collected results are displayed in Table 2, number N next to each author represents the total number of publications that have been evaluated using

Table 2. Identified requirement elicitation techniques

		Anwar et al. (2022) N=29	Amin et al. (2021) N=50	Pacheco et al. (2018) N=109	Alflen & Prado (2021) N=61	Sum of articles
Interviews	13	23	12	30	45	123
Survey/Questionnaire	10	21	17	7	34	89
Brainstorming	5	20	12	5	24	66
Prototyping	8	20	7	4	25	64
Observation	10	40	2	0	11	63
Scenarios	0	19	4	10	23	56
Ethnography	4	21	2	4	10	41
JAD	5	9	10	0	10	34
Introspection	0	0	31	0	2	33
Use cases	0	0	0	3	25	28
Social analysis	3	23	0	0	0	26
Workshop	0	0	5	3	17	25
Focus groups	0	0	5	3	17	24
User stories	0	0	0	8	12	20
Similar system	3	16	0	0	0	19

a systematic literature review, whereas the cells represent the number of articles that addressed particular techniques.

As can be seen in Table 2, interviews not only appear in the greatest number of articles in total, but they are also the most commonly used technique. However, before starting a detailed examination of the data for the purpose of comparison, it is essential to investigate which elicitation techniques are widely used and recognized on the market.

A Guide to the business analysis body of knowledge (BABOK Guide) as a standard for the practice of business analysis states nine of the most popular requirements elicitation techniques in software engineering (Apriorit.com, 2022):

- Brainstorming
- Document analysis
- Focus group
- User interface analysis
- Interviews
- Observation
- Prototyping
- Requirements workshops
- Survey/questionnaire.

When comparing the acquired results to BABOK's list of the most often used elicitation techniques, it can be noted that the majority of the mentioned techniques are recognized in the selected articles. The only techniques with non-overlapping results are user interface analysis and document analysis, which were not recognized as widespread techniques in a thorough review of the literature. To make it easier to interpret the previous data, a graphical representation of the results is provided. Accordingly, Figure 1 illustrates the most often used elicitation techniques in descending order of citation frequency in selected articles.

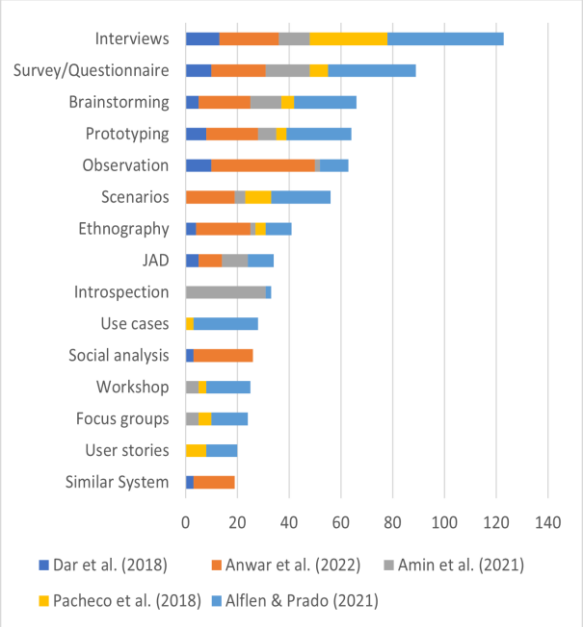


Figure 2. Frequency of elicitation technique

As shown in Figure 2, interviews are by far the most commonly used technique for requirement elicitation, followed by a survey, and finally, brainstorming, prototyping, and observation, which appear in nearly equal proportions. This alone illustrates that interviews in the form of asking questions to elicit responses from users, as well as surveys, are still widely used to identify requirements.

Only the five most popular elicitation strategies chosen will be further explained, with reference to the key results of other techniques.

The characteristics of the five most frequently used elicitation techniques are discussed further below.

- **Interviews:** an interview is a method of direct communication between the interviewer and the user (Dar et al., 2018). From Pacheco et al. (2018) point of view, interviews are the most effective technique since they allow analysts to get far more data than the other elicitation techniques.
- **Survey/questionnaire:** a survey is a traditional elicitation technique for collecting data by asking important questions with the goal of understanding user opinion (Pacheco et al., 2018). It is typically used in the first stages of requirements elicitation to gather as many requirements as possible from various stakeholders who may be located in various locations (Alflen & Prado, 2021). It was also a useful tool during the pandemic era, when face-to-face interaction was impossible (Anwar et al., 2022).
- **Brainstorming:** a brainstorming is a meeting in which each member is free to voice any thoughts they have on the system's needs (Alflen & Prado, 2021). It is often used in the case of extracting tacit knowledge (Amin et al., 2021).
- **Prototyping:** a prototype, as a simplified version of the software, is used when there is a lot of uncertainty regarding the requirements, with the goal of getting a clear idea of how the software will perform in practice and encouraging feedback from stakeholders (Pacheco et al., 2018). When clients want to be involved in the creation of the system and use it, prototyping works better (Dar et al., 2018).
- **Observation:** an analyst examines consumers in their natural environment through observation. It could be active, with analysts asking questions, or passive, with user activities being observed (Amine t al., 2021).

Although there is little research on the subject, the authors strain the significance of picking an appropriate technique for requirements elicitation in order to gather the requirements that will be used to build an appropriate software architecture (Gillani et al., 2022). For example, tactic requirements can be elicited by brainstorming, scenario planning, and prototyping (Anwar et al., 2022).

Introspection, as a technique in which system analysts use their experience to identify user requirements, according to Amin et al. (2021), should be utilized as a starting point in the requirements elicitation process, particularly for large projects. It would be possible to discover all known needs in this manner, significantly decreasing the time and cost of the software development process. Ethnography is used to discover problems; for that purpose, the observer enters the user's environment to collect detailed observations, while the scenario technique can assist in determining which functional needs should be incorporated into the system (Amin et al., 2021). Social analysis is done when it is unknown what the user's surroundings, especially the political and social ones, are like (Dar et al., 2018).

4.4 Identification of key stakeholders

According to Dar et al. (2018), there is a need to manage stakeholders' requirements due to the difficulty in identifying acceptable stakeholders and the challenges that arise as a result of their involvement in the product development process. According to Lewellen (2020), a critical stakeholder is someone whose late disclosure of requirements may disrupt the early software planning and project development timeframe. Specifically, many software development projects fail precisely because of incorrect stakeholder identification, which results in an inability to identify the true system requirements and hence a software product that does not fulfil the true wants and demands of customers (Khan et al., 2022). However, due to the vast number of stakeholders, it is frequently impossible to identify all significant stakeholders, despite the fact that their accurate identification should ensure comprehension of the most critical needs throughout the software planning phase (Lewellen, 2020).

Khan et al. (2022) identify a potential solution in the compilation of a list of various stakeholders or organizations that may have an impact on the software development project. Then, from that list, interesting stakeholders having direct or indirect contact with software development, such as analysts, journalists, and so on, should be identified. Following the identification of all stakeholders, the process of eliminating those who are not significant and relevant to the software product would begin, leaving just the vital and relevant stakeholders.

5 Conclusion

A business analyst's initial task is to obtain requirements from the client. It refers to an organized strategy aiming at "drawing out" information and forging consensus regarding application/software development needs. Elicitation is the foundation of any project because it is responsible for bringing project requirements to the table. Elicitation errors, according to scientists and engineers, are one of the most common reasons of project failures and abandonment, which have a negative influence on the bottom line. Simply expressed, the purpose of requirement elicitation is to discover all assumptions, risks, and needs associated with any project.

The literature was analysed to get insight into the most prevalent problems that arise throughout the request elicitation process, and it was discovered that the problems are connected. Problems with inaccurate identification of relevant stakeholders or a lack of engagement with stakeholders, in particular, can result in ambiguous or incomplete requirements. Similarly, if

the proper request elicitation approach is not used, it is more difficult to find the user's actual requests, which leads to frequent changes in the user's requests. By reviewing the findings, it is clear that the most prevalent issues originate with stakeholders, requirements, and elicitation approaches. However, in order to solve the difficulties, each of these three major factors must be solved.

References

Aldave, A., Vara, J. M., Granada, D., & Marcos, E. (2019). Leveraging creativity in requirements elicitation within agile software development: A systematic literature review. *Journal of Systems and Software, 157*, 110396.

Alflen, N. C., & Prado, E. P. V. (2021). Requirements elicitation techniques for software development: a systematic review of literature. *AtoZ: novas práticas em informação e conhecimento; v. 10, n. 1 (2021): jan./abr.*; 39-49, 24(2), 49-39.

Amin, T. U., Shahzad, B., & Shoaib, M. (2021). Economical Requirements Elicitation Techniques during COVID-19: A Systematic Literature Review. *Computers, Materials and Continua*, *67*(2), 2665-2680.

Anwar, H., Khan, S. U. R., Iqbal, J., & Akhunzada, A. (2022). A Tacit-Knowledge-Based Requirements Elicitation Model Supporting COVID-19 Context. *IEEE Access, 10*, 24481-24508.

Apriorit.com (2022). *Requirements Elicitation: Reasons, Stages, Techniques, and Challenges*. [Blog post]. Retrieved May 10, 2023, from https://www.apriorit.com/white-papers/699-requirement-elicitation

Clancy, T. (1995). The standish group report. *Chaos report.*

Conrad, A. (2020). *How to Build Effective Customer Feedback Surveys.* [Blog post]. Retrieved May 10, 2023, from https://www.gartner.com/en/digital-markets/insights/how-to-build-effective-customer-feedback-surveys

Dar, H., Lali, M. I., Ashraf, H., Ramzan, M., Amjad, T., & Shahzad, B. (2018). A systematic study on software requirements elicitation techniques and its challenges in mobile application development. *IEEE Access, 6*, 63859-63867.

Ferreira Martins, H., Carvalho de Oliveira Junior, A., Dias Canedo, E., Dias Kosloski, R. A., Ávila Paldês, R., & Costa Oliveira, E. (2019). Design thinking: Challenges for software requirements elicitation. *Information, 10*(12), 371.

García-López, D., Segura-Morales, M., & Loza-Aguirre, E. (2020). Improving the quality and quantity of functional and non-functional requirements obtained during requirements elicitation stage for the development of e-commerce mobile applications: an alternative reference process model. *IET Software, 14*(2), 148-158.

Gillani, M., Niaz, H. A., & Ullah, A. (2022). Integration of Software Architecture in Requirements Elicitation for Rapid Software Development. *IEEE Access, 10*, 56158-56178.

Khalimonchuk, K. (2022). *Software Development: How it Saves Time & Money*. [Blog post]. Retrieved May 10, 2023, from https://fulcrum.rocks/blog/discovery-phase-software-development#what-is-a-discovery-phase?-1

Khan, F. M., Khan, J. A., Assam, M., Almasoud, A. S., Abdelmaboud, A., & Hamza, M. A. M. (2022). A comparative systematic analysis of stakeholder's identification methods in requirements elicitation. *IEEE Access, 10*, 30982-31011.

Lewellen, S. (2020). Identifying key stakeholders as part of requirements elicitation in software ecosystems. In *Proceedings of the 24th ACM International Systems and Software Product Line Conference-Volume B* (pp. 88-95).

Pacheco, C., García, I., & Reyes, M. (2018). Requirements elicitation techniques: a systematic literature review based on the maturity of the techniques. *IET Software*, 12(4), 365-378.

The impact of using ChatGPT on employees of Croatian IT Companies

Donata Szombathelyi, Tomislav Jakopec, Josipa Selthofer
Josip Juraj Strossmayer University of Osijek
Faculty of Humanities and Social Sciences in Osijek
Lorenza Jagera 9, 31000 Osijek, Hrvatska
{dszombathelyi, tjakopec, jselthofer}@ffos.hr

Abstract. *More efficient technologies and adaptations in work mainly create challenges for employees in the IT sector, who, in addition to search engines to solve problems, have recently turned more and more to chatbots. Chatbots are software systems whose goal is to automate and simplify certain tasks, and considering their widespread use today, a study was conducted that tries to answer the question of what impact ChatGPT creates on the work of employees in Croatian IT companies; in which IT area is its use the most dominant, what are its advantages and disadvantages, do search engines or ChatGPT dominate when it comes to solving problems, and what is the level of security and trust, and thus the perception of ChatGPT's development in the future. Using a survey questionnaire, it was confirmed that the full-stack development has used ChatGPT the most so far, its wide application was singled out as an advantage, while limited knowledge was also confirmed as a disadvantage. Conventional search engines still dominate in solving business challenges, and in a comprehensive review, the existing results, (and the IT population so far in the world), are very cautious about their use and have good attitudes about security and trust. The latest research show that the constant progress and improvement of ChatGPT is expected because they have encouraged their growing integration in different IT sectors, which will result in even more significant improvements in terms of access to problem solving for employees.*

Keywords. GPT, ChatGPT, IT, business

1 Introduction

We live in very interesting times from the point of view of the use of technology. Company's websites, expert sources, and articles in the media point to the battle in which companies are fighting for supremacy in the Large language model market as the next step to artificial intelligence. In 2022, Meta, the umbrella company of Facebook, announced a device and platform for using virtual realities known as Metaverse. Like any business venture, they presented it as the next big thing in our lives. However, the effect of everyday business and personal use of GUI (windows 95, 1995) or smartphones (iPhone, 2007), which have become integral parts of people's lives, did not occur. Not long after that, the Large language models developed by the company OpenAI, called GPT-3, became available to end users via the OpenAI Chat-GPT interface. Soon the topic of virtual reality fell into the background and the world of technology began to talk about the possibilities of the Generative Pre-trained Transformers (GPT). Compared to Meta's efforts to introduce virtual reality, besides rumours of Apple tackling that field, none of the big players (Google, Microsoft, Oracle and others) presented their competitors. In the case of Large language models, shortly after the presentation of ChatGPT, Google Bard was also presented as an immediate response in the usage of this technology in the search results. In contrast, there are reports that others will get into that race as well. It is pretentious to say after such a short time that the next big thing has just occurred, but there is a high probability that we are currently living at the next big thing point in time that will change our lives. Therefore, this paper aims to examine the employees of Croatian IT companies regarding the use of OpenAI ChatGPT technology as the first representative of Large language models technology widely available. Accordingly, the paper first looks at existing research on the use of ChatGPT in the world, followed by the described research (conducted in Croatian IT companies), goals, overall methodology, results, discussion and finally the conclusion, together with references.

2 Previous research

Previous research in the field of Large language models, GPT and artificial intelligence chatbot, particularly ChatGPT is rare and new since an OpenAI was released first in November 2022. Since that time, it became a global issue. It has been argued whether chatbots can write student essays, be an author of

scientific articles etc. (Lee, 2023). Looking at the state of the Art in artificial intelligence chatbot literature, it is obvious the lack of an interdisciplinary approach. In a bibliometric analysis from 2018 based on the information contained in the Scopus database, 314 documents were analysed. The first publication appeared in 2002, and in 2016 there was a significant increase in the number of publications. Probably due to fact of progress in performance, hardware, market visibility and potential investments in the artificial intelligence area (Bernardini at all, 2018).

Artificial intelligence is aimed to skip human interaction into problem solutions or task execution. According to the author Jatin Borana, artificial intelligence can by analyzing data be useful in the gaming industry (chess), heavy industries (in jobs dangerous for humans), weather forecasting (by analysing patterns they predict the weather), expert systems (statistical analysis), data mining or knowledge extraction and knowledge representation (based on data mining they establish patterns, associations and classification; all needed in the process of data visualisation) (Borana 2016).

In the business world, there are some concerns regarding the wide popularity of ChatGPT. For example, if a user request is replaced by a chat engine, certain websites could lose their purpose (De Vynck, 2023). There are reports of companies banning the use of ChatGPT for business purposes (Vincent, 2023; Tilley and Kruppa, 2023).

In a paper published in 2018, authors Anghelescu and Nicolaescu (2018), focused on the implementation and design of the Chatbot system using search engines and teaching techniques.

Further, authors in 2023 researched how ChatGPT may serve in management theories and concepts. Their findings are that some management theories and concepts could be used in artificial intelligence and implemented at strategic, functional, or administrative levels of business (Korzynski, Mazurek, Altmann et all, 2023).

Arguing on ChatGPT's legibility, there is also a question of its maliciousness. As an open tool, it is vulnerable to attacks from various actors (Murphy, Du & Suarez, 2023).

In the world of business, artificial intelligence platforms could be a new opportunity, for employers, as well as employees. As ChatGPT is evolving question is how it will affect people's businesses. It is a common belief that only human is capable to create an idea and possess creativity. Group collaboration and the exchange of knowledge, emotional intelligence and joy that brings working in a human environment can't be replaced. So, embracing new technology could be an opportunity for improving jobs and thinking in different ways. For sure, some of the jobs will be replaced by artificial intelligence, but also other job opportunities will appear (Garg and Khan, 2023).

Easy and free to use, ChatGPT is becoming frequently used by small and medium-sized companies, as well as start-ups. It can be used to identify product features, develop marketing strategies, or create pricing models. Although of the potential, many companies are cautious, mainly in terms of legal issues. EU plans to regulate artificial intelligence (European Commission, 2023). The plan is to divide applications into different categories so that they could be marked as risks or as potential for the society and economy (Büchel & Mertens, 2023).

3 Research

3.1 Aim and research questions

This research aimed to examine the employees of Croatian IT companies about the use of ChatGPT and to provide answers to the following research questions:

1. In which areas of business have Croatian IT companies used ChatGPT the most so far?
2. What are the advantages and disadvantages of ChatGPT for employees?
3. When solving business challenges, which dominates more, conventional search engines or ChatGPT?
4. Do employees have sufficient security and trust in ChatGPT?
5. What attitudes do employees have regarding the development of ChatGPT in the future?

3.2 Methodology

This research was conducted from March 31st to April 14th, 2023, and it included all employees of IT companies in the Republic of Croatia (including students and interns in IT companies). For this research, an online questionnaire was created in Lime Survey[1] and sent to employees, students, and interns via the official website of the Faculty of Humanities and Social Sciences Osijek (FFOS), via Gmail, LinkedIn, and WhatsApp. The questionnaire was composed of 18 questions[2], of which were single and multiple-answer types. In the first group of questions, the respondents were asked to confirm their gender, age and to indicate the job position in which they are employed. Moreover, the second group of questions required respondents to state which search engines they use and how useful the results are. The third group of questions refers to the use, purpose and usefulness of the results obtained by ChatGPT. Furthermore, the fourth group refers to the ease of use of ChatGPT and its advantages and limitations. Also, the fifth group of questions refers to trust, security, and negative experience in using ChatGPT. Lastly, the sixth group of questions asked respondents to state which they use more - ChatGPT or search engines, to indicate their

[1] https://puh.srce.hr/s/ofeiQPEbcpombgA

[2] https://puh.srce.hr/s/NqxGtCPtiKzb2ey

perception regarding the development of ChatGPT in the future, and to give their opinion on ChatGPT as a replacement for conventional search engines. It should be noted that respondents who have never used ChatGPT and those who intend to use it could only answer up to the third group of questions and 2 questions in the sixth group (perception of the development of ChatGPT and ChatGPT as a substitute for search engines).

3.3 Results

Out of a total of 166 solved questionnaires, 140 were complete answers [3], while the remaining 26 were incomplete. Regarding gender, a total of 62% (87) of respondents were male and 37% (51) were female while 1% (2) did not reveal their gender as shown in Fig 1.

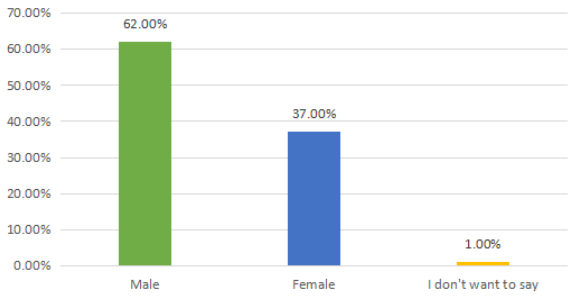


Figure 1. Gender of respondents

Fig. 2. shows the grouping age of the very diverse respondents. In the first group aged 18-20, there were a total of 4% (6 respondents), while in the second group, there were a total of 47% (66) aged 20-30. Furthermore, the third group included respondents aged 30 to 40 and there was a total of 35% (49 of them). Also, there were 10% (14 respondents) between the ages of 40 and 50 and a total of 4% (5 respondents) between the ages of 50 and 60. There were no respondents older than 60 years.

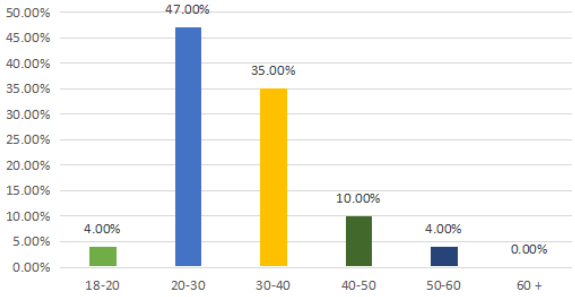


Figure 2. Age of respondents

Also, Fig. 3. shows which business areas the respondents belong to and gives the answer that full stack development is the area that uses ChatGPT the most. The total number of positions is 140: CTO 2% (3), CEO 5% (7), Project Manager 8% (11), Back-End Developer 8% (11), Front-End Developer 9.3% (13), Full Stack Developer 15% (21), iOS Developer 5% (7), DevOps Engineer 3% (4), UX/UI designer 4.3% (6), Product Owner 2% (3), Scrum Master 0,7% (1), Social Media Manager 3% (4), Content Writer 1.4% (2), Marketing Manager 5% (7), Quality Assurance 6.5% (9), Cloud System Administrator (0%), Cloud Security Specialist (0%), Database Administrator 1.4% (2), Office Manager 3% (4), Human Resources 1.4% (4)), Business Analyst 2.14% (3), Data Analyst 3.6% (5), Data Scientist 2.14% (3) and Other 7% (10).

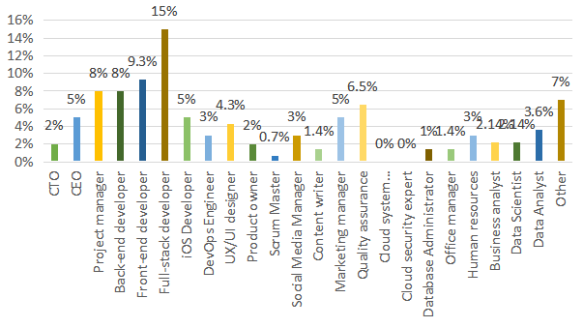


Figure 3. Business areas that use ChatGPT

Furthermore, in Fig. 4. is visible that respondents use various conventional search engines such as Google Search 87.14% (122), Yahoo 2.14% (3), Bing 1.42% (2), DuckDuckGo 7.14% (10), Ask 1.42 % (2) and Baidu 0.70% (1).

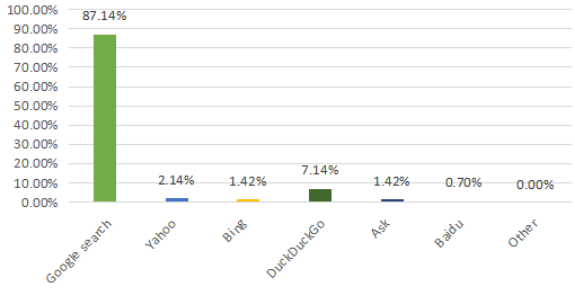


Figure 4. Use of search engines in work - Google Search, Yahoo, Bing, DuckDuckGo, Ask, Baidu and other

Fig. 5. points that a total of 46% (64 respondents) stated that the results were very useful, 30% (42 respondents) said that they were moderately useful and 24% (33 respondents) that they were completely useful. Only 1 respondent (0.70%) confirmed that the results were hardly useful.

[3] https://puh.srce.hr/s/sX6PEFXqX5kAkDK

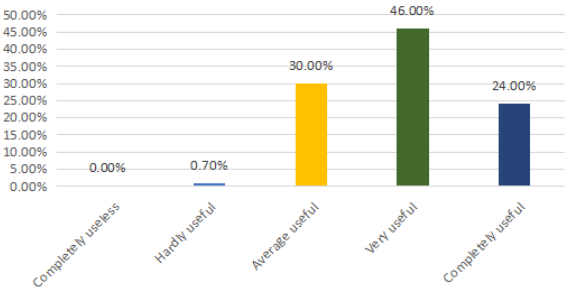


Figure 5. The usefulness of obtained data essential for work

Also, Fig. 6 indicates that a total of 4.30% (6 respondents) stated that they did not use a chatbot at all and did not plan to use one. 17.10% (24) confirmed that they did not use it, but plan to use it, while a total of 14.30% (20 users) probably used the chatbot but did not use it after that. Also, 48 respondents (34.30%) use a chatbot only, when necessary, i.e. when a certain problem needs to be solved, and a total of 30% (42 respondents) use it often because they like to experiment with artificial intelligence.

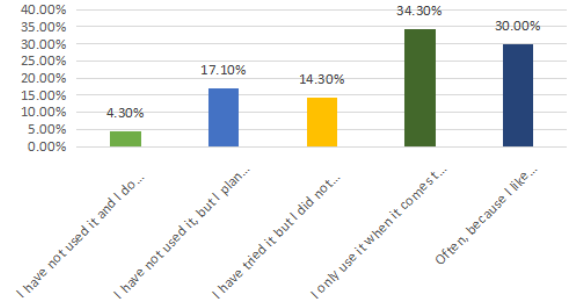


Figure 6. Using ChatGPT in work so far

Fig. 7. shows that a total of 38.70% (53 respondents) have used ChatGPT for debugging purposes, while 33.70% (46 respondents) have used it for code interpretation. For writing an essay, a total of 19.30% (27 respondents) are using it, and for writing an e-mail 24.30% (34 respondents). Also, 12.10% (17 respondents) used ChatGPT for text translation, and 41.40% (58 respondents) for a kind of testing. "Other" purposes were marked by 12.10% of respondents. This question allowed multiple answers.

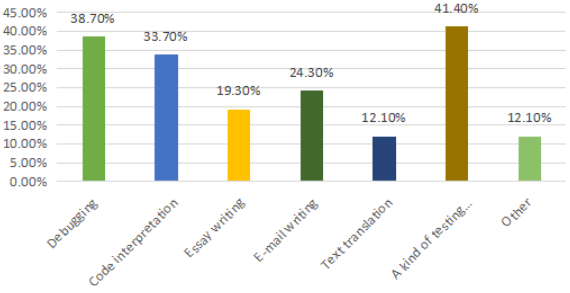


Figure 7. Purposes of using ChatGPT in work

Moreover, Fig. 8. is showing how a total of 3% (3 respondents) believe that the results are completely useless, 3% (3 of them) stated that they are barely useful, while a total of 22% (25 respondents) confirmed that they are moderately useful. As many as 32% (35 respondents) think they are very useful and 40% (44 respondents think they are completely useful.

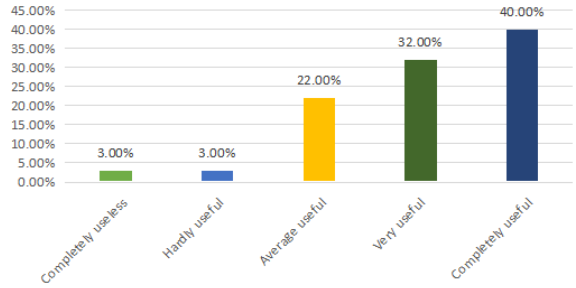


Figure 8. The usefulness of obtained data generated by the ChatGPT

Fig. 9. represents how 1% (1 respondent) thinks that ChatGPT is complicated to use, while a total of 12% (13 respondents) stated that it is moderate. 33% (33 respondents) think that it is easy to use, and a total of 54% (59) that it is very easy.

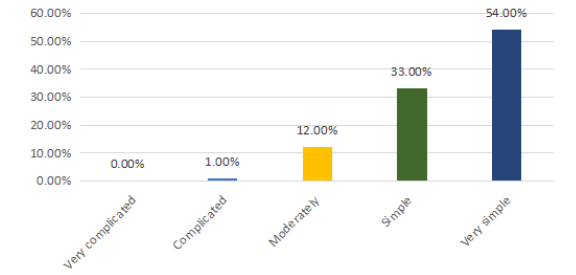


Figure 9. The ease of use of ChatGPT in work

Fig. 10. shows how wide application is the biggest advantage of 42% (46 respondents). 17% (19 respondents) think that it is the possibility of subsequent corrections, and 14% (16 of them) that it is previous interactions with the user. Also, a total of 13% (14 respondents) stated that for them the biggest advantage is the imitation of human conversation, while for the remaining 12% (13 respondents), it is the possibility to expand the features. Option "Other" have marked 2% of respondents (2).

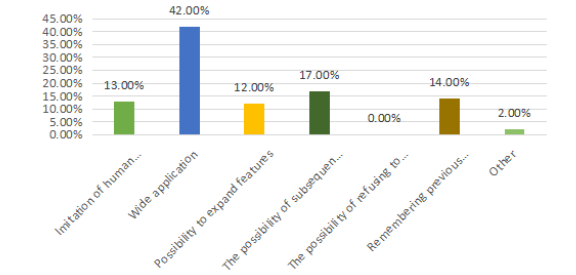


Figure 10. The biggest advantage of using ChatGPT

On the other hand, Fig. 11. represents that limited knowledge is considered the biggest disadvantage by 41% (45 respondents). Also, a total of 32% (35 respondents) believe that inaccuracy and ambiguity in giving answers are the biggest drawbacks, while for 11% (12 respondents), ethics is problematic. Legal issues are main disadvantage for 14% (15 respondents) and 2% (3 respondents) have marked "Other" disadvantages.

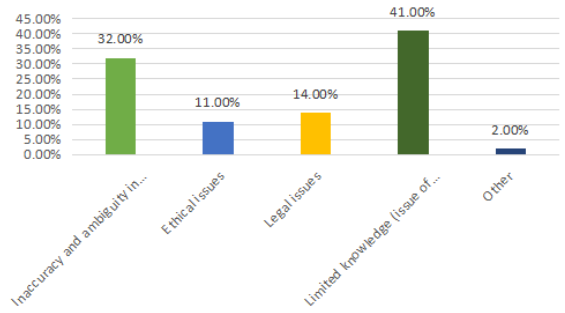


Figure 11. The biggest disadvantage of using ChatGPT

Furthermore, Fig. 12. point that 37.20% (41 respondents) think that ChatGPT is very safe and 31% (34 of them) think that it is only safe. Also, 24.50% (27 respondents) said that it is relatively safe, 6.40% (7 of them) said that it is not safe, and only 0.90% (1 respondent) thinks that it is very unsafe.

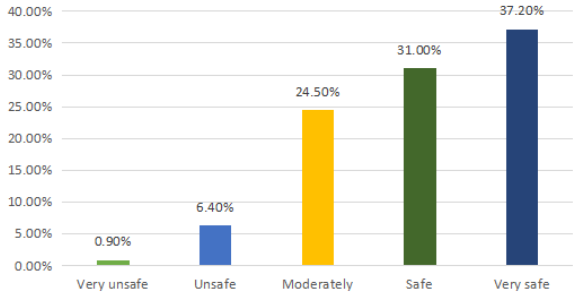


Figure 12. Safety of using ChatGPT

Also, Fig. 13. impersonates that a total of 92 respondents (84%) think they have enough trust, while 16% respondents (18) think they have no trust.

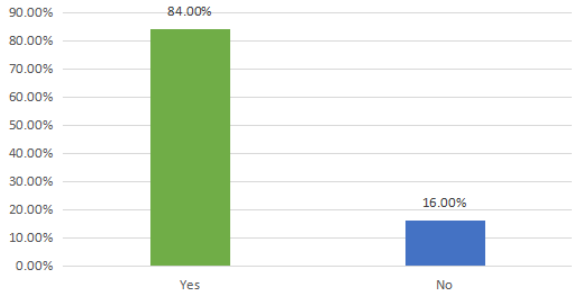


Figure 13. Trust in using ChatGPT

Likewise, Fig. 14. indicate that 73 (66.40%) respondents have never had a negative experience, 29 respondents (26.40%) had it only once, and 8 respondents (7.20%) had it more than once. No one has had a negative experience every time so far.

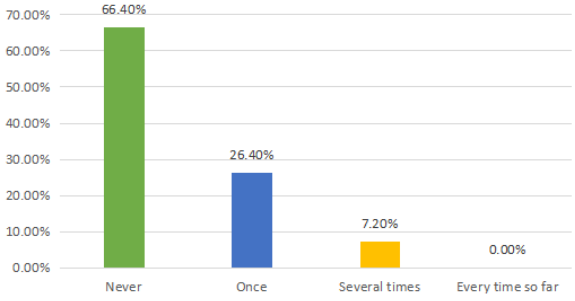


Figure 14. A negative experience by using ChatGPT

Also, Fig. 15. represents that 74 respondents (67%) use their search engine first when solving problems and 36 of them (33%) said the opposite.

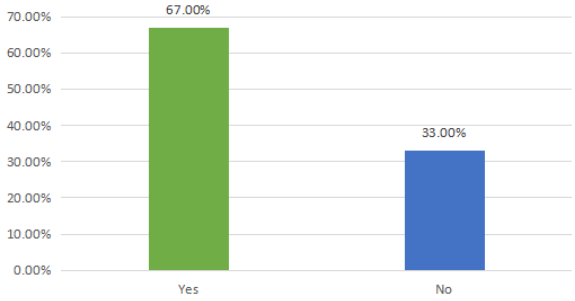


Figure 15. The first tool as a problem solver at work

When asked what they use more, ChatGPT or search engines, points to Fig. 16. 19% (26 respondents) answered that they use ChatGPT more, while 45% (63 of them) confirmed the opposite. As many as 15% (21 respondents) stated that they use both tools equally.

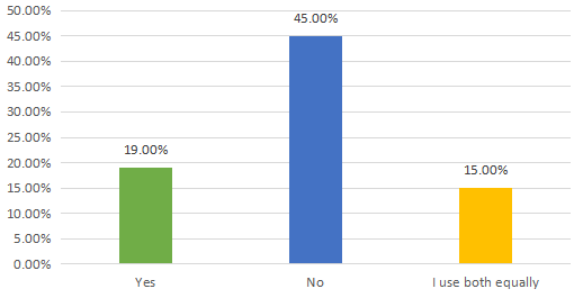


Figure 16. Domination of using - ChatGPT vs. search engines in work

The Fig. 17. shows that a total of 69% (96 respondents) have a positive perception regarding the development of ChatGPT in the future, and 2% (4

respondents) have a negative perception. 29% (40 respondents) have a neutral perception.

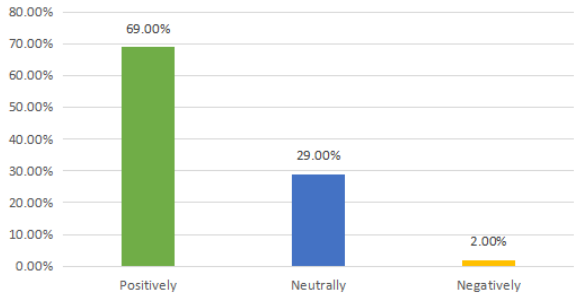


Figure 17. Perception of ChatGPT in future

In the last question, which is open-ended type, the survey confirmed that respondents believe that ChatGPT, despite its popularity, will not replace existing search engines (*"He won't, because he doesn't know how to think critically."*). Also, some stated that their relationship will be united and that they will function together (*"They will improve, and I think they will work together."*).

3.4 Discussion

Compared to the existing literature, the results give a good insight into the use of ChatGPT by employees of the IT sector in the Republic of Croatia. Considering that only a few months have passed since the start of using this chatbot, the results about the effects of its use are really good. A high-quality answer to each research question was obtained.

Based on the findings of the study, it was determined that full stack development with 15% is imposed as an answer to the first research question which is the predominant business area of using chatbots so far. The second research question aimed to explore the advantages and disadvantages of using ChatGPT, revealing that its key advantage is its wide application, which was acknowledged by 42% of respondents. On the other hand, limited knowledge of ChatGPT was identified as a disadvantage, as indicated by 41% of respondents. Regarding usage as part of the third research question, the respondent confirms that search engines still dominate as the preferred tool for finding information over ChatGPT, according to 67% of respondents. In particular, several respondents emphasized that ChatGPT is not expected to replace existing search engines, but rather is seen as a positive trend in technological progress. An analysis of the sense of security and confidence in using ChatGPT, as addressed in the fourth research question, highlights that 31% of respondents stated that they feel safe when using the system. Finally, the attitude of the employees on the future development of ChatGPT was investigated. A total of 69% of respondents had a positive perception, while 29% of respondents maintained a neutral attitude, and only 2% of respondents expressed a negative perception.

Regarding the limitations of this research, it should be pointed out that the survey questionnaire as a research instrument can be organized differently. Observing the results, moreover, it may contain some more questions. e.g. *"Do you use ChatGPT premium version", "Do you intend to pay for free ChatGPT services in the future?", "Do you use any other artificial intelligence tools in your work?"* etc. Moreover, the research instrument (if the sample is smaller) can be also an interview in order to find out more precisely the satisfaction of the users. Conducting interviews with a diverse group of IT employees would provide more detailed impressions of ChatGPT's performance and limitations. Also, the research limited the time. For detailed research, time for conducting should be longer. Likewise, sample should also be minimized. This sample is very large, namely IT employees throughout the Republic of Croatia. In this way, all areas are included at once, and not just one (e.g. not only Data Science).

As for the implications of this research, on the one hand, the theoretical contribution stands out, in the context of understanding the connection between man and machine and the increasing reliance on technology of this kind. In addition, by researching the impact of this chatbot, the so-called framework (model) for predicting trends, this is also confirmed by the fact that 30% of IT employees in the Republic of Croatia have used it because they like to experiment. Also, the progress of such a tool, in general, is an important item. By studying the requirements of employees in the Republic of Croatia and the purposes of their use of ChatGPT, employees will soon be able to get better results, thanks to the improved performance of the chatbot. But on the other hand, research implies numerous ethical and social issues. In other words, many are becoming concerned about their privacy, especially when it comes to key personal information. Also, it is implied that it is important to establish a proper implementation of the policy of using and implementing chatbots in Croatian IT companies. For example, Italy banned the use of this tool in business activities, where in fact the holistic approach to the integration of artificial intelligence in the workplace was lost, and thus the well-being and development of the skills of their employees. In addition, this coincides with De Vynck's fact that certain websites have lost their purpose due to the excessive integration of artificial intelligence tools. The goal is to create a positive and empowering environment where artificial intelligence tools complement employee capabilities, not replace or diminish them in any way.

Therefore, this research provides a general picture of the impact of chatbots on the work of employees in Croatian IT companies. For future research in Croatian IT companies, if a deeper approach and analysis of chatbot impact is required, it is necessary to specify the sample to only one area in the work (for example,

Quality Assurance or UX/UI design). In this way, all the nuances will be revealed because it is easier to focus on only one area and monitor the impact of the chatbot. In addition, it is easier to discover possible challenges that are specific to a certain position at work, and thus the necessary level of support for the employee can be better provided. Ultimately, classification into one area will enrich the understanding of the functioning of today's work not only for experts, but also for employees, and other research will probably build on the general findings of broader studies.

4 Conclusion

Artificial intelligence can be useful in the gaming industry, heavy industries, weather forecasting, expert systems, data mining or knowledge extraction and knowledge representation. In the business world, there are some concerns regarding the wide popularity of ChatGPT.

Considering the existing literature and already conducted similar research in the field, this research explored the field of use, attitudes, and perception of ChatGPT from an employee's perspective. Full stack development together with frontend and backend development is the most common area of business that uses ChatGPT. This fact was expected because developers are the ones who are supposed to come up with a solution to problems as quickly as possible. From the point of advantages and disadvantages, people are still exploring technology. It is reasonable that the biggest advantage of ChatGPT is its wide application.

ChatGPT can indeed incorporate into every segment of today's business. In addition, the problem of limited knowledge of this chatbot is a big minus for respondents, which would mean that many respondents probably turn to using their usual search engines. Security and trust in ChatGPT are quite high, and people are putting their trust in technology. The question is how quickly ChatGPT can complete tasks with the data provided in a limited time. At this moment, It is yet to see if ChatGPT will remain extremely positive and replace existing search engines. GPT technology is new, compelling, and doing things differently, and people like new and different.

As results show, developers are the dominant users so far. That is understandable because they are the most competent to understand and take the most from this new technology. The lack of an interdisciplinary approach is obvious when talking about the entire population and as seen in the previous research in the field regarding the use of chatbots and artificial intelligence in general. The easily accessible and widely accepted ChatGPT doesn't mean that humans are prepared in any way for such a transition to humanoid artificial intelligence interaction in such a short time. Many technologies over the centuries have changed the way people think, work, and interact with each other. So, understanding new ways in which humans can use new technology is crucial for knowing their best purpose. To achieve that, many new challenges are placed in front of employers, as well as employees.

References

Anghelescu, P., Nicolaescu, S. V. (2018). Chatbot Application using Search Engines and Teaching Methods. *Electronics, Computers and Artificial Intelligence, ECAI 2018 - International Conference – 10th Edition*, Iasi, ROMÂNIA, 28 June -30 June.

Borana, J. (2016). Applications of Artificial Intelligence & Associated Technologies. In (Eds.) *Proceeding of International Conference on Emerging Technologies in Engineering*, Biomedical, Management and Science [ETEBMS-2016], 5-6 March.

Bernardini, A.A., Sônego, A.A. & Pozzebon, E. (2018). Chatbots: An Analysis of the State of Art of Literature. In (Eds.) *Proceedings of the 1st Workshop on Advanced Virtual Environments and Education (WAVE2 2018)*, Florianópolis, Brazil, 4–5 October.

Büchel, J. & Mertens, A. (2023). *How companies use Chat GPT*. IW-Kurzbericht 32/2023.

European Commission. (2023). *Regulatory framework proposal on artificial intelligence*. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (2023-06-05)

Garg, A., Khan, S. (2023). ChatGPT and the future of work. (2023). *The Eightfold Talent Intelligence Platform*. Retrieved from https://eightfold.ai/blog/chat-gpt-future-work/ (2023-06-05)

Korzynski, P., Mazurek G., Altmann, A. et all. (2023). Generative artificial intelligence as a new context for management theories: analysis of ChatGPT. *Central European Management Journal*, Emerald Publishing Limited. DOI 10.1108/CEMJ-02-2023-0091. Retrieved from https://www.emerald.com/insight/2658-0845.htm

Lee, J. Y. (2023). Can an artificial intelligence chatbot be the author of a scholarly article? *Journal of Educational Evaluation for Health Professions*, 20 (6). Retrieved from https://doi.org/10.3352/jeehp.2023.20.6

Tilley, A, Kruppa, M. (2023). Apple Restricts Employee Use of ChatGPT, Joining Other Companies Wary of Leaks. The wall street journal. Retrieved from

https://www.wsj.com/articles/apple-restricts-use-of-chatgpt-joining-other-companies-wary-of-leaks-d44d7d34 (2023-06-05)

Murphy, A., Du, K. & Suarez, C. (2023). ChatGPT and Privacy Risks: Generative artificial intelligence technology chatbots rely heavily on the accuracy of vast quantities of online data, Chicago Medicine. Retrieved from www.cmsdocs.org.

Vincent, J. (2023). Apple restricts employees from using ChatGPT over the fear of data leaks. The Verge. Retrieved from https://www.theverge.com/2023/5/19/23729619/apple-bans-chatgpt-openai-fears-data-leak (2023-06-05)

Vynck, De G. (2023). Google is changing the way we search with AI. It could upend the web. *The Washington Post*, May 10. Retrieved from https://www.washingtonpost.com/technology/2023/05/10/google-io-ai-search-keynote/ (2023-06-05)

Semantic Interoperability of Digital Twins in Smart Cities

Ahmed Shareef, Boris Tomaš, Neven Vrček
University of Zagreb
Faculty of Organization and Informatics
Pavlinska ul. 2, 42000, Varaždin
{ashareef, btomas, nvrcek}@foi.hr

Abstract. *Semantic Interoperability of digital twins in smart cities is crucial for improving the efficacy and compatibility of digital twins' communication, which will result in improved smart cities and lifestyles. Without constant interoperability improvement, it will be impossible to effectively communicate and share information between digital twins. This paper examines the possibility of interoperability enhancement focusing Microsoft Azure Digital Twin Definition Language (DTDL) and discusses future research ideas that could contribute to the enhancement of Azure DTDL.*

Keywords. digital twin, interoperability, smart city, azure dtdl

1 Introduction

Semantic interoperability refers to the ability of different devices to communicate and exchange information with each other in a meaningful way. In other words, these systems can understand and interpret the data they receive from each other, even if they use different languages or formats to represent those data. Achieving semantic interoperability is essential to ensure that different systems can work together seamlessly and efficiently, which is particularly important in fields such as healthcare, finance and government, where accurate and timely data exchange is critical (Nilsson & Sandin, 2018).

Digital Twins of Smart Cities (DTSC) are virtual replicas of physical cities that use data to simulate and analyse the behaviour of the city in real time. These digital models can be used to improve city operations, improve urban planning, and improve the overall quality of life of citizens. The digital twins of smart cities rely on semantic interoperability to gather and analyze data from various sources, such as sensors, cameras, and other IoT devices, to provide a comprehensive view of the city's activities (Deren et al., 2021).

Semantic interoperability is a crucial component in the development and implementation of DTSC, as it enables the seamless exchange and analysis of data from multiple sources, ultimately leading to more efficient and sustainable urban environments.

By achieving semantic interoperability, digital twins of smart cities can provide accurate and timely insights into the functioning of a city, allowing policymakers and urban planners to make informed decisions and implement effective strategies for sustainable development (Raghavan et al., 2020)(Yang et al., 2021).

In addition to improving decision-making, semantic interoperability also facilitates enhanced collaboration between different stakeholders involved in the development and management of smart cities. With a shared understanding of the data and its meaning, different departments and agencies can work together more effectively towards common goals, leading to greater efficiency and effectiveness in city operations. Furthermore, semantic interoperability also promotes transparency and accountability, as all stakeholders have access to the same data and can work together to ensure that the city is being managed in the best possible way (Raghavan et al., 2020).

Semantic interoperability can enable the automation of processes and workflows, reducing the time and effort required to perform tasks (Raghavan et al., 2020). By enabling different systems to work together, semantic interoperability can help foster innovation and the development of new solutions and services (Yang et al., 2021).

The existing problems regarding semantic interoperability in the context of smart cities and digital twins include technical issues related to data collection, retrieval, exchange, analysis, and processing, data sovereignty issues, and the lack of a unified approach to data exchange that takes into account the semantic relationships between different systems and their data. These challenges make it difficult to achieve seamless data exchange and integration between different systems and stakeholders in smart cities and digital twins.

According to Lehtola et al (Lehtola et al., 2022) in the context of smart cities and digital twins, there are some existing problems regarding semantic interoperability. These problems include the need for semantic recognition of data pieces that represent objects in the city, such as buildings. It also shows

challenges of ensuring that different digital twins used in the city are interoperable and can communicate with each other effectively. Therefore the need for standardization of data formats and communication protocols to ensure interoperability between different digital twins and smart city systems.

This paper focuses on achieving semantic interoperability of digital twins in relation to DTDL implementation while relying on an ontology-based approach supported by MS Azure DTDL. The rest of the paper is organized as follows: Section 2 describes the current state of digital twins, smart cities, and the challenges of achieving semantic interoperability. Section 3 represents the reference model presented in the paper for building an interoperable system of systems supporting the implementation of smart cities and digital twins. Section 4 uses the reference model as a guide to demonstrate a framework for achieving semantic interoperability in digital twins of smart cities. Section 5 proposes future research ideas to improve the interoperability. To conclude, Section 6 summarizes the key points of the article.

2 Background

2.1 Digital twins and smart cities

A digital twin is a virtual copy of a physical system, which can be designed, tested, manufactured, and applied in a virtual environment. It is a set of virtual information constructs that are designed to fully describe a potential or existing physical manufactured product. In simpler terms, it is a digital representation of a physical object that contains all the attributes and behaviors of the real-life object through modeling and data.

A smart city is a city that uses advanced technologies, such as the Internet of Things (IoT), artificial intelligence (AI), and data analytics, to improve the quality of life for its citizens, enhance sustainability, and optimize urban services. In the context of the paper, a smart city is conceptualized as a digital twin that includes the infrastructure, human dynamics, spatial and temporal information flow, and physical and virtual connectivity of a city.

According to Du et al (Du et al., 2020), using a digital twin model to collect cognitive data and analyze cognitive patterns under different information cues to create a personalized information system that reduces the risk of cognitive overload for people in a smart city. The personalized information system will tailor information based on the cognitive digital twin model of each person. This model will tell the system what information format to use, when to use it, and how to use it. This will make information delivery in a smart city more efficient and effective, making life better for the people who live there and making the most of urban services. So, the digital twin makes a smart city better by making it possible to create a personalized information system that reduces mental overload and improves the way information is delivered.

The definition of "semantic interoperability" is the use of the same language or framework between different systems and devices so that they can communicate with each other and share data. It is important to use the same frameworks, ontologies, and standards to make sure everything works together. But there are still problems that need to be fixed in order to make semantic IoT work better in smart cities (Nahhas, 2023).

2.2 Examples of existing architectures and standards for data exchange in smart cities and digital twins

The concept of "Common European data spaces", as in Figure 1, which refers to a rich pool of data that varies in accessibility and can flow freely across sectors and countries while fully respecting the General Data Protection Regulation (GDPR). It has several aspects related to the management of data that need to be considered in order to make smart cities and digital twins a reality. These aspects include technical issues related to collecting, retrieving, exchanging, analyzing, and processing data, data sovereignty issues, and data semantics, features, and metadata specifications. *For example, the European Union data strategy describes the creation of "data spaces", and explicitly conceives these as multiple interoperating data spaces for different domains* (Atkinson et al., 2022).

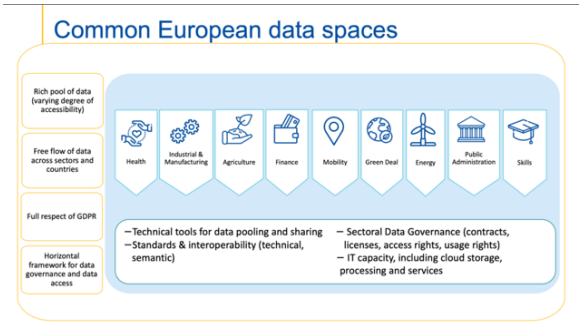


Figure 1. Common European data spaces (Dataspaces, 2023)

Common European data spaces refer to a concept of creating a shared platform for data exchange and collaboration among different sectors and countries in Europe. The Figure 1 shows a visual representation of the Common European data spaces, highlighting the different components involved in maintaining the flow of information among them.

DPA stands for Data Privacy-preserving Automation architecture. It is a backend computing architecture proposed in the research paper (Xiao et al., 2018) to facilitate online privacy-protection processing automation and secure data privacy. The DPA architecture can be seamlessly integrated with

companies' principal application system in an interruption-free manner, allowing for adaption to flexible models and quality of service (QoS) guarantee for cross-entity data exchange.

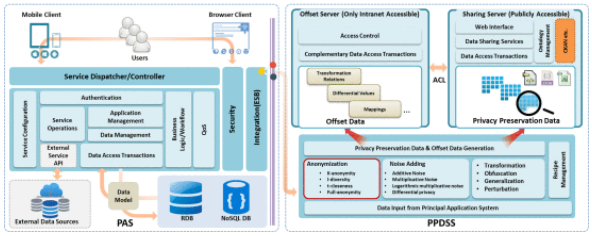


Figure 2. Examples of existing architectures and standards for data exchange in smart cities and digital twins (Xiao et al., 2018)

Data Privacy-Preserving Automation Architecture (DPA) is designed to automate the process of protecting data privacy while sharing information across different entities. The goal is to create a cooperative information sharing ecosystem that can bridge the gaps between different domains and enable advanced intelligence for smart cities (Xiao et al., 2018).

2.3 Challenges of achieving semantic interoperability

The main concern of the DT is the data and data management systems. Much research has been conducted focusing on the issue specifically based on IoT sensors and its data collection. *The main technological infrastructure is indeed dedicated to sensors and data management like FIWARE or Azure DT* (Sottet & Pruski, 2023).

However, ensuring data quality remains a challenge, as the data collected from different sources may vary in accuracy, completeness, and consistency. Another challenge is the interoperability of different data management infrastructures, which may use different data formats and standards (Nilsson & Sandin, 2018). Overcoming these challenges requires collaboration between different stakeholders, including government agencies, private companies, and academic institutions, to establish common standards and protocols for data collection, management, and sharing. With the right approach, semantic interoperability can unlock the full potential of digital twins.

Considerably crucial issues include data security. DTSC would remain to be attracted to hackers as it surrounds by potential information from all around the world. At the same time, DTSC is extremely hard to secure as each point of vulnerability can open a gate to overall DTSC (Raghavan et al., 2020).

Data analysis is essential as it enhances both interoperability and DT autonomy generally. The challenge arises from the data itself, which poses problems during collection, cleaning, and structuring before being fused for deeper meaning and used to process intelligent tasks (Raghavan et al., 2020)(Boje et al., 2020).

Without a doubt, implementing semantic interoperability can be costly, especially for smaller cities or organizations with limited resources (Yang et al., 2021). However, financial perspective of implementation is beyond the scope of this paper. Rather, this research focuses on ontology of semantic interoperability.

Semantic interoperability is important because it makes it possible for different systems, such as DT and non-DT systems like Computerized Maintenance Management System (CMMS) or Enterprise Resource Planning (ERP), to communicate and share information in a useful way. It makes it possible to combine data and models from different fields, which is necessary for making decisions in asset maintenance management that are based on DT. Without semantic interoperability, people may not agree on what data and models mean, which can lead to wrong interpretations and bad decisions. Because of this, semantic interoperability is a necessity for achieving efficient and proactive maintenance strategies that can improve the availability of assets and reduce downtime in a way that is both cost-effective and time-efficient (Ariansyah & Pardamean, 2022).

3 Reference Model

Atkinson et al (Atkinson et al., 2022) in their paper, proposed a reference model (Figure 3) for building an interoperable system of systems (SoS) that supports the implementation of smart cities and digital twins. The model consists of three components: Global Domain, Operational SoS, and Consumer System. The Global Domain refers to the overall environment in which the SoS operates, including the physical and digital infrastructure, policies, regulations, and stakeholders. The Operational SoS represents the interconnected systems that make up the smart city or digital twin, such as sensors, data platforms, and analytics tools. The Consumer System refers to the end-users who interact with the SoS, such as citizens, businesses, and government agencies. The proposed reference model aims to address the technical, semantic, and governance issues related to data management in smart cities and digital twins. This includes issues such as data collection, retrieval, exchange, analysis, and processing, as well as data sovereignty, semantics, features, and metadata specifications.

The authors Atkinson et al (Atkinson et al., 2022) also noted that, while current projects and frameworks have explored some of these issues and proposed solutions for subsets of them, there is no overall framework that connects all aspects in a unique model. The proposed reference model builds on current experiences and considers OGC standards and initiatives to provide open solutions for specific parts of the framework.

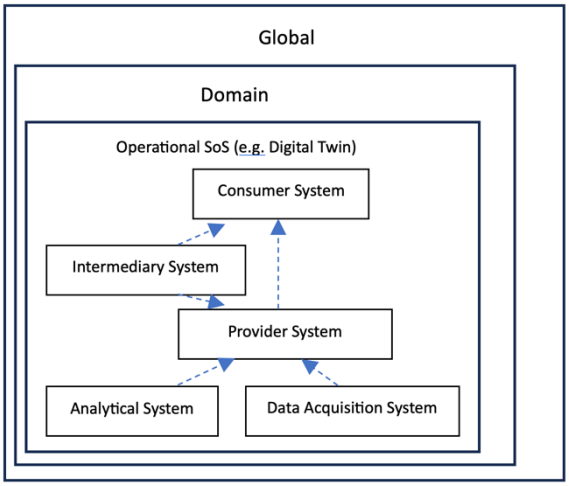


Figure 3. Scope Characterisation for SOS (Atkinson et al., 2022)

Boje et al (Boje et al., 2020) discusses the use of Building Information Modelling (BIM) and the concept of a Construction Digital Twin to improve the design, construction, and operation of buildings. While BIM provides a standardised way to represent building components and systems, it lacks semantic completeness in areas such as control systems and sensor networks. The authors propose a Construction Digital Twin as a more holistic and process-oriented approach that leverages the synchronicity of cyber-physical bi-directional data flows.

The paper by J. B. Correia, M. Abel, and K. Becker (Correia et al., 2023) suggests that semantic interoperability can be applied in smart cities by addressing the need for interoperability at all levels (semantic, data, and others) to provide fully integrated and optimized services for citizens. The interaction between DTs from different subdomains (smart buildings, urban planning) highlights the importance of data management between DTs.

To ensure consistency and accuracy in data interpretation, standardizing data models and ontologies that can be applied across various domains (Raghavan et al., 2020). This will also aid in the integration of different systems and facilitate communication between them.

Before expanding to larger areas, digital twin implementations are tested and improved in pilot projects (Chang & Jang, 2021).

4 Achieving Interoperability

There are numerous varieties of semantic interoperability, each of which enables distinct actions on the interoperability. This paper highlights on DTDL-based ontology.

The DTDL-based ontology for smart cities provides a modeling guideline for creating new entities, using English terms, camel case syntax for attribute names, and capital letters for entity type names. It also allows describing relationships between twins, which are digital representations of real-world environments brought to life with real-time data from sensors and other data sources.

The ontology is used to build Azure Digital Twins-based solutions and bring them to life in a live execution environment. The ontology is open source and aims to drive openness and interoperability (Bloch et al., n.d.).

4.1 Implementing DTDL Guideline

The guidelines for creating new entities in the DTDL-based ontology for smart cities recommend checking if the entity already exists in the repository before creating a new one. The ontology uses English terms, preferably American English, and camel case syntax for attribute names. Entity type names must start with a capital letter, for example, Streetlight. The ontology also allows describing relationships between twins, which are digital representations of real-world environments brought to life with real-time data from sensors and other data sources.

Telemetry refers to the data emitted by any digital twin, be it a regular stream of sensor readings, a computed stream of data, such as occupancy, or an occasional alert or information message. The following table details the properties that Telemetry may possess.

Property	Required	Data type	Limits	Description
`@type`	required	IRI		This must at least be "Telemetry"; it can also include a semantic type.
`@id`	optional	DTMI	max 2048 characters	An identifer for the Telemetry. If no @id is provided, one will be assigned automatically.
`comment`	optional	*string*	max 512 characters	A comment for model authors.
`description`	optional	localizable *string*	max 512 characters	A localizable description for display.
`displayName`	optional	localizable *string*	max 64 characters	A localizable name for display.
`name`	required	*string*	max 64 characters; contains only alphanumerics and underscore, starting with a letter, ending with alphanumeric; must be unique for all contents in Interface	The programming name of the element.
`schema`	required	Schema	must be *double*, *float*, *integer*, or *long* when a semantic type is present	The data type of the Telemetry, which is an instance of Schema.
`unit`	required when semantic type is present; disallowed otherwise	Unit		The unit type for data associated with the element.

Figure 4. Telemetry Table (Douceur, 2023)

The following example demonstrates a simple Telemetry definition for a currency using the double data type, declared based on the properties described above table.

```
{
  "@type": "Telemetry",
  "name": "eur",
  "schema": "double"
}
```

Figure 5. Telemetry Definition

This example illustrates the serialized Telemetry data for the aforementioned Telemetry model definition when JSON is utilized to serialize Telemetry data.

```
eur: 42.5
```

Figure 6. JSON Serialized

The following example demonstrates a Telemetry definition that includes a Currency semantic type and a unit property.

```
{
  "@type": ["Telemetry", "Currency"],
  "name": "eur",
  "schema": "double",
  "unit": "EUR"
}
```

Figure 7. Telemetry Definition

Figure 9 represents an example of telemetry definition in DTDL that includes a currency semantic type and a unit property. The example includes the following properties: "@type": specifies the semantic type of the telemetry, which in this case is "Currency". "name": specifies the name of the telemetry, which in this case is "eur". "schema": specifies the data type of the telemetry, which in this case is "double". "unit": specifies the unit of measurement for the telemetry, which in this case is "EUR".

4.2 Authoring DTDL ontology

Microsoft's Azure Digital Twins' instructional document (Learn, 2023a) describes ontology as a collection of models that comprehensively describe a particular domain, such as manufacturing, building structures, IoT systems, smart cities, energy grids, and web content, etc.

DTDL plays a significant role in the semantic interoperability of digital twins of smart cities. DTDL provides a common modeling language that enables developers to describe twins in terms of the telemetry they emit, the properties they report or synchronize, and the commands they respond to (Berhane Russom, 2021)(Azure, 2021). DTDL also allows describing the relationship between twins, which is essential for interoperability and enabling data sharing between multiple domains (Berhane Russom, 2021)(Azure, 2021). DTDL-based ontologies can be used to provide a common representation of places, infrastructure, and assets, which is paramount for interoperability and data sharing (Berhane Russom, 2021)(Azure, 2021). The Microsoft Digital Twins Definition Language (DTDL) has a “context model” which is similar to the NGSI-LD (Digital Twin Hub, 2023). Using DTDL to describe any digital twin's abilities enables the platform and solutions to leverage the semantics of each digital twin (Alina Cartus et al., 2022). DTDL version 3 is being developed, along with a new open-source parser that supports v2 and v3 (Minguez Pablos (RIDO), 2023). Therefore, DTDL is a crucial tool for achieving semantic interoperability of digital twins of smart cities.

As mentioned in the reference model section, according to Atkinson et al (Atkinson et al., 2022), there are key aspects which has to be addressed to achieve interoperability. Microsoft Azure DTDL reference model aims to address the technical, semantic, and governance issues related to data management in smart cities and digital twins. This includes issues such as data collection, retrieval, exchange, analysis, and processing, as well as data sovereignty, semantics, features, and metadata specifications. DTDL provides a common modeling language that enables developers to describe twins in terms of the telemetry they emit, the properties they report or synchronize, and the commands they respond to (Learn, 2023b)(Berhane Russom, 2021). DTDL also allows describing the relationship between twins, which is essential for interoperability and enabling data sharing between multiple domains (Learn, 2023b).

5 Future Research

As DTDL is a language for describing digital twin models of smart devices, assets, spaces, and more, contributing to DTDL will enhance the semantic interoperability of the digital twins. Future research could propose new ways of using DTDL or suggest improvements to the language. According to Cavalieri et al (Cavalieri & Gambadoro, 2023) purpose of proposing new ways of using DTDL is to enhance the interoperability of Digital Twins through integration into the OPC UA (Open Platform Communications Unified Architecture) domain. The authors (Cavalieri & Gambadoro, 2023) believe that interoperability between Digital Twins and the OPC UA communication standard should be enabled to meet the main requirements of Industry 4.0.

6 Conclusion

Semantic interoperability certainly plays a significant role in the implementation and integration of digital twins in smart cities. It facilitates data exchange between digital twins and enhances the analysis of data. The review paper explored the topic of semantic interoperability of digital twins in smart cities. It highlights the importance of improving interoperability for communication and data exchange between different twins. Achieving semantic interoperability is essential as it provides accurate and

timely insights. And this leads to the existence of well-functioning smart cities. The paper introduces reference models that can be reliable in developing interoperability between digital twins in smart cities. It addressed technical, semantic, and governance issues related to interoperability. Further, it discussed challenges in achieving semantic interoperability. The paper focuses on Microsoft Azure Digital Twin Definition Language (DTDL) to enhance interoperability. It highlights the DTDL and demonstrates the practical application of DTDL in achieving semantic interoperability. It discussed future research ideas to further improve Azure DTDL.

Acknowledgments

This work has been supported by the Croatian Science Foundation under the project IP-2019-04-4864.

References

Alina Cartus, Both, M., Maisch, N., Müller, J., & Diedrich, C. (2022). Interoperability of semantically heterogeneous digital twins through Natural Language Processing methods. *CLIMA 2022*. https://doi.org/10.34641/clima.2022.143

Ariansyah, D., & Pardamean, B. (2022). Enhancing Interoperability of Digital Twin in the Maintenance phase of Lifecycle. *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 514–518. https://doi.org/10.1109/ICITISEE57756.2022.10057666

Atkinson, R. A., Zaborowski, P., Noardo, F., & Simonis, I. (2022). SMART CITIES – SYSTEMS OF SYSTEMS INTEROPERABILITY AND OGC ENABLERS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-4/W3-2022*, 19–26. https://doi.org/10.5194/isprs-annals-X-4-W3-2022-19-2022

Azure. (2021). *DTDL based ontology for Smart Cities*. Github. https://github.com/Azure/opendigitaltwins-smartcities

Berhane Russom, M. (2021). *Smart Cities Ontology for Digital Twins*. https://techcommunity.microsoft.com/t5/internet-of-things-blog/smart-cities-ontology-for-digital-twins/ba-p/2166585

Bloch, O., Russam, M. B., & Tant, G. D. (n.d.). *IoT Show: Smart Cities Ontology for Digital Twins* [Interview]. https://www.youtube.com/watch?v=GrwI4GIp7nI&ab_channel=MicrosoftIoTDevelopers

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction*, *114*, 103179. https://doi.org/10.1016/j.autcon.2020.103179

Cavalieri, S., & Gambadoro, S. (2023). Proposal of Mapping Digital Twins Definition Language to Open Platform Communications Unified Architecture. *Sensors*, *23*(4), 2349. https://doi.org/10.3390/s23042349

Chang, Y. S., & Jang, I. S. (2021). Technology Trends in Digital Twins for Smart Cities. *Electronics and Telecommunications Trends*, *36*(1), 99–108. https://doi.org/10.22648/ETRI.2021.J.360111

Correia, J. B., Abel, M., & Becker, K. (2023). Data management in digital twins: A systematic literature review. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-023-01870-1

Dataspaces. (2023). COMMON EUROPEAN DATA SPACES. *Dataspaces*. http://dataspaces.info/common-european-data-spaces/#page-content

Deren, L., Wenbo, Y., & Zhenfeng, S. (2021). Smart city based on digital twins. *Computational Urban Science*, *1*(1), 4. https://doi.org/10.1007/s43762-021-00005-y

Digital Twin Hub. (2023). A Survey of Industry Data Models and Reference Data Libraries. *Digital Twin Hub*. https://digitaltwinhub.co.uk/website_archive/a-survey-of-idms-and-rdls-telecommunication_ontologies_for_dts/

Douceur, J. (2023). *Digital Twins Definition Language (DTDL)*. Azure. https://github.com/Azure/opendigitaltwins-dtdl/blob/master/DTDL/v2/DTDL.v2.md#introduction

Du, J., Zhu, Q., Shi, Y., Wang, Q., Lin, Y., & Zhao, D. (2020). Cognition Digital Twins for Personalized Information Systems of Smart Cities: Proof of Concept. *Journal of Management in Engineering*, *36*(2), 04019052. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000740

Learn, M. (2023a). *Manage Azure Digital Twins models*. https://learn.microsoft.com/en-us/azure/digital-twins/how-to-manage-model

Learn, M. (2023b). Learn about twin models and how to define them in Azure Digital Twins. *Microsoft Learn*. https://learn.microsoft.com/en-us/azure/digital-twins/concepts-models

Lehtola, V. V., Koeva, M., Elberink, S. O., Raposo, P., Virtanen, J.-P., Vahdatikhaki, F., & Borsci, S. (2022). Digital twin of a city: Review of technology serving city needs. *International Journal of Applied Earth Observation and Geoinformation*, *114*, 102915. https://doi.org/10.1016/j.jag.2022.102915

Minguez Pablos (RIDO), R. (2023). Digital Twins Definition Language (DTDL) updates. *Microsoft Tech Community*. https://techcommunity.microsoft.com/t5/internet-of-things-blog/digital-twins-definition-language-dtdl-updates/ba-p/3751705

Nahhas, S. (2023). Potentials of semantic internet of things in smart cities: An overview and roadmap. *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*, 1–6. https://doi.org/10.1109/ICAISC56366.2023.10085121

Nilsson, J., & Sandin, F. (2018). Semantic Interoperability in Industry 4.0: Survey of Recent Developments and Outlook. *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*, 127–132. https://doi.org/10.1109/INDIN.2018.8471971

Raghavan, S., Simon, B. Y. L., Lee, Y. L., Tan, W. L., & Kee, K. K. (2020). Data Integration for Smart Cities: Opportunities and Challenges. In R. Alfred, Y. Lim, H. Haviluddin, & C. K. On (Eds.), *Computational Science and Technology* (Vol. 603, pp. 393–403). Springer Singapore. https://doi.org/10.1007/978-981-15-0058-9_38

Sottet, J.-S., & Pruski, C. (2023). Data and Model Harmonization Research Challenges in a Nation Wide Digital Twin. *Systems*, *11*(2), 99. https://doi.org/10.3390/systems11020099

Xiao, Z., Fu, X., & Goh, R. S. M. (2018). Data Privacy-Preserving Automation Architecture for Industrial Data Exchange in Smart Cities. *IEEE Transactions on Industrial Informatics*, *14*(6), 2780–2791. https://doi.org/10.1109/TII.2017.2772826

Yang, D., Karimi, H. R., Kaynak, O., & Yin, S. (2021). Developments of digital twin technologies in industrial, smart city and healthcare sectors: A survey. *Complex Engineering Systems*. https://doi.org/10.20517/ces.2021.06

Using Card Sorting to Classify Elements of the Security Risk Management Program for Medical Devices

Nadica Hrgarek Lechner
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
nhrgarek@foi.hr

Abstract. *Establishing a security risk management program for medical devices can be challenging. Medical device manufacturers must understand which elements of the program are required, optional, or unnecessary to make effective decisions. By prioritizing the most important elements, manufacturers can establish a robust security risk management program.*
This paper identifies 40 elements of the security risk management program for medical devices. Using closed card sorting, professionals within the medical device sector classified these elements into predefined categories. The study involved 53 participants who completed two card sorts using an online tool over a period of five weeks.

Keywords. card sorting, card sort, closed cart sorting, cybersecurity, medical devices, security risk management

1 Introduction

In recent years, regulators have been increasing their focus on cybersecurity of medical devices. Medical device cybersecurity regulations, standards, and guidelines keep rising worldwide. To cope with the new situation, medical device companies are adjusting the processes and structures they need to ensure compliance with relevant regulations.

Medical device manufacturers need to perform security risk management for medical devices that are subject to cybersecurity during their total product life cycle. This applies to new product development as well as to legacy medical devices that cannot be reasonably protected against current cybersecurity threats. Due to increasing regulatory requirements, establishing and maintaining a comprehensive security risk management program for medical devices can be a daunting task. An effective program needs to be integrated into the manufacturer's quality (management) system, as well as be robust and flexible to respond to changes.

This paper provides 40 elements to be considered when establishing and maintaining the security risk management program for medical devices. The elements have been categorized by study participants into predefined categories using closed card sorting. *Card sorts* are the simplest form of sorts, in that the entities being sorted are simply names on cards (Rugg & McGeorge, 1997, p. 84). There are two primary types of card sorts: open and closed. Open card sort allows the participants to create their own categories. In a closed card sort, participants have to sort a list of items into a predefined set of category names provided by a researcher. Participants are constrained and cannot update category names or add new categories.

The paper is organized as follows. Section 2 provides an overview of the security risk management program for medical devices. The research approach is described in section 3. Section 4 presents and discusses the key results of the study. The final section of the paper restates the research problem, summarizes the main findings from the study, discusses the implications of the research findings and the limitations related to the research problem, and proposes new directions for future research.

2 Security Risk Management Program for Medical Devices

Effective cybersecurity management is intended to reduce the risk to patients by decreasing the likelihood that device functionality is intentionally or unintentionally compromised by inadequate cybersecurity (FDA, 2014, p. 2). Ray (2021, p. 19) states that an alternative risk modeling approach for medical device cybersecurity is needed, with a threat modeling approach driving the identification of risk factors.

According to FDA guidances (2014, 2016), security risk management is applicable to:

- Medical devices that contain software (including firmware) or programmable logic,

- Software that is a medical device (including mobile medical applications),
- Medical devices that are considered part of an interoperable system, and
- Legacy devices.

Security risk management for medical devices is an ongoing process and not a one-time activity. The security risk management process shall include the following elements: security risk analysis, security risk evaluation, security risk control, evaluation of overall security residual risk acceptability, security risk management review, and production and post-production activities (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023, pp. 10–11). According to (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023, p. 12), production and post-production activities include: vulnerability monitoring process associated with manufacturer-developed software and third-party components, establishing a threat intelligence program or using membership to ISAOs or similar organizations of threat intelligence sources, security incident response plan, vulnerability disclosure and communication plans, establishment of a customer communication process, periodic reviews of security risk controls and the security landscape, and identification of vulnerabilities and development, testing, and deployment of security patches.

Having the security risk management program in place is necessary to take a systematic approach to managing security risks throughout the total medical device life cycle and to achieve regulatory compliance. The program helps to determine which security risks have the highest impact, and to mitigate risks and minimize damage, when a medical device security incident occurs. Once implemented, the security risk management program needs to be continuously updated and improved to keep up with regulatory changes in the medical device industry.

3 Research Approach

Card sorting is one of techniques that can be used for knowledge elicitation (Barrett & Edwards, 1995). According to Fincher & Tenenberg (2005, p. 89), card sorting is a categorization task. Spencer (2009, p. 69) recommends using 30–100 cards for card sorting. U.S. Department of Health and Human Services (2013) recommends 30 to 40 cards at the absolute outside, especially for an open sort. Tullis and Wood (2004) conducted a study to assess the minimum number of participants for a card-sorting study. They found that reasonable structures are obtained from 20–30 participants. Lantz et al. (2019, p. 654) found that the most efficient number of participants for method of card sorting or pairwise comparisons was approximately 10–15. Nielsen (2004) recommends to test 15 users for card sorting to reach a correlation of 0.90, 20 users to reach 0.93, and 30 users to reach 0.95.

Up to the author's best knowledge, this study is the first attempt to use card sorting to classify elements of the security risk management program for medical devices. The study was guided by the following research questions:

1. How can a security risk management program for medical devices be characterized?
2. How professionals within the medical devices sector from various backgrounds and roles would classify elements of the security risk management program for medical devices into logical categories?
3. What is the order of perceived needs for elements of the security risk management program for medical devices?
4. What are the respondents' views on the elements of the security risk management program for medical devices?

To answer the first research question, a narrative literature review was conducted to identify elements of the security risk management program for medical devices. After reviewing cybersecurity standards, technical information reports, and guidances for medical devices and taking into consideration the results of a scoping study (Hrgarek Lechner, 2021), recommendations for creating a Product Cybersecurity Organization (Ray, 2021, pp. 22–23) and defining roles and responsibilities (Wirth et al., 2020), author's experience with the security risk management, and the recommended number of cards for card sorting, 40 elements of the security risk management program for medical devices were derived (refer to Table 1). Preparing the list of elements to produce a set of cards with brief descriptions was the most challenging and time-consuming activity.

The second, third, and fourth research questions were answered by conducting an online survey and analysing and interpreting collected quantitative data and qualitative data including participant comments. The survey was developed with Qualtrics Surveys tool. The survey was reviewed by one subject matter expert who is familiar with the security risk management process of medical devices. Two experts from academia were consulted to get feedback about the content and questionnaire design. After received feedback was addressed, the survey link was sent to potential survey participants. The survey contained two unmoderated card sorting activities where participants were asked to sort the same set of cards using a different criterion for the sorting each time. The order of cards was randomized to guard against sorting bias. Each card was presented using the carousel view in Qualtrics Surveys. After choosing a category for a card, a participant automatically progressed to the next card. The participants were allowed to skip cards.

Collecting responses from survey participants took five weeks and no personally identifiable information

and other identifiers (i.e., IP address, location data) were collected. Data was collected from a convenience sample of cybersecurity professionals with domain knowledge within the medical devices sector (i.e., practitioners, consultants, auditors, and academic researchers who are involved in security risk management tasks of medical devices and have at least one year of relevant experience in this area).

LinkedIn professional network was primarily used to identify and recruit qualified participants for the study. The survey link was sent individually to 328 potential participants via LinkedIn and to 4 potential participants via e-mail. It was anticipated that up to 20 survey respondents will be recruited. Survey scams and phishing links are common concerns when using online surveys, especially when cybersecurity professionals participate in online surveys. Few contacted people expressed security concerns about clicking on the provided survey link and refused to participate. One participant asked a question that only the survey author could answer to get confidence that the author's LinkedIn account was not hacked and the survey link is safe.

The total number of individuals who attempted the card sort was 57. The results presented in this paper are from 53 survey participants who truly attempted the sort. Section 4 summarizes and discusses the results of the card sorting exercises.

Table 1. Elements of the security risk management program for medical devices

Nr.	Element	Source
1	**Compliance with applicable cybersecurity laws, regulations, standards, and guidances** for medical devices	Author's own work
2	**Monitoring of regulatory changes and developments** in cybersecurity laws, regulations, standards, and guidances for medical devices	Author's own work
3	**A dedicated team of product security professionals**	Derived from (Ray, 2021)
4	**Qualified and trained personnel** performing security risk management tasks	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023)
5	**Organizational roles and responsibilities** for security risk management	Derived from (*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023), and (Wirth et al., 2020)
6	**Secure product development life cycle/framework**	(*IEC 81001-5-1: Health software and health IT systems safety, effectiveness and security – Part 5-1: Security – Activities in the product life cycle*, 2021), (FDA, 2022)
7	**Secure product development training for employees**	Author's own work
8	**Glossary of terms and definitions** relating to medical device cybersecurity	Author's own work
9	**Security risk management process** that is coordinated with other medical device risk management processes	(*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023)
10	**Integrated security risk management process into a quality (management) system**	(*IEC 81001-5-1: Health software and health IT systems safety, effectiveness and security – Part 5-1: Security – Activities in the product life cycle*, 2021), (FDA, 2022)
11	**Integrated security risk management process with accompanying processes** such as vulnerability handling, vulnerability disclosure, incident response, etc.	(FDA, 2016), (FDA, 2022)
12	**Internal/quality audits** of the security risk management program	Derived from (FDA, 2022)
13	**Measures and metrics** for processes that reduce the number and severity of vulnerabilities in products	(FDA, 2022)
14	**Software tools** to support security risk management tasks	Derived from (FDA, 2016)
15	**Security risk assessment** of a product	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016)
16	**Security risk assessment of third-party software/firmware components** incorporated within a product	(FDA, 2022)
17	**Security assessment of unresolved software anomalies** that exist in a product at the time of regulatory submission	(FDA, 2022)
18	**Threat modeling process**	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (FDA, 2016),

Nr.	Element	Source
		(*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019), (Medical Device Coordination Group, 2020), (*IEC 81001-5-1: Health software and health IT systems safety, effectiveness and security – Part 5-1: Security – Activities in the product life cycle*, 2021), (FDA, 2022), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023)
19	Determined **security risk controls** to reduce security risks	(FDA, 2014), (*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (FDA, 2022), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023)
20	Implemented and tested **security risk controls**	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (FDA, 2022)
21	**Security architecture** providing the security context and trust boundaries of a medical device system	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (FDA, 2022)
22	**Security requirements** for the product under development	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (FDA, 2022)
23	**Labeling** to communicate relevant security information to users of medical devices	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (FDA, 2022)
24	**Software Bill of Materials (SBOM)**	(*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019), (Medical Device Coordination Group, 2020), (*IEC 81001-5-1: Health software and health IT systems safety, effectiveness and security – Part 5-1: Security – Activities in the product life cycle*, 2021), (FDA, 2022), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023), (FDA, 2023)
25	**Premarket security testing** to identify and address potential vulnerabilities prior to exploitation	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (Medical Device Coordination Group, 2020), (*IEC 81001-5-1: Health software and health IT systems safety, effectiveness and security – Part 5-1: Security – Activities in the product life cycle*, 2021), (FDA, 2022)
26	**Security risk management artifacts** (e.g., security risk management plan, security risk analysis, security risk management report, etc.)	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023)
27	**Security risk management documentation** for regulatory submissions	(FDA, 2014), (FDA, 2022)
28	**Security training for users** of medical devices	Derived from (FDA, 2016) and (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023)
29	**Post-market periodic security testing**, including penetration testing	Derived from (FDA, 2022)
30	**Monitoring of third-party software/firmware components** incorporated within a product to identify and detect potential vulnerabilities	(FDA, 2016), (*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019)
31	**Monitoring of cybersecurity information sources** to identify and detect potential security threats and vulnerabilities that may affect medical devices	(FDA, 2016), (*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019)
32	Medical device manufacturers participation in a health focused **Information Sharing Analysis Organization (ISAO)**	(FDA, 2016), (*AAMI TIR57: Principles for medical device security–Risk management*, 2016), (*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023)
33	**Vulnerability management process**	Derived from (FDA, 2016) and (*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019)
34	**Vulnerability management plans**	(FDA, 2022)
35	**Security risk assessment of post-market vulnerabilities**	(FDA, 2016)
36	**Patch management process** for providing post-market security patches and updates	(*AAMI TIR57: Principles for medical device security–Risk management*, 2016),

Nr.	Element	Source
		(*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019), (*IEC 81001-5-1: Health software and health IT systems safety, effectiveness and security – Part 5-1: Security – Activities in the product life cycle*, 2021)
37	**Product security incident response process**	Derived from (FDA, 2016) and (*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019)
38	**Coordinated vulnerability disclosure process**	(FDA, 2016), (*AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers*, 2019), (*ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers*, 2023), (FDA, 2023)
39	**Post-market surveillance system** including cybersecurity considerations	(Medical Device Coordination Group, 2020)
40	**Vigilance process** for reporting serious incidents and field safety corrective actions related to cybersecurity incidents	(Medical Device Coordination Group, 2020)

4 Results and Discussion

As shown in Fig. 1, a total of 68 Qualtrics Survey responses were recorded during the data collection period from May 8th, 2023 to June 11th, 2023.

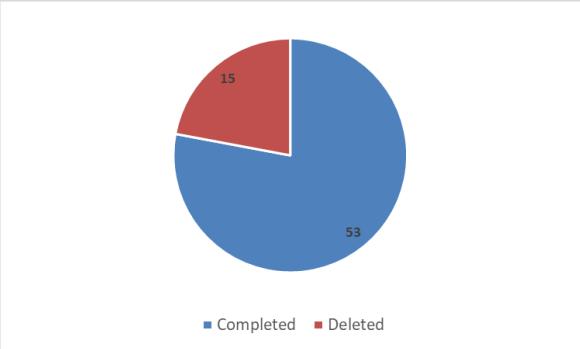


Figure 1. Recorded responses by status

Prior to data analysis, all recorded responses in Qualtrics Survey were screened. Survey data cleaning was performed to identify and remove preview responses as well as incomplete and suspicious responses from participants who don't match target audience criteria, did not complete the card sorting exercises, had only a small number of cards 'sorted' leading to incomplete sorts, or offered nonsensical feedback in open-ended questions.

15 recorded responses were deleted from dataset (Fig. 1) in the following cases:

- 2 preview responses were used only for testing the survey and not to record real data.
- 1 participant did not give consent and no data was collected.
- 2 participants gave consent, but did not answer any survey question.
- 1 participant had less than 1 year of experience and participated only in 1 medical device project that involved cybersecurity. In addition, the participant placed 23 out of 40 cards consecutively in one category in the first card sorting exercise and only sorted one card in the second card sorting exercise.
- 6 participants answered the demographic and context questions, but did not complete any card sorting exercise.
- 3 participants only partially completed card sorting exercises leading to incomplete sorts.

Participants who do not have adequate familiarity with the items being sorted may reduce the effectiveness of card sorting. Two participants had less than 1 year of experience in security risk management activities of medical devices and did not meet target audience criteria. However, they were familiar with the security risk management process of medical devices and participated in 2-5 medical device projects that involved cybersecurity. For these reasons, their responses were not deleted from dataset.

Demographic characteristics of survey participants are illustrated in Fig. 2 – Fig. 6. Fig. 7 and Fig. 8 present context data.

As shown in Fig. 2, the large majority of the respondents (69.8%) represented companies with more than 1000 employees.

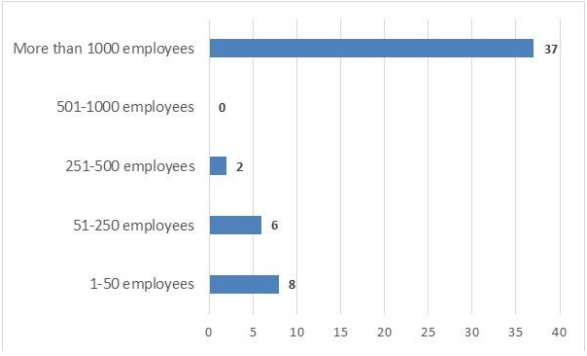


Figure 2. Responses per company size

Fig. 3 represents the geographic location of company's headquarters.

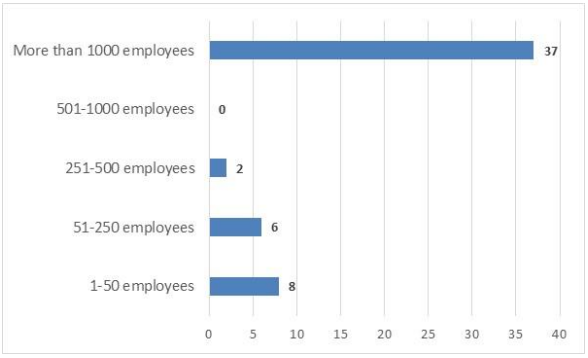


Figure 3. Responses per company headquarters location

Security risk management of medical devices involves participation of stakeholders from various areas. Participants were asked to select one or more applicable job roles and their answers are displayed in Fig. 4. Four respondents specified other job roles as follows: Quality Management, Head of Development, Regulatory Consultant, Director of Product Security.

Figure 4. Responses per current job role within the organization

As shown in Fig. 5, the majority of respondents were very familiar (45.2%) and extremely familiar (39.6%) with the security risk management process for medical devices.

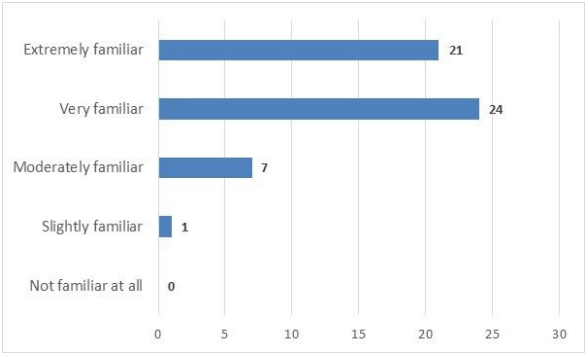


Figure 5. Responses per familiarity with the security risk management process for medical devices

As depicted in Fig. 6, 39.6% of respondents had more than 6 years of experience in security risk management activities of medical devices.

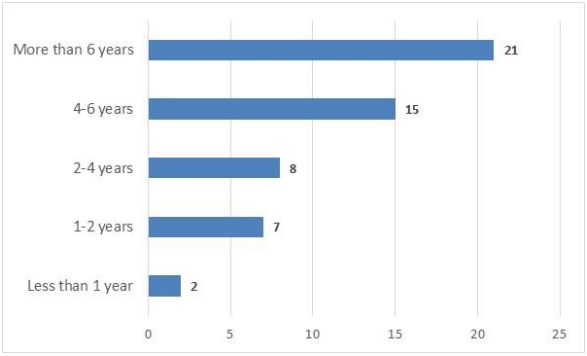


Figure 6. Responses per years of experience in security risk management activities of medical devices

As illustrated in Fig. 7, 50.9% of respondents participated in more than 10 projects involving cybersecurity.

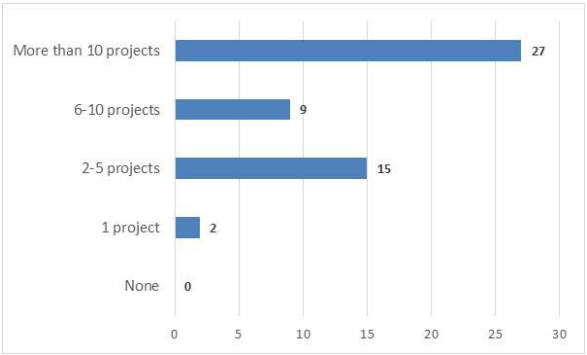


Figure 7. Responses per number of medical device projects involving cybersecurity

According to Fig. 8, the majority of respondents used hybrid approach (59.6%) to assess security risk in medical device projects. Quantitative approach was used by only 13.5% of respondents. 1 of 53 participants did not provide answer.

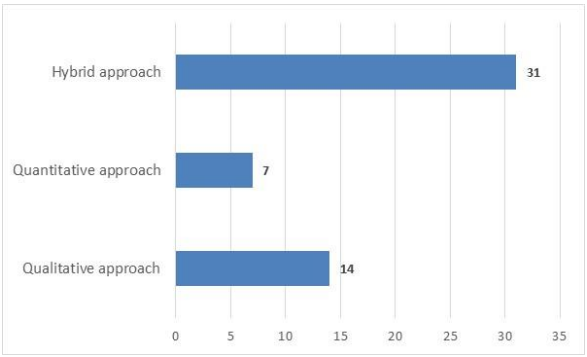


Figure 8. Responses per approach to assess security risk in medical device projects

Survey participants were asked to categorize a set of 40 items (i.e., cards) representing elements of the security risk management program for medical devices into predefined categories using a different criterion for the sorting each time. The goal was to understand how the participants would group these items into categories. In total, 52 responses were collected for the first card sorting exercise with five predefined categories and 51 responses were collected for the second card sorting exercise with three predefined categories.

In the first card sort, survey participants were asked to categorize a set of 40 cards into five predefined categories: "Regulatory compliance", "Product security governance", "Security risk management", "Total product life cycle", *"Uncategorized"*. Participants were instructed to use a special category *"Uncategorized"* for cards that don't fit with anything. Table 2 summarizes what percentage of respondents placed each card into each predefined category and shows the most popular clusters of cards for each category based on where respondents placed the cards most often. The cards in the matrix were reordered according to which cards have the highest percentages among all of the categories and the cards with the same popular category are clustered together. Within the clusters, the cards are sorted in descending order by the percentage of respondents who placed them in that category. The percentage of each grouping within the cluster is also visually represented by different shades of grey colour as follows: ☐ 0%, ☐ 1–25%, ☐ 26–50%, ☐ 51–75%, ☐ 76–100%. The darker the shade, the more the percentage. In two cases, a card was placed into two different categories at equal percentages (refer to table cells marked with a dashed rectangle). Individual responses were screened and the cards were clustered to the category that received responses from a higher number of respondents who rated their familiarity with the security risk management process for medical devices as extremely familiar.

Table 2 provides an overview about all categories that a card was placed. It shows four clusters with higher agreement rates for the specific cards most respondents agree to belong to those predefined categories. "Regulatory compliance" and "Security risk management" are the most popular categories. Out of total 40 cards, the respondents placed 15 cards (37.5%) into the special category *"Uncategorized"*. Most cards were placed in a number of categories. This may indicate that categories overlap or card labels are not clearly defined. Feedback was received from two participants who attempted the first card sort that some of the categories are overlapping.

Table 2. Agreement matrix of the first card sorting exercise

Card #	Card name	Category				
		Regulatory compliance	**Product security governance**	**Security risk management**	**Total product life cycle**	***Uncategorized***
27	**Compliance with applicable cybersecurity laws, regulations, standards, and guidances** for medical devices	86.5%	9.6%	0%	3.8%	0%
10	**Monitoring of regulatory changes and developments** in cybersecurity laws, regulations, standards, and guidances for medical devices	78.8%	7.7%	3.8%	7.7%	1.9%
30	**Security risk management documentation** for regulatory submissions	67.3%	5.8%	19.2%	7.7%	0%
14	**Labeling** to communicate relevant security information to users of medical devices	48.1%	11.5%	19.2%	17.3%	3.8%
6	**Vigilance process** for reporting serious incidents and field safety corrective actions related to cybersecurity incidents	46.2%	19.2%	11.5%	23.1%	0%
38	**Internal/quality audits** of the security risk management program	40.4%	26.9%	13.5%	11.5%	7.7%
31	**Organizational roles and responsibilities** for security risk management	11.5%	53.8%	28.8%	1.9%	3.8%
34	**Measures and metrics** for processes that reduce the number and severity of vulnerabilities in products	1.9%	51.9%	26.9%	15.4%	3.8%
2	**A dedicated team of product security professionals**	1.9%	50%	26.9%	21.2%	0%
7	**Secure product development training for employees**	17.3%	44.2%	7.7%	23.1%	7.7%
39	**Coordinated vulnerability disclosure process**	17.3%	38.5%	21.2%	19.2%	3.8%
40	**Vulnerability management plans**	3.8%	36.5%	30.8%	26.9%	1.9%
28	**Glossary of terms and definitions** relating to medical device cybersecurity	13.5%	34.6%	9.6%	9.6%	32.7%

Card #	Card name	Category				
		Regulatory compliance	**Product security governance**	**Security risk management**	**Total product life cycle**	***Uncategorized***
12	Medical device manufacturers participation in a health focused **Information Sharing Analysis Organization (ISAO)**	19.2%	30.8%	19.2%	15.4%	15.4%
13	**Product security incident response process**	23.1%	25%	25%	23.1%	3.8%
17	Determined **security risk controls** to reduce security risks	1.9%	5.8%	84.6%	7.7%	0%
21	**Security risk assessment** of a product	3.8%	11.5%	78.8%	5.8%	0%
24	**Security risk management artifacts** (e.g., security risk management plan, security risk analysis, security risk management report, etc.)	13.5%	7.7%	75.0%	3.8%	0%
5	**Threat modeling process**	3.8%	11.5%	71.2%	11.5%	1.9%
22	Implemented and tested **security risk controls**	0%	7.7%	63.5%	25%	3.8%
37	**Security risk assessment of third-party software/firmware components** incorporated within a product	3.8%	17.3%	61.5%	17.3%	0%
19	**Security risk assessment of post-market vulnerabilities**	9.6%	9.6%	59.6%	21.2%	0%
3	**Security risk management process** that is coordinated with other medical device risk management processes	11.5%	3.8%	55.8%	25%	3.8%
23	**Security assessment of unresolved software anomalies** that exist in a product at the time of regulatory submission	11.5%	9.6%	53.8%	21.2%	3.8%
20	**Software tools** to support security risk management tasks	0%	17.3%	51.9%	19.2%	11.5%
33	**Integrated security risk management process with accompanying processes** such as vulnerability handling, vulnerability disclosure, incident response, etc.	3.8%	32.7%	42.3%	19.2%	1.9%
15	**Monitoring of cybersecurity information sources** to identify and detect potential security threats and vulnerabilities that may affect medical devices	1.9%	30.8%	42.3%	25%	0%
36	**Vulnerability management process**	3.8%	25%	40.4%	28.8%	1.9%
18	**Qualified and trained personnel** performing security risk management tasks	11.5%	36.5%	38.5%	9.6%	3.8%
26	**Premarket security testing** to identify and address potential vulnerabilities prior to exploitation	7.7%	25%	38.5%	28.8%	0%
9	**Security architecture** providing the security context and trust boundaries of a medical device system	1.9%	30.8%	34.6%	32.7%	0%
1	**Secure product development life cycle/framework**	1.9%	13.5%	7.7%	76.9%	0%
35	**Patch management process** for providing post-market security patches and updates	1.9%	13.5%	19.2%	65.4%	0%
25	**Monitoring of third-party software/firmware components** incorporated within a product to identify and detect potential vulnerabilities	7.7%	19.2%	26.9%	44.2%	1.9%
4	**Security requirements** for the product under development	13.7%	15.7%	25.5%	43.1%	1.9%
32	**Post-market periodic security testing**, including penetration testing	11.5%	17.3%	30.8%	38.5%	1.9%
11	**Post-market surveillance system** including cybersecurity considerations	36.5%	11.5%	15.4%	36.5%	0%
8	**Software Bill of Materials (SBOM)**	25%	19.2%	19.2%	32.7%	3.8%
16	**Security training for users** of medical devices	21.2%	21.2%	13.5%	26.9%	17,3%
29	**Integrated security risk management process into a quality (management) system**	19.2%	21.2%	25%	26.9%	7.7%

In the second card sort, survey participants were asked to categorize the same set of 40 cards into three predefined categories: “Required”, “Optional”, “Not necessary”. Table 3 shows the agreement matrix. A

value of 100% means that all respondents placed the card in the category. The agreement matrix shows two clusters with higher agreement rates for the specific cards most respondents agree to belong to those predefined categories. Out of total 40 cards, the respondents placed 4 cards (10%) into the category "Optional". It was expected that the card "**Security training for users** of medical devices" will be categorized as "Required". However, 55% of the respondents categorized this card as "Optional".

Table 3. Agreement matrix of the second card sorting exercise

Card #	Card name	Category		
		Required	**Optional**	**Not necessary**
1	**Secure product development life cycle/framework**	100%	0%	0%
17	Determined **security risk controls** to reduce security risks	100%	0%	0%
21	**Security risk assessment** of a product	100%	0%	0%
22	Implemented and tested **security risk controls**	100%	0%	0%
25	**Monitoring of third-party software/firmware components** incorporated within a product to identify and detect potential vulnerabilities	100%	0%	0%
27	**Compliance with applicable cybersecurity laws, regulations, standards, and guidances** for medical devices	100%	0%	0%
3	**Security risk management process** that is coordinated with other medical device risk management processes	100%	0%	0%
30	**Security risk management documentation** for regulatory submissions	100%	0%	0%
4	**Security requirements** for the product under development	100%	0%	0%
9	**Security architecture** providing the security context and trust boundaries of a medical device system	100%	0%	0%
11	**Post-market surveillance system** including cybersecurity considerations	98%	2%	0%
19	**Security risk assessment of post-market vulnerabilities**	98%	2%	0%
24	**Security risk management artifacts** (e.g., security risk management plan, security risk analysis, security risk management report, etc.)	98%	2%	0%
35	**Patch management process** for providing post-market security patches and updates	98%	2%	0%
13	**Product security incident response process**	96%	4%	0%
18	**Qualified and trained personnel** performing security risk management tasks	96%	4%	0%
37	**Security risk assessment of third-party software/firmware components** incorporated within a product	96%	4%	0%
6	**Vigilance process** for reporting serious incidents and field safety corrective actions related to cybersecurity incidents	96%	4%	0%
33	**Integrated security risk management process with accompanying processes** such as vulnerability handling, vulnerability disclosure, incident response, etc.	94%	6%	0%
23	**Security assessment of unresolved software anomalies** that exist in a product at the time of regulatory submission	92%	6%	2%
36	**Vulnerability management process**	92%	6%	2%
40	**Vulnerability management plans**	90%	10%	0%
10	**Monitoring of regulatory changes and developments** in cybersecurity laws, regulations, standards, and guidances for medical devices	88%	12%	0%
15	**Monitoring of cybersecurity information sources** to identify and detect potential security threats and vulnerabilities that may affect medical devices	88%	12%	0%
39	**Coordinated vulnerability disclosure process**	88%	12%	0%
14	**Labeling** to communicate relevant security information to users of medical devices	88%	10%	2%
5	**Threat modeling process**	88%	10%	2%
29	**Integrated security risk management process into a quality (management) system**	86%	14%	0%
26	**Premarket security testing** to identify and address potential vulnerabilities prior to exploitation	86%	12%	2%
8	**Software Bill of Materials (SBOM)**	84%	14%	2%
31	**Organizational roles and responsibilities** for security risk management	78%	22%	0%
32	**Post-market periodic security testing**, including penetration testing	78%	22%	0%
7	**Secure product development training for employees**	78%	20%	2%
38	**Internal/quality audits** of the security risk management program	73%	27%	0%
34	**Measures and metrics** for processes that reduce the number and severity of vulnerabilities in products	69%	31%	0%

Card #	Card name	Category		
		Required	Optional	Not necessary
2	**A dedicated team of product security professionals**	67%	31%	2%
20	**Software tools** to support security risk management tasks	33%	65%	2%
12	Medical device manufacturers participation in a health focused **Information Sharing Analysis Organization (ISAO)**	31%	59%	10%
16	**Security training for users** of medical devices	37%	55%	8%
28	**Glossary of terms and definitions** relating to medical device cybersecurity	41%	51%	8%

At the end of the card sorting exercises, the survey participants were asked to provide information what key elements of the security risk management program for medical devices are missing in the survey and to share any other thoughts about the security risk management. 7 of 53 respondents (13.2%) answered that no key elements are missing.

The following comments list the missing key elements according to the respondents:

- "Relevant Standards and Guidances"
- "Perhaps a specific register of cybersecurity regulations / standards that is used to define requirements for TPLC activities."
- "Check of regulatory requirements fulfilled"
- "secure design best practices, secure coding standards and their enforcement (e.g. code reviews, SAST), security event monitoring (for certain product types) if not already covered by PMS, software configuration management, release artifact archiving, security of IT infrastructure used for product/software development, test, production, delivery…, security of software distribution and software updates security (code integrity and security of code signing keys), supply chain security risk management (beyond assessment of 3rd-party components)"
- "Threat Modeling, Risk Management System, Risk Management Score, supply chain"
- "Security risk scoring, prioritisation, treatment of risk whether eliminated, mitigated, accepted or transferred."
- "evaluation of security risk with potential safety impact, legacy device security risk managment, product security continue support plan including retirement and obsolescence"
- "alignment with IEC14971 and Patient risk management process"
- "Who and how to connect security risks with safety."
- "Design Reviews, Digital Signature handling (File Integrity and Private Key Protection)"
- "Cryptography / protocol design considerations. It's a bad idea to "roll your own" or even implement your own instead of using a library, but correct library usage should still be monitored, the correct choice of algorithms and key lengths should be as well, and if implementing your own for whatever reason, a cryptography exert should be consulted on the protocol design and a cryptography implementation expert should be consulted on implementation details."
- "…Only suggestion would be to differentiate between pre- and post-market risk and vulnerability management. Although there is some overlap there are also some very specific differences with pre-market being more tied in with engineering processes whereas post-market being more about customer communication and management. Also, there was a question about post-market pen testing. I would see pen testing more as a pre-market activity, although one could make the argument of periodically repeating certain assessment activities. So ... I am not necessarily disagreeing, just wanted to clarify."
- "Multi-year architecture strategy to prepare for long product life, SBOM based monitoring for new vulnerabilities in post-market, Strategy for secure connectivity over Internet"
- "SLAs"
- "…Penetration testing wasn't covered"
- "Maybe clearly identify penetration and fuzz testing and software composition analysis."
- "The first section differentiated between qualitative and quantitative risk severity assessments, but I think it could have been mentioned in the following two sections."

The following key elements of the security risk management program for medical devices were mentioned by respondents as missing, but they were present in the card set:

- threat modeling (card #5);
- secure design best practices, secure coding standards, code reviews, software configuration management, design reviews, and security of IT infrastructure are best practices for a secure product development life cycle/framework (card #1);
- connecting security risks with safety, evaluation of security risk with potential safety impact, alignment with ISO 14971 (*ISO 14971: Medical devices – Application of risk management to medical devices*, 2019) and patient risk management process, legacy device security risk management, security risk scoring, risk management score, prioritisation, and security risk control options are associated with the security risk management process (card #3);

- penetration and fuzz testing were implicitly covered (card #26);
- penetration testing was explicitly covered (card #32);
- SBOM based monitoring for new vulnerabilities needs to be conducted during both premarket and post-market phases (card #25);
- differentiation between premarket and post-market risk management (card #19, #21);
- vulnerability management (card #11, #36, #39, #40).

Relevant cybersecurity standards and guidances were not written on the individual card #27 on purpose to keep it concise. Some key elements were mentioned as missing (e.g., security event monitoring, cryptography, code integrity, file integrity, secure software updates), but they are examples of security risk controls (card #17). Each manufacturer can choose its approach for performing a security risk assessment (e.g., quantitative, etc.) and the label of card #21 did not contain this information. One respondent mentioned a service-level agreement (SLA) which is a contract between the service provider (e.g., medical device manufacturer) and the customer (e.g., health delivery organization). Such agreement may contain specific security requirements that may not be considered during design and development. Supply chain risk management was also missing in the card set.

The following comments list the thoughts about the security risk management from the respondents:

- "follow some guidelines TIR57, NIST RMF, NIST CSF"
- "The security risk management should be implemented as a part of the overall product and process risk management"
- "It was hard for me to separate security risk management from total product lifecycle. For me, the latter includes the first ;-)"
- "When asked to categorize stuff under "most appropriate" labels, there were a number of items that were either detail-dependent (e.g. nature of the product) and/or fit equally well into two or more labels."
- "Security risk management and product security governance may all form a combined organization. The main decision point for certain aspects, like a dedicated team of experts central vs dispersed in product organizations entirely depends on the company and their ability to maintain expertise dispersed but consistent. There is no standard most effective model; key is for all elements of the program to be present, consistent and coordinated."
- "Contracts are a very important aspect of system security."
- "The current methodology is still very confusing and difficult to implement, especially for medical devices that privacy/data protection is also involved."
- "Risk transfer is a key element, in case a risk cannot be handled at manufacturer. Risk should not be accepted on behalf of end customer. Risk transfer along with suggestions to address the risk at user's end should be well documented."
- "Security controls should be appropriate for the risk introduced by the device into the system it will operate in and the function it performs. This may be best achieved by isolating clinical control from external comms to the maximum extent possible"
- "There is a serious gap in subject matter experts when it comes to cybersecurity in medical devices, most people think of corporate cybersecurity and not product security which leads to gaps in knowledge when trying to hire."
- "…I think we are all waiting to see what, if any, risk considerations greater AI integration in medical devices will bring. In the US the FDA rapidly advancing the regulatory criteria and it is something that is important to keep an eye on."

5 Conclusion and Future Work

This paper presents the use of the closed card sorting technique to classify 40 elements of the security risk management program for medical devices. The study used individual card sorting and involved a total of 40 cards. Every card had a name which represented an element of the security risk management program for medical devices. The analysis of closed card sort data and collected qualitative data was useful to elicit knowledge and to find out how the practitioners would distribute the cards within the predefined categories. Card sorts were helpful to better understand the connection between categories and various elements of the security risk management program for medical devices, and to determine which elements are suitable, missing, or need improvement.

LinkedIn professional network can be used to identify and recruit qualified participants for the survey. LinkedIn profiles provide valuable information about members, including their job title, location, industry, and experience, all of which are publicly available. When recruiting a large number of potential participants, a LinkedIn Premium account is required to browse for profiles without limits and to use InMail messages to directly message another LinkedIn member that you are not connected to.

When conducting an online survey research that addresses sensitive topics such as cybersecurity of medical devices, it is recommended to use an anonymous survey. The respondents will feel freer and more comfortable raising their concerns and giving honest, detailed answers with no inhibitions.

When using online surveys, it is recommended to use security survey options afforded by the tool (e.g.,

bot detection, preventing multiple submissions by the same person, preventing security scanners from accidentally starting a new session on the survey, etc.). While convenient and efficient, online surveys are subject to security concerns (e.g., survey scams, phishing links) from the potential survey participants, especially when the recipients of a survey link don't know the person who sent the link.

Card sort results and qualitative data including respondent comments will be considered when designing a new conceptual framework for managing security risks of medical devices. Categorizing with overlapping categories needs further investigation.

Acknowledgments

The author extends their gratitude to the survey participants, the reviewers of the survey, and the anonymous reviewers of the paper for their invaluable support of the research.

References

AAMI TIR57: Principles for medical device security–Risk management. (2016).

AAMI TIR97: Principles for medical device security–Postmarket risk management for device manufacturers (2019).

ANSI/AAMI SW96: Standard for medical device security–Security risk management for device manufacturers. (2023).

Barrett, A.R. & Edwards, J.S. (1995). Knowledge elicitation and knowledge representation in a large domain with multiple experts. Expert Systems with Applications, 8(1), 169–176.

FDA. (2014). Content of Premarket Submissions for Management of Cybersecurity in Medical Devices – Guidance for Industry and Food and Drug Administration Staff. Retrieved from https://www.fda.gov/media/86174/download

FDA. (2016). Postmarket Management of Cybersecurity in Medical Devices – Guidance for Industry and Food and Drug Administration Staff. Retrieved from https://www.fda.gov/media/95862/download

FDA. (2022). Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions – Draft Guidance for Industry and Food and Drug Administration Staff. Retrieved from https://www.fda.gov/media/119933/download

FDA. (2023). Cybersecurity in Medical Devices: Refuse to Accept Policy for Cyber Devices and Related Systems Under Section 524B of the FD&C Act – Guidance for Industry and Food and Drug Administration Staff. Retrieved from https://www.fda.gov/media/166614/download

Fincher, S. & Tenenberg, J. (2005). Making sense of card sorting data. Expert Systems, 22(3), 89–93.

Hrgarek Lechner, N. (2021). The Key Elements of a Risk-based Product Security Program for Medical Devices: A Scoping Study. In N. Vrček, E. Pergler, & P. Grd (Eds.), Proceedings of the 32nd Central European Conference on Information and Intelligent System (CECIIS 2021) (pp. 257–264). University of Zagreb, Faculty of Organization and Informatics Varaždin.

IEC 81001-5-1: Health software and health IT systems safety, effectiveness and security – Part 5-1: Security – Activities in the product life cycle. (2021).

ISO 14971: Medical devices – Application of risk management to medical devices. (2019).

Lantz, E., Keeley, J. W., Roberts, M. C., Medina-Mora, M. E., Sharan, P., & Reed, G. M. (2019). Card sorting data collection methodology: How many participants is most efficient?. Journal of Classification, 36(3), 649–658.

Medical Device Coordination Group. (2020). MDCG 2019-16 Rev. 1 Guidance on Cybersecurity for medical devices. Retrieved from https://health.ec.europa.eu/system/files/2022-01/md_cybersecurity_en.pdf

Nielsen, J. (2004, July 18). Card Sorting: How Many Users to Test. Nielsen Norman Group. https://www.nngroup.com/articles/card-sorting-how-many-users-to-test/

Ray, A. (2021). Cybersecurity for Connected Medical Devices. Eastbourne: Academic Press.

Rugg, G. & McGeorge, P. (1997). The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. Expert Systems, 14(2), 80–93.

Spencer, D. (2009). Card sorting: Designing usable categories. New York: Rosenfeld Media.

Tullis, T. & Wood, L. (2004, June 7–11). How Many Users Are Enough for a Card-Sorting Study? [Conference paper]. Usability Professionals Association (UPA) 2004 Conference, Minneapolis, MN, United States.

U.S. Department of Health and Human Services. (2013, October) Card Sorting. https://www.usability.gov/how-to-and-tools/methods/card-sorting.html

Wirth, A., Gates, C., & Smith, J. (2020). Medical Device Cybersecurity for Engineers and Manufacturers. Boston: Artech House.

AI techniques and tools in Agile Software Development: Preliminary research

Dijana Peras, Zlatko Stapić, Mislav Matijević
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{dijana.peras, zlatko.stapic, mislav.matijevic}@foi.unizg.hr

Abstract. *In recent years, there has been significant growth in research related to Agile software development (ASD), accompanied by a notable rise in the adoption of artificial intelligence (AI) tools and techniques. AI is believed to have the potential to bring about a transformation in agile software development, leading to improved product quality, increased production efficiency, and higher project success rates. Consequently, there is a compelling need to incorporate AI methods and tools into the agile software development process. This study presents the results of a literature review and answers research questions on AI techniques and tools, their purposes, and their benefits when used in the context of ASD. In a multi-phased process, a total of 374 documents were gathered and examined. 28 papers satisfied the inclusion and quality assessment requirements. A total of 7 different AI techniques were identified, of which machine learning (ML) was the one predominantly used. The purposes and benefits of AI techniques were identified and discussed. Recommendations for future research were given to tackle detected research gaps.*

Keywords. Artificial Intelligence, AI, Agile Software Development, Agile Framework, Agile Methodology

1 Introduction

The effective creation of high-quality software has become essential for the software industry as software applications are becoming increasingly popular across every aspect of society. To deliver better software, boost efficiency, and assure project success, agile software development (ASD) could be enhanced with artificial intelligence (AI) solutions. Software development teams may take advantage of AI in a variety of ways, including automation of specific tasks, delivery of project insights, identification of phases of the agile process that are prone to risks, prioritization of effort, etc. (Sofian et al., 2022).

AI techniques like machine learning (ML), heuristic algorithms (HA), deep learning (DL), data mining (DM), data analytics (DA), and natural language processing (NLP) have been widely explored in various software engineering (SE) phases, which involve different activities across all stages of the Software Development Lifecycle (SDLC) (Sofian et al., 2022). Several systematic and mapping literature reviews were published (Biesialska et al., 2021; Perkusich et al., 2020; Sofian et al., 2022), exploring the overall trends in AI techniques and their application to SE. However, no clear distinction was made in AI usage within traditional and agile software development processes, which operate under entirely different assumptions and therefore require special approaches.

This paper aims to provide a preliminary review of AI techniques and tools used in ASD, as well as to investigate the purposes for which AI is used in the context of ASD and the benefits it provides for businesses and users. The distinction is made among agile processes and activities to inspect the scope of AI integration. A broader goal of the paper is to identify research gaps and to build the foundation for the proposal of AI supported framework for agile software development.

The paper consists of five sections. First, a short overview of relevant background on AI and ASD is given. In the second section, the methodology is presented. The third section presents the results of the research. The discussion is made in the fourth chapter. The last chapter presents the main conclusions.

2 Research methodology

A literature review was conducted by three researchers. To maximize the efficiency of the research, two researchers were engaged in each phase of the process. The literature review aimed to identify AI techniques and tools used within ASD. The Literature Review guidelines proposed by Templier and Paré (Templier & Paré, 2015) were followed. The study's protocol is outlined in the sections that follow.

2.1 Formulating the problem

The goal of this paper is to synthesize and analyse the recent research in the area of applying AI to ASD, which is a specific software development paradigm. Namely, ASD is a change-driven methodology for creating software in the context of unpredictable requirements (Perkusich et al., 2020). ASD emphasizes the needs of the customers, collaboration, and fast delivery of software. Due to the presence of the customers, who always provide feedback in accordance with market trends or company values, there is a need for constant decision-making and optimization. Scrum, Extreme Programming, Crystal, and Adaptive Software Development are a few popular agile methodologies.

To clarify the scope of the research, the difference between software process and activities will be briefly explained. A software process (also known as a software methodology) is a set of related activities that leads to the production of the software (Shehab et al., 2020). ASD process consists of three phases, as shown in Figure 1.

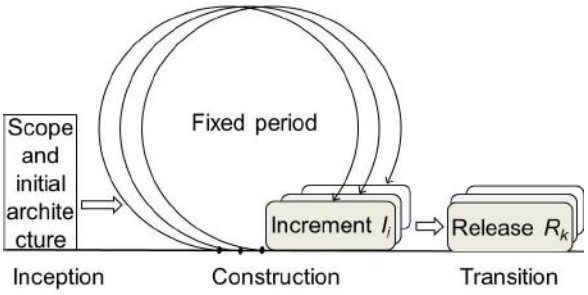


Figure 1. The Agile Software Development Process (ben Othmane et al., 2014)

The project's scope and the initial architecture are determined at the inception phase. The software is developed through a series of iterations during the construction phase. For every iteration, a group of user stories is selected. The requirements are elicited, and the design is updated to address the requirements. Then, a software increment that addresses the requirements is coded. At the end of the iteration, results are presented and the team's efficiency is evaluated. In the transition phase product is tested and prepared as a release for use in production. Multiple activities take place during the ASD process, and each has its own workflow. Some of these activities are task planning, agile estimating (e.g., the time to completion), user stories development, backlog management, agile scheduling, etc.

Today's software is increasingly using artificial intelligence (AI). The ability to quickly, automatically, and effectively make decisions and predictions are considered to be the main benefits of AI techniques and tools, which has led to their adoption in software engineering (SE) (Sofian et al., 2022). In light of the increased interest in AI application in the SE domain, this study seeks to find the answers to the following research questions:

- RQ1: Which AI techniques and tools are used in the context of ASD?
- RQ2: For which purposes are AI techniques and tools used in the context of ASD?
- RQ3: What are the benefits of using AI techniques and tools in the context of ASD?

2.2 Searching the literature

For the selection of primary studies, three databases were used: Web of Science Core Collection (WOSCC), Scopus, and IEEE Xplore (IEEE). The following query was applied to selected databases: *("artificial intelligence" OR "AI")* AND *("agile software development" OR "agile development" OR "agile process" OR "agile processes" OR "agile framework" OR "agile method" OR "agile methodology").* The search period was limited to the last 10 years (2013 – 2023) and to the subject area of Computer Science. Search fields were limited to title, abstract, and keywords. Only articles and conference papers written in English were included in the research.

2.3 Screening for inclusion

Gathering the papers from databases and identifying possibly relevant studies by title and abstract comprised the first two phases of the research selection procedure. Two researchers were included in this phase of the process. One researcher was responsible for screening all papers retrieved during the search phase, while the other author was responsible for screening 15% of the papers from each database. Afterward, the results were compared to see the rate of match between the authors. The following inclusion and exclusion criteria were used:

Inclusion criteria:

- It is clear from the article's title that it is about the use of AI in ASD.
- It seems obvious from the abstract that the article describes and elaborates on the application of AI in ASD.
- It is not clear from the title or abstract to what extent the article covers the application of AI in ASD, but it might be relevant to the research.

Exclusion criteria:

- The title suggests that the article is only marginally related to AI or ASD.
- The abstract suggests that the article is only marginally related to AI or ASD.
- The abstract's focus is not on using AI in ASD, but rather on another topic (i.e., the adoption of agile methodology, development of the ontology, etc.).

After the removal of duplicated papers, the final set of 37 papers was selected for the quality assessment phase, as presented in Table 1.

Table 1. Sources and obtained papers

Source	Initial query	Potentially relevant by title	Potentially relevant by abstract
WOSCC	56	18	10
Scopus	169	27	14
IEEE	149	35	28
Total	374	80	52
Duplicates			15
Final			37

2.4 Assessing quality

Based on the titles and abstracts of 37 papers that appeared to be possibly relevant, the following set of questions for assessing quality was developed:

- Q1: Does the article describe how certain AI technique/tool was used in the context of ASD?
- Q2: Does the article describe how certain activities or processes of ASD can benefit from using AI?
- Q3: Does the article provide an example (i.e., use case) of integrating AI in ASD process?
- Q4: Does the paper only describe the theoretical implications of using AI in ASD?

Two authors were included in this phase of the research. A ratio of 1:2 was used to distribute the papers among them. Negative responses to questions Q1 to Q3, as well as positive response to question Q4, were eliminatory. The quality assessment analysis resulted in 28 papers which were included in the following phase. Afterward, one paper was excluded from further analysis because the full text could not be retrieved.

2.5 Extracting data

In this stage of the research, 27 papers were examined by two researchers. The number of papers per author was distributed equally. Before the start of this phase, a short discussion was made to assure there are no misunderstandings about the data extraction method. The inductive method was used to answer the defined research questions. Neither the AI techniques and tools nor the ASD processes and activities were predetermined; rather, they emerged through the research, as well as the stated benefits of using AI techniques and tools in the context of ASD. Applicable information from each of the primary studies included in this stage of the review was gathered. Spreadsheets and reference management software were utilized to manipulate data during the analysis. The research results are based on the answers given in these papers to the stated research questions.

3 Research results

The last step of the methodology consists of analysis and synthesis of previously extracted information. Our research findings are displayed in accordance with the stated research questions. The systematization of results and references will be presented in Table 2.

3.1. Which AI techniques and tools are used in ASD?

Answers to the first research question are presented in Figure 2. A total of 7 different AI techniques were identified during the research process.

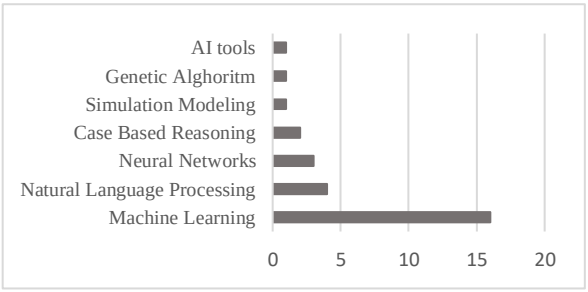


Figure 2. Identified AI techniques and tools used in the context of ASD

The primary AI technique that SE frequently employs is machine learning (ML), which appeared in 16 papers. ML was applied both independently or integrated with other AI techniques to create hybrid techniques. While supervised ML algorithms (e.g., Decision Tree, Linear regression, Logistic Regression, Support Vector Machine, Naïve Bayes, Predictive Analytics, etc.) were mostly used to improve the accuracy of predictions, unsupervised ML algorithms (e.g., K-Means Clustering, KNearest Neighbor, K-medoids, etc.) were used less often, mostly for data segmentation, and usually integrated with supervised ML algorithms. Decision Trees were identified as the most commonly used AI technique (6 papers). Natural Language Processing techniques followed, with 4 papers reporting on this AI technique. This technique was mostly used to automate the derivation of textual information, mainly as a standalone technique. Neural networks were mentioned in 3 papers, Case Based Reasoning in 2 papers, while other AI techniques (Simulation Modeling, Genetic Algorithm) and AI tools (Google Coral devices, TFL mobile SSD object recognition, DLIB correlation tracking) were mentioned in 1 paper.

3.2 For which purposes are AI techniques and tools used in ASD?

As for the second research question, several different purposes of AI techniques and tools were identified, as shown in Figure 3.

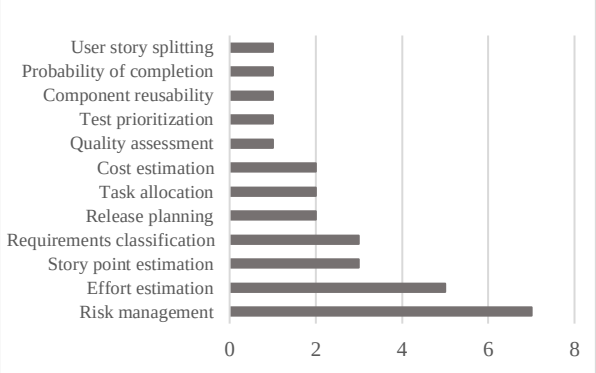


Figure 3. Identified purposes of AI techniques and tools used in the context of ASD

Risk management was identified as the most frequently reported purpose of using AI (7 papers), which is not surprising since the risk of project failure, and time and budget overruns are among the most relevant problems of ASD. Various AI techniques were used to assess the risk, predict failures, prioritize issues, automate the risk management process, and to detect software vulnerabilities. Another almost equally important purpose was effort estimation (5 papers). As the set of requirements for the software to be developed is usually incomplete, effort estimation is uncertain activity, and the use of AI techniques can help improve its accuracy. For similar reasons, AI techniques were used in story point estimation (3 papers). AI techniques were also found to be useful for the purposes of requirements classification (3 papers), release planning (2 papers), task allocation (2 papers), and cost estimation (2 papers). Finally, several different purposes were mentioned only once, including component reusability, quality assessment, test prioritization, probability of completion of features, and user story splitting.

3.3 What are the benefits of using AI techniques and tools in ASD?

The answers to the third research question are shown in Figure 4.

There are 10 groups of identified benefits of using AI in ASD. Improvement of the prediction accuracy (of costs, time, effort, failures, task allocation, etc.) was identified as the most reported benefit of AI techniques and tools in the context of ASD (14 papers), followed by decreased risk (5 papers), and decreased development time (4 papers). The decreased cost was stated as a benefit in 3 papers, followed by improved performance and process support which were stated as benefits in 2 papers. Finally, increased value, improved planning, improved quality, and process automation were reported only once as benefits of using AI in ASD.

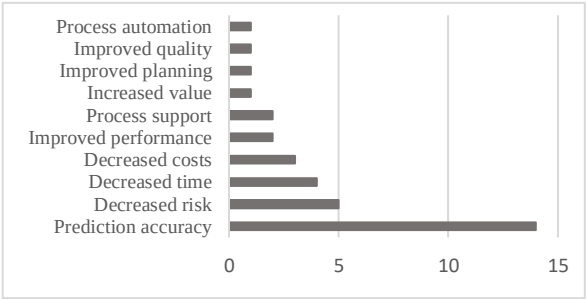


Figure 4. Identified benefits of using AI techniques and tools in the context of ASD

4 Discussion

The value of this paper lies in the identification of the types, purposes, and benefits of AI techniques used in ASD. Contrary to the similar literature reviews which were found during the identification of the need for this research (see Introduction), this paper is the first that tackles agile software development. Since traditional software development (TSD) depends on comprehensive execution of each phase by covering all aspects of the system which is under construction, the choice of AI techniques differs from those that are used to support ASD (Kulkarni & Padmanabham, 2017). As far as we are aware, this literature review is the first that distinguishes agile from traditional software development, as well as the first that deals with AI techniques and tools used within ASD. However, it is important to mention that this paper does not discuss the difference between the AI techniques used in both types of software development, since its primary purpose is narrower and more limited to ASD.

According to the results of the literature review, a variety of AI techniques have been adopted and successfully implemented in ASD. As stated in the previous section, machine learning is without a doubt the most frequently employed AI technique. ML was used for various purposes, such as risk management, effort estimation, story point estimation, requirements classification, task allocation, etc. It was reported that ML has seriously improved prediction accuracy related to cost and effort, as well as decreased risk of failures. On the other hand, other AI techniques were used more rarely, mainly because of their limited scope of impact, but also because of the limited research related to their use within software development. Namely, while ML is well-known and widely recognized, the rise of NLP and neural networks within software development has just recently begun. However, thanks to the promising results of their use in ASD, the expansion of the scope of their use is expected in the next few years.

Table 2. Systematization of results

AI techniques and tools	Purposes	Benefits	References
Machine Learning (Decision Tree, Random Forest and AdaBoost)	Effort estimation, Cost estimation	Prediction accuracy (effort and cost)	(Rodríguez Sánchez et al., 2023)
Machine Learning (Predictive analytics)	Risk management	Prediction accuracy (failures)	(Batarseh & Gonzalez, 2018)
Machine Learning	Quality assessment	Process support, improved quality	(Poth et al., 2019)
Genetic Algorithm	Release planning	Increased business value	(A. Kumar et al., 2014)
Google Coral devices, TFL mobile SSD object recognition, DLIB correlation tracking	Cost estimation	Decreased costs	(Jüngling et al., 2020)
Natural Language Processing (NLP)	Release planning	Improved planning	(Sharma & Kumar, 2019)
Natural Language Processing (NLP)	Requirements classification	Decreased time and risk	(Almanaseer et al., 2022)
Machine Learning (Decision Tree, Stochastic Gradient, Boosting, and Random Forest)	Effort estimation	Prediction accuracy	(Satapathy & Rath, 2017)
Machine Learning (Logistic Regression)	Probability of completion	Prediction accuracy (time)	(Ameta et al., 2022)
Machine Learning	Risk management	Decreased risk, improved process	(Khanna et al., 2021)
Machine Learning (Random Forest, Decision tree, SGB, and Neural Networks)	Effort estimation	Prediction accuracy (effort estimation)	(S. Kumar et al., 2022)
Machine Learning (Support Vector Machine, Logistic Regression)	Task allocation	Prediction accuracy (task allocation)	(William et al., 2021)
Software Process Simulation Modeling (Monte Carlo simulation)	Risk management	Decreased risk, improved process	(Lunesu et al., 2021)
Machine Learning (K-medoids, and K-means Clustering)	User story splitting	Decreased time, improved performance	(B. Kumar et al., 2022)
Machine Learning	Task allocation	Prediction accuracy (task allocation)	(Singh et al., 2021)
Bi-directional Long Short-Term Memory (BiLSTM), Tree-structured Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN)	Story point estimation	Prediction accuracy (story points)	(Marapelli et al., 2020)
Machine Learning (k-Nearest Neighbors algorithm, Linear Regression)	Effort estimation	Prediction accuracy (effort estimation)	(Sanchez et al., 2022)
Case based reasoning	Risk management	Decreased risk	(Jabeen et al., 2014)
Case Based Reasoning	Risk management	Decreased risk, decreased time	(Mukhtar et al., 2013)
Machine Learning (J48 Decision Tree Classifier)	Test prioritization	Prediction accuracy (defects)	(Butgereit, 2019)
Machine Learning (Deep-SE)	Story point estimation	Prediction accuracy (story points)	(Abadeer & Sabetzadeh, 2021)
Graph Neural Network Text Classification	Story point estimation	Prediction accuracy (story points)	(Phan & Jannesari, 2022)
Natural Language Processing (NLP)	Risk management	Decreased risk	(Shafiq et al., 2021)
Machine Learning	Component reusability	Prediction accuracy (reusability)	(Khan & Lingala, 2022)
Machine Learning (Text Classification Algorithm)	Requirements classification	Improved performance	(Magalhaes et al., 2021)
Natural Language Processing (NLP) and Machine learning	Requirements classification	Process automation	(Lano et al., 2021)
Machine Learning (Naïve Bayes, J48 Decision Tree, Random Forest, Logistic Model Tree)	Effort estimation	Prediction accuracy (effort estimation)	(Moharreri et al., 2016)
Graph Method, Deep Neural Network	Risk management	Process support, decreasing risk	(Srivastava et al., 2022)

During the research process, some major concerns were detected. First of all, only one paper has reported the use of AI tools (Google Coral devices, TensorFlow Lite (TFL) mobile Single Shot Detection (SSD) object recognition, and Dlib correlation tracking) (Jüngling et al., 2020). All other papers were experimenting with the use of different AI techniques on specific activities within the ASD, but without providing concrete tools.

To benefit from this kind of research, businesses should first implement AI techniques according to the stated guidelines, and then monitor and measure the progress of implementation. Since this task is time and resource-consuming, some businesses might be discouraged to get it done. To overcome this obstacle, concrete AI tools should be provided in future research. Secondly, most of the time, the purpose of the AI technique was isolated, i.e., the focus and application of a specific AI technique were very narrow. Out of 28 research papers that were included in the phase of data extraction, 26 of them were using one or two types of AI techniques (see Table 2) for specific activities of the ASD process, e.g., effort estimation, task allocation, prediction accuracy, test prioritization, etc. Only two papers were dealing with the processes of software development and the impact of AI techniques on the ASD process. One of them (Poth et al., 2019) has determined the improvement of the quality of the test process by using ML as a support technique, while the other (Jüngling et al., 2020) has shown how AI tools can be used in the system development lifecycle to save costs. Finally, while the coverage of some activities is satisfactory, other activities (e.g., backlog management, user story development, etc.) are completely neglected. As a consequence, the impact of AI on the whole agile software development process is still unknown. To help overcome detected concerns, our future work should focus on the development of AI supported framework for agile software development. In contrast to the stated benefits of previous research, which were related to specific individual activities of ASD, the purpose of AI supported framework would be to improve the quality of the agile software development process.

5 Conclusion

This paper provided a preliminary overview of AI techniques and tools used in agile software development. It also described the purposes for which AI techniques are used in the context of ASD, as well as the benefits they provide for businesses and users. The results of this paper will be mostly valuable for researchers, as a starting point to further investigate AI techniques and tools used in ASD, but also for detecting gaps and promising areas of future research.

The paper's findings demonstrate the diversity of the literature on this topic, describing several AI methods used in ASD. The majority of the papers that were analysed focused on the use of ML techniques within specific activities of the software development process. Also, a great share of papers was dealing with the small set of activities of the software development process (e.g., risk management, agile estimating, task planning, etc.). It is therefore assumed that many more activities of ASD could greatly benefit from the adoption of AI techniques.

During the literature review, a lack of AI tools and a lack of papers dealing with the impact of AI techniques on the ASD process were detected. To help overcome the detected concerns, the authors strive to design and develop AI supported framework for agile software development. Namely, ASD is more than a collection of activities and techniques. It consists of processes, which encompass different events, roles, and artifacts.

Such framework could help the software development teams to maximize the potential of the agile development by maximizing the strengths and opportunities in the team and given context as well as helping to reduce the weaknesses and threats during the whole process. The journey in design and development of such framework is ahead of us, and to determine the overall impact of AI on ASD, the impact of AI on different components of the process should also be determined.

As we have determined that specific AI tools that can be practically applied are lacking, future research could focus on developing concrete AI tools and examining their impact on the entire ASD process. The integration of these tools into the mentioned framework would also be possible and highly desirable.

References

Abadeer, M., & Sabetzadeh, M. (2021). Machine Learning-based Estimation of Story Points in Agile Development: Industrial Experience and Lessons Learned. *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, 106–115. https://doi.org/10.1109/REW53955.2021.00022

Almanaseer, A. M., Alzyadat, W., Muhairat, M., Al-Showarah, S., & Alhroob, A. (2022). A proposed model for eliminating nonfunctional requirements in Agile Methods using natural language processes. *2022 International Conference on Emerging Trends in Computing and Engineering Applications (ETCEA)*, 1–7. https://doi.org/10.1109/ETCEA57049.2022.10009796

Ameta, U., Patel, M., & Sharma, A. K. (2022). Scaled Agile Framework Implementation in Organizations', its Shortcomings and an AI Based Solution to Track Team's Performance. *2022 IEEE 3rd Global Conference for Advancement in*

Technology (GCAT), 1–7. https://doi.org/10.1109/GCAT55367.2022.9971968

Batarseh, F. A., & Gonzalez, A. J. (2018). Predicting failures in agile software development through data analytics. *Software Quality Journal*, *26*(1), 49–66. https://doi.org/10.1007/s11219-015-9285-3

ben Othmane, L., Angin, P., Weffers, H., & Bhargava, B. (2014). Extending the agile development process to develop acceptably secure software. *IEEE Transactions on Dependable and Secure Computing*, *11*(6), 497–509.

Biesialska, K., Franch, X., & Muntés-Mulero, V. (2021). Big Data analytics in Agile software development: A systematic mapping study. *Information and Software Technology*, *132*, 106448.

Butgereit, L. (2019). Using Machine Learning to Prioritize Automated Testing in an Agile Environment. *2019 Conference on Information Communications Technology and Society (ICTAS)*, 1–6. https://doi.org/10.1109/ICTAS.2019.8703639

Jabeen, J., Motla, Y. H., Abbasi, M. A., Batool, D.-B., Butt, R., Nazir, S., & Anwer, S. A. (2014). Incorporating artificial intelligence technique into DSDM. *Asia-Pacific World Congress on Computer Science and Engineering*, 1–8. https://doi.org/10.1109/APWCCSE.2014.7053838

Jüngling, S., Peraic, M., & Martin, A. (2020). *Towards AI-based Solutions in the System Development Lifecycle*.

Khan, F., & Lingala, G. (2022). Machine Learning Techniques For Software Component Reusability. *2022 3rd International Conference for Emerging Technology (INCET)*, 1–6.https://doi.org/10.1109/INCET54531.2022.9824063

Khanna, E., Popli, R., & Chauhan, N. (2021). Artificial Intelligence based Risk Management Framework for Distributed Agile Software Development. *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*, 657–660. https://doi.org/10.1109/SPIN52536.2021.9566000

Kulkarni, R. H., & Padmanabham, P. (2017). Integration of artificial intelligence activities in software development processes and measuring effectiveness of integration. *IET Software*, *11*(1), 18–26. https://doi.org/10.1049/iet-sen.2016.0095

Kumar, A., Nagar, R., & Baghel, A. S. (2014). A genetic algorithm approach to release planning in agile environment. *2014 International Conference on Information Systems and Computer Networks (ISCON)*, 118–122. https://doi.org/10.1109/ICISCON.2014.6965230

Kumar, B., Tiwari, U., & Dobhal, D. C. (2022). User Story Splitting in Agile Software Development using Machine Learning Approach. *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 167–171. https://doi.org/10.1109/PDGC56933.2022.10053226

Kumar, S., Arora, M., Sakshi, & Chopra, S. (2022). A Review of Effort Estimation in Agile Software Development using Machine Learning Techniques. *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 416–422. https://doi.org/10.1109/ICIRCA54612.2022.9985542

Lano, K., Yassipour-Tehrani, S., & Umar, M. A. (2021). Automated Requirements Formalisation for Agile MDE. *2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*, 173–180. https://doi.org/10.1109/MODELS-C53483.2021.00030

Lunesu, M. I., Tonelli, R., Marchesi, L., & Marchesi, M. (2021). Assessing the Risk of Software Development in Agile Methodologies Using Simulation. *IEEE Access*, *9*, 134240–134258. https://doi.org/10.1109/ACCESS.2021.3115941

Magalhaes, C., Araujo, J., & Sardinha, A. (2021). MARE: An Active Learning Approach for Requirements Classification. *2021 IEEE 29th International Requirements Engineering Conference (RE)*, 516–521. https://doi.org/10.1109/RE51729.2021.9714537

Marapelli, B., Carie, A., & Islam, S. M. N. (2020). RNN-CNN MODEL:A Bi-directional Long Short-Term Memory Deep Learning Network For Story Point Estimation. *2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA)*, 1–7. https://doi.org/10.1109/CITISIA50690.2020.9371770

Moharreri, K., Sapre, A. V., Ramanathan, J., & Ramnath, R. (2016). Cost-Effective Supervised Learning Models for Software Effort Estimation in Agile Environments. *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, 135–140. https://doi.org/10.1109/COMPSAC.2016.85

Mukhtar, M., Motla, Y. H., Riaz, M., Khan, M. A., Ahmed, M., Abbas, M. A., Naz, H., & Batool, A. (2013). A hybrid model for agile practices using case based reasoning. *2013 IEEE 4th International Conference on Software Engineering and Service Science*, 820–823. https://doi.org/10.1109/ICSESS.2013.6615431

Perkusich, M., Chaves E Silva, L., Costa, A., Ramos, F., Saraiva, R., Freire, A., Dilorenzo, E., Dantas, E., Santos, D., Gorgônio, K., Almeida, H., &

Perkusich, A. (2020). Intelligent software engineering in the context of agile software development: A systematic literature review. *Information and Software Technology*, *119*, 106241. https://doi.org/10.1016/j.infsof.2019.106241

Phan, H., & Jannesari, A. (2022). Story point level classification by text level graph neural network. *Proceedings of the 1st International Workshop on Natural Language-Based Software Engineering*, 75–78. https://doi.org/10.1145/3528588.3528654

Poth, A., Beck, Q., & Riel, A. (2019). Artificial Intelligence Helps Making Quality Assurance Processes Leaner. In A. Walker, R. V. O'Connor, & R. Messnarz (Eds.), *Systems, Software and Services Process Improvement* (Vol. 1060, pp. 722–730). Springer International Publishing. https://doi.org/10.1007/978-3-030-28005-5_56

Rodríguez Sánchez, E., Vázquez Santacruz, E. F., & Cervantes Maceda, H. (2023). Effort and Cost Estimation Using Decision Tree Techniques and Story Points in Agile Software Development. *Mathematics*, *11*(6), 1477.

Sanchez, E. R., Maceda, H. C., & Santacruz, E. V. (2022). Software Effort Estimation for Agile Software Development Using a Strategy Based on k-Nearest Neighbors Algorithm. *2022 IEEE Mexican International Conference on Computer Science (ENC)*, 1–6. https://doi.org/10.1109/ENC56672.2022.9882947

Satapathy, S. M., & Rath, S. K. (2017). Empirical assessment of machine learning models for agile software development effort estimation using story points. *Innovations in Systems and Software Engineering*, *13*(2–3), 191–200. https://doi.org/10.1007/s11334-017-0288-z

Shafiq, S., Mashkoor, A., Mayr-Dorn, C., & Egyed, A. (2021). NLP4IP: Natural Language Processing-based Recommendation Approach for Issues Prioritization. *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 99–108. https://doi.org/10.1109/SEAA53835.2021.00022

Sharma, S., & Kumar, D. (2019). Agile Release Planning Using Natural Language Processing Algorithm. *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 934–938. https://doi.org/10.1109/AICAI.2019.8701252

Shehab, M., Abualigah, L., Jarrah, M. I., Alomari, O. A., & Daoud, M. Sh. (2020). (AIAM2019) Artificial Intelligence in Software Engineering and inverse: Review. *International Journal of Computer Integrated Manufacturing*, *33*(10–11), 1129–1144. https://doi.org/10.1080/0951192X.2020.1780320

Singh, M., Chauhan, N., & Popli, R. (2021). Framework for Machine Learning Based Task Allocation in DASD Environment. *2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, 29–34. https://doi.org/10.1109/CCICT53244.2021.00017

Sofian, H., Yunus, N. A. M., & Ahmad, R. (2022). Systematic Mapping: Artificial Intelligence Techniques in Software Engineering. *IEEE Access*, *10*, 51021–51040. https://doi.org/10.1109/ACCESS.2022.3174115

Srivastava, D. K., Makhija, R., & Batta, A. (2022). Software Vulnerabilities Detection in Agile Process using graph method and Deep Neural Network. *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 1–7. https://doi.org/10.1109/ASSIC55218.2022.10088314

Templier, M., & Paré, G. (2015). A Framework for Guiding and Evaluating Literature Reviews. *Communications of the Association for Information Systems*, *37*. https://doi.org/10.17705/1CAIS.03706

William, P., Kumar, P., Chhabra, G. S., & Vengatesan, K. (2021). Task Allocation in Distributed Agile Software Development using Machine Learning Approach. *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, 168–172. https://doi.org/10.1109/CENTCON52345.2021.9688114

POEMA: a Personal Cloud for Inclusive Education

Aránzazu San Juan Llano
Orange Innovation
44 avenue de la République
92320 Châtillon, France
LIP6 - Sorbonne Université, CNRS
4 place Jussieu, 75005 Paris, France
aranzazu.sanjuanllano@orange.com

Thi-Mai-Trang Nguyen
L2TI - Université Sorbonne Paris Nord
99 avenue Jean-Baptiste Clément
93430 Villetaneuse, France
LIP6 - Sorbonne Université, CNRS
4 place Jussieu, 75005 Paris, France
thi-mai-trang.nguyen@lip6.fr

Daniel Rodríguez
Universidad de Alcalá
Escuela Politécnica Superior, Campus Universitario
28805 Alcalá de Henares, Madrid, Spain
daniel.rodriguezg@uah.es

Matthieu Chardy
Orange Innovation
44 avenue de la République
92320 Châtillon, France
matthieu.chardy@orange.com

Abstract. *The COVID-19 pandemic has evolved the way that education takes place. Distant or hybrid learning has confirmed the importance of cloud computing and network infrastructure for maintaining education activities in this situation. However, lessons learned from these experiences also show the problems of education exclusion and digital infrastructure limitations. In this paper, we analyze these two problems and propose a solution named POEMA, a Personal Cloud for inclusive education based on the cloud continuum concepts and beyond with explicit consideration of inclusiveness workloads by design. Likewise we propose an Inclusiveness Education Key Value Indicator (KVI) definition compatible with the 6G Key Value Indicator concept of Hexa-X.*

Keywords. cloud continuum, cloud computing, edge computing, inclusive education, 5G/6G, personal cloud, key value indicator, inclusiveness KVI

1 Introduction

COVID-19 pandemic outbreaks accelerated the digital revolution. Telecommunications operators' role has been shown to be essential for the continuity of educational and economic activities[1]. Most worldwide governments have closed schools during the COVID-19 pandemic (*Education: An Unprecedented Crisis | UNESCO* 2022). Responses to the outbreaks in the education domain have been heterogeneous both on solutions to ensure continuity of learning and on the modularity of applied restrictions and the adopted systems, either for full-time distance learning or to ensure continuity of education with a blended learning mode, mixing distance learning with degraded face-to-face learning (Huang, R.H., Liu, D.J., Tlili, A., Yang, J.F., Wang, H.H., et al., 2020).

As Custers (Custers, 2022) proposes, it is necessary to consider the new fundamental rights which would have to create in the digital age and reflect upon the rights that would have to be extended. In particular, he suggests introducing the right to Internet access, which would be strongly correlated to the evolution from the right to education[2] through digital education right, focused on reducing the digital divide and digital illiteracy.

Our research should be seen from this perspective of evolution from the right to education through the right to digital education centered on the impacts of telecommunication infrastructures on these rights, especially on digital inclusion. We would analyze its impact to model the open and flexible Information and Communication Technology (ICT) infrastructures that will sculpt the digital twins of schools of tomorrow in order to propose an open, inclusive, intelligent, and sustainable solution tailored to the education domain. We aim mainly at the context that requires substantially modifying the practice of learning, traditionally taught in a face-to-face manner and a dedicated building like the COVID-19 pandemic context.

Regarding the worldwide educational responses to the COVID-19 pandemic, experiences on teaching continuity based on distance learning reported from several countries highlight the importance of ICT infrastructures as a guarantee of reliable networks to simul-

[1] *The impact of COVID-19 on the Global Telecommunications Industry*. (2020). https://www.ifc.org/wps/wcm/connect/Industry_EXT_Content/IFC_External_Corporate_Site/Infrastructure/Resources/covid-19+impact+on+the+global+telecommunications+industry.

[2] *Le droit de l'enfant à l'éducation; 1979 - 188684freo.pdf*. (1979). https://unesdoc.unesco.org/ark:/48223/pf0000188684.

taneously provide millions of users with smooth online learning, as well as the necessity of collaboration among different sectors as an enabler of inclusive access to education in emergency contexts (Huang, R.H., Liu, D.J., Tlili, A., Yang, J.F., Wang, H.H., et al., 2020). Likewise, unequal access to this infrastructure at home was a major challenge for most education systems, amplified by the electricity concerns in lower-income territories. In order to ensure access to education, many online learning solutions were adopted. Globally, a national online learning platform was used by 75% education systems. This percentage decreases through 62% in lower-middle income territories (*National education responses to COVID-19: summary report of UNESCO's online survey - UNESCO Bibliothèque Numérique* 2020).

We were wondering what is the impact of network infrastructure currently used in education on learning and inclusive education and how to build a cloud continuum infrastructure at the lowest cost for learning and inclusive education depending on the context (out-of-coverage areas, rural areas, border areas, COVID-19 contexts, and post-COVID-19 contexts). In the same way, we also wonder how to use data mining techniques in a cloud continuum infrastructure as an education-inclusive vector applied to existing use cases and the new use cases that will emerge with 5G/6G technology.

To address the education infrastructure concerns, we propose an open cloud continuum infrastructure for inclusive education based on the new concept of the personal cloud, Personal Open Education Mobile Appliance (POEMA), extended beyond the edge of the network through the user devices (laptop, tablet, mobile phone, or IoT devices). In addition, following the 6G Key Value Indicator definition (KVI) as defined by the 6G European initiative, Hexa-X (*Hexa-X H2020 5G-PPP project. Deliverable D1.2, Expanded 6G vision, use cases, and societal values* 2021), we propose a new Inclusiveness Education KVI.

The remainder of this paper is organized as follows. Section 2 introduces the cloud continuum infrastructures background. Section 3 provides a overview of related work. Section 4 presents the POEMA architecture. Finally, we conclude and present the future directions in Section 5.

2 Cloud Continuum Infrastructures Background

Since the 60s, the evolution of computing paradigms has been characterized by alternating between paradigm waves of centralized computation and paradigm waves of distributed computation. In the 2000s, a paradigm of centralized computation named cloud computing emerged; this paradigm has been followed by its distributed alternative, the edge computing paradigm, to converge below in the cloud continuum or edge continuum (Satyanarayanan, 2017; *State of the Edge Report 2021 - State of the Edge* 2021). We will address the paradigms conforming Everything-as-a-service (XaaS) continuum, meaning the different services proposed to manage computing, storage, or networks, and what are the advantages and the limitations that present depending on their context. Finally, we will present a thriving approach, WebAssembly on the server side, which we find extremely promising for the new challenges of the cloud continuum and potentially relevant for the new challenges of the education vertical.

The concept of cloud has been widely diffused. The amateurs and even the users without any computer skills have embraced it through the confluence of smartphone popularization and the rise of free online storage offers to data synchronizing in an ultra-connected and multi-device society. From a formal point of view, cloud computing consists of sharing a cloud provider's physical and virtual resources among multiple customers based on a multi-tenant architecture scalable on demand. These resources may be, for instance, computing, storage, or bandwidth (Bernstein, 2014). The fundamental characteristics of cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service (*Cloud computing | CNIL* n.d.; Mell and Grance, 2011). All these features generate the illusion of unlimited capacity for the customers.

Cloud computing is the technical response to a globalized society that pushes us to consume more remotely than nearby. On the one hand, this consumption of computing resources all over the planet, all these data transiting by inter-continental networks, raises many societal questions about privacy, sovereignty, sustainability, and digital inclusion; on the other hand, cloud computing does not allow to address some use cases like those in the Internet of Things (IoT) field (Khan et al., 2019). Regarding the education vertical, the COVID-19 pandemic fostered the emergence of new online educational models, revealing the limits of cloud computing to respond to a large number of simultaneous requests with reasonable latency times. For these reasons, we find it relevant to contextualize our research on the continuum of the cloud, edge, and beyond (Duan, 2021).

2.1 The Cloud Computing paradigms

There are many stakeholders in the industry gathered around cloud continuum challenges, of which Cloud Native Computing Foundation (CNCF) and Linux Foundation Edge (LF Edge) (*CNCF* n.d.; *LF Edge* n.d.). Both of them are hosted by the Linux Foundation (LF) (*Linux Foundation - Decentralized innovation, built with trust* n.d.).

The leading actor in the cloud continuum is the

CNCF, whose principal objective is to make cloud-native computing ubiquitous. Cloud-native technologies like containers or mesh services foster the observability and resilience of the system it implements.

The most popular CNCF project is Kubernetes (*Kubernetes* n.d.), an open-source system for automating deployment, scaling, and managing containerized applications. We are interested in lightweight Kubernetes distributions like K3S or MicroK8S, which would be more relevant for the network edge (*K3s* n.d.; *MicroK8s - Zero-ops Kubernetes for developers, edge and IoT | MicroK8s* n.d.; *Small Kubernetes for your local experiments: k0s, MicroK8s, kind, k3s, and Minikube – Flant blog* n.d.).

Regarding the concept of edge computing, many definitions and nomenclatures exist in the literature (edge, fog, cloudlet) (Ceselli et al., 2017). This paradigm addresses workloads with reduced latency, mobility scenarios, and critical-time response applications (Khan et al., 2019). LF Edge proposes a formal taxonomy, which is being adopted by the industry. Via this taxonomy, edge computing is presented as a continuum that extends from central data centers hosting cloud servers through edge devices located on the user side while traversing edge servers located on the telecommunications operator side.

2.2 The Serverless/FaaS paradigm

Cloud providers offer customers different models of services to manage the infrastructure where their applications will be deployed. In recent years, the dominant cloud computing service has been Infrastructure-as-a-Service (IaaS) (Long et al., 2021). This service is used, for example, by the public clouds of Web giants such as Amazon Web Services (AWS), Google App Engine (GCE) or Microsoft Azure. On the Open Source side, OpenStack (*Open Source Cloud Computing Infrastructure* n.d.) is the most famous platform based on an IaaS model; it allows the creation of public and private clouds. The explosive extension of Kubernetes has oriented the market from the Insfrastructure-as-a-Service (IaaS), the dominant model last years, to the Container-as-a-Service (CaaS), both of which present limitations regarding edge computing workloads. The serverless or Function-as-a-Service(FaaS) approach allows for addressing this point. Existing serverless/FaaS solutions, some are based on clasic virtual machines (VMs); others on the micro-virtual machines (microVMs), like the AWS Lambda Functions; other on Docker containers like the IBM Cloud Functions; or still on a compromise between several approaches of VMs and containers depending on the context, such as Microsoft Azure Functions or Google Cloud Functions. However, existing FaaS solutions based on containers or Virtual Machine (VM) approaches do not take advantage of the benefits of this model (Gadepalli et al., 2020).

The new FaaS stacks would be designed to avoid these limits improving performance and scalability, the startup time, and the footprint for every function call. Additionally, it could support a more accurate billing model (Long et al., 2021).

2.3 WebAssembly on the server side

In order to lighten the execution environment to run the functions characterizing a FaaS platform, the literature proposes fickler virtualization technologies. Among these technologies, WebAssembly (Wasm) traverses the mirror of the Web through the server side, offering remarkable advantages for web applications (De la Rocha, 2020). Solomon Hykes, Docker's founder, says that if WebAssembly had existed in 2008, Docker would not have been created[3].

The WebAssembly and WebAssembly System Interface (WASI) (*WebAssembly* n.d.) standards have been developed by the Bytecode Alliance (*Bytecode Alliance* n.d.). The objective is to develop new runtimes and language toolchains that provide security, efficiency, modularity, and portability based on a fine-grained sandboxing approach, improving the software composability and cross-device feature in the future. It transposes hence ubiquity in interoperability of the Web through the pervasive computing world. With near-native performance, the literature shows that Wasm improves the cold start time compared to other runtimes like Docker on different public cloud platforms. However, despite all hope for Wasm, it presents several limits. In effect, it is necessary to consolidate features such as the network and the multithreading management tools. Otherwise, to fully benefit from the LLVM chain, the support for programming languages must be completed (Gackstatter et al., 2022; Long et al., 2021).

2.4 The 6G Key Value Indicators

The fundamental principles of 6G, still in an inchoative state, shift the center of gravity of cellular networks towards a dual vision composed of technology and society. Hexa-X, the most significant 6G European initiative, means digital inclusion as three 6G key value with sustainability and reliability. Its vision of 6G includes the concept of the Key Value Indicator (KVI), which aggreges evolution of the classic notion of Performance Indicator (KPI) like the connection density with the new indicators as the integrated intelligence or the artificial intelligence/machine learning (AI/ML) for the optimization of network operations. Regarding the KVI, Hexa-X underlines the difficulty of measuring and quantifying societal values and the perceived performance value. To quantify the KVI is suggested to use softer indicators. It would be linked, for example,

[3] *Solomon Hykes sur Twitter* [Twitter]. (2019). https://twitter.com/solomonstre/status/1111004913222324225.

to a specific use case or an emergent trend that would allow affecting a relative weight to a given KPI, and their values would be considered from a global point of view. Regarding the perceived performance and standardizing these subjective perceptions set modeled by the personal or collective environments of each individual, the most suitable is to be based on the Sustainable Development Goals (SDGs) of the United Nations Organisation (UNO)[4], that is the most open, and inclusive toolset that exists. For education uses cases, SDG 4, piloted by UNESCO[5], is focused on ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all.

3 Related work

Concerning research in cloud continuum infrastructures for education, we made an exploratory phase to provide an overview of works studying the transition from the centralized cloud computing model to the edge computing model and their integration with artificial intelligence technologies (AI), especially the techniques emerging from the machine learning (ML) subfield. AI combined with other technologies like blockchain or genetic algorithms is very promising for research in education infrastructures.

Several studies present a cloud continuum infrastructure for education based on the edge computing paradigm, including an Internet of Things (IoT) layer. As described in the Introduction section, the centralized cloud computing approach presents some limitations concerning IoT devices, especially for latency. In (Siddiqui et al., 2023), the authors highlight the need for cloud-edge infrastructures to use industry 4.0 technologies to improve education infrastructures. They propose an education framework based on an IoT edge infrastructure, describe some limitations of these approaches, and make some proposals to address them. These limitations are the security concerns like authentication, manual configurations, and privacy protection; the network management across heterogeneous devices; the time for data aggregation; the lack of expertise and adequate infrastructure of educational institutions.

Security concerns could be addressed with blockchain technology, an up-and-coming technology for the development of smart campuses which brings cyber-security and redundancy to the processes and data managed (Fernández-Caramés and Fraga-Lamas, 2019).

Educational data or infrastructure data from the IoT layer were exploited with AI/ML techniques. In an experience carried out in Pakistan, (Razzaq et al., 2023) a cost-effective infrastructure based on Convolutional Neural Networks (CNN) was implemented to build a digital twin framework for performance monitoring of teachers and attendance monitoring of students and teachers without human intervention. Moreover, the authors consider to include blockchain technology in their future works to provide a more secure framework.

In (Guo, 2022), in addition to the IoT layer, some functions of education governance are available for parents and students through mobile internet. A total of 91.3 % of parents gave positive feedback on this intelligent home control system. Techniques based on genetic algorithms were proposed in two of explored works. In (Liu et al., 2022) a central cloud approach was used complemented with an edge node adaptive data type sensing mechanism based on counting bloom filter, optimized using genetic algorithms techniques and implemented applying the blockchain agricultural products trading process. Additionally, in (Duan, 2021) an offloading mechanism also based on genetic algorithms was presented in order to minimize the computing delay of terminal tasks of an edge infrastructure for preschool education.

To the best of our knowledge, POEMA is the first proposition in terms of personal cloud for education. There exist other solutions for edge cloud such as the mesh router of Green Communications (*Green Communications – Internet-of-Edges (IoE)* n.d.). Applications such as education applications can be embeded into these mesh routers. However, the main objective of mesh routers is to provide a network infrastructure with distributed embedded applications. They don't have the cloud continuum feature based on serverless technology and a specific objective of personal cloud dedicated to inclusive education as our proposed solution.

4 The POEMA Architecture

4.1 Our Personal Cloud concept: POEMA

The United Nations Educational, Scientific and Cultural Organization (UNESCO) and the European Agency for Special Needs and Inclusive Education (EASNIE) have formulated several recommendations. In 2019, UNESCO adopted the first international recommendation regarding the field of openly licensed educational materials and technologies in education (*Ressources éducatives libres (REL) accessibles: document d'information - UNESCO Bibliothèque Numérique* n.d.) and defines the concept of Open Educational Resources (OER) [6]. It addresses all formal and informal education sector stakeholders, including

[4] *THE 17 GOALS | Sustainable Development*. (n.d.). https://sdgs.un.org/goals.

[5] *Leading SDG 4 - Education 2030 | UNESCO*. (n.d.). https://www.unesco.org/en/education2030-sdg4.

[6] This norm defines the concept of Open Educational Resources (OER) as "learning, teaching, and research materials in any format and medium that reside in the public domain or are under the copyright that has been released under an open license, that permits no-cost access, re-use, re-purpose, adaptation and redistribution by others."

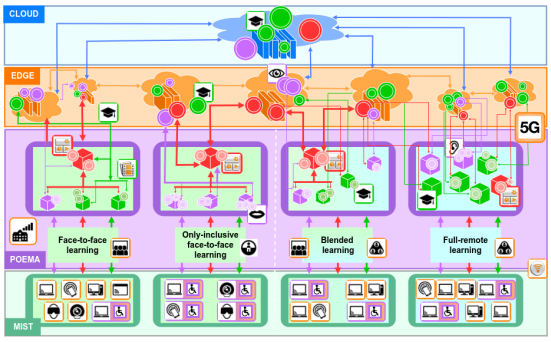


Figure 1: POEMA architecture

ICT infrastructure providers or cloud providers. In 2021, EASNIE proposed several recommendations in the context of the COVID-19 pandemic, ensuring free and inclusive access to digital resources in all remote learning processes.

As discussed previously, the COVID-19 pandemic revealed the limits of infrastructures of education, increasing the digital divide: central cloud model presents limits for the uses cases of massive distance or hybrid learning like the COVID-19 or the future classroom technology with extended reality (xR), namely augmented reality (AR), virtual reality (VR) and mixed reality (MR).[7] Ṡome limits are the data sovereignty, enormous latency, energy consumption, or the commercial factors related to the traditional domain of telecommunications infrastructures based on vendor-locked hardware infrastructures (Akcin, 2023). Edge computing infrastructures partially address these concerns, but the education vertical needs more flexible, mobile, extensible, and resilient infrastructures in order to guarantee the continuity of education in planetary scenarios like the COVID-19 pandemic, natural disasters, or human migrations.

Taking into account the UNESCO and the EASNIE recommendations and in order to try to answer the research questions we described previously, we suggest an extension of the notion of OER, at the moment focused on education contents, through the Open Educational Resources and Inclusive Infrastructures (OERII). It could provide open solutions for the digital inclusion challenges in education, especially for identified contexts with a critical inclusion fracture like the out-of-coverage areas, rural areas, border areas, and COVID-19 contexts. For this purpose, we propose a tiers architecture inclusive by design with the new concept of Personal Cloud named Personal Open Education Mobile Appliance (POEMA), mainly based on the FaaS paradigm and a new Inclusiveness Education KVI. [8].

POEMA is a new tier in the cloud continuum spectrum, which would be located beyond the edge tier, between the user device (tablet, laptop, and IoT devices) and the base station of the telecom provider or the access point. To begin with, we consider a scenario in which the POEMA nodes could connect to the telecom provider edge nodes only with 5G gNodeB.

In Fig. 1, we present a high-level vision of the POEMA architecture that we describe briefly below.

- **Cloud tier:** This tier is represented in a blue color range. It hosts the nodes of the central data centers. This tier addresses the traditional distance or hybrid learning education use cases.

- **Edge tier:** This tier is represented in an orange color range. It hosts the nodes of the edge data centers. This tier addresses sovereignty, latency, and energy consumption issues of traditional distance or hybrid learning education use cases.

- **POEMA tier:** This tier is represented in a purple color range. It hosts the nodes corresponding to POEMA devices. This tier provides flexibility and extensibility required by the inclusion uses cases and guarantees the continuity of education. Its flexibility and extensibility could contribute to enhancing inclusion with more personalized education infrastructures and also to implement easily not traditional education use cases like flipped classrooms, outdoor classrooms, or gamified classrooms.

- **Mist tier:** This tier is represented in a green color range. It hosts the nodes of the user devices.

In Fig. 2, we present a sequence diagram for the default face-to-face scenario.

POEMA workloads

- **POEMA technical functions:** The set of functions implementing technical actions like authentication, education environment build, or orchestration. This set is represented in a red color range.

- **POEMA educational functions *(POEMA OER functions)*:** The set of educational functions which compose the OERs, as an open-source video conferencing application, for instance, a BigBlueButton alternative tool mainly based on a FaaS model. This set is represented in a green color range.

- **POEMA inclusiveness functions:** The set of functions implementing actions to improve the inclusiveness of the POEMA infrastructure. For instance, the functions related to simultaneous translation for face-to-face, blended, or full remote learning activities for refugees of different origins and studying in regular schools. This set is represented in a purple color range.

[7] staff, A. B. (n.d.). *xR, AR, VR, MR: What's the Difference in Reality?* Arm Blueprint. https://www.arm.com/blogs/blueprint/xr-ar-vr-mr-difference.

[8] The word "poema" means poem in Spanish and proceeds from the Latin *poēma*

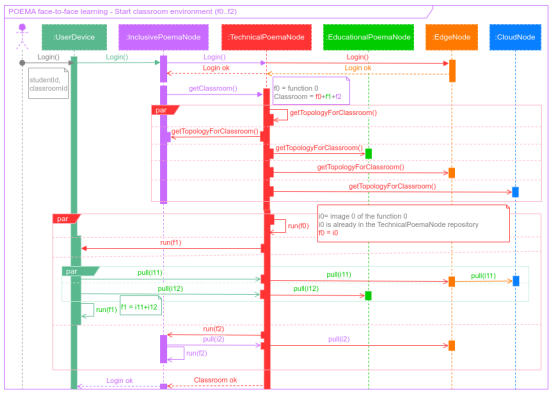


Figure 2: POEMA sequence diagram

4.2 Our new Inclusiveness Education KVI

We propose a new Inclusiveness Education KVI, compatible with the 6G KVI definition proposed by Hexa-X. This soft inclusiveness indicator will be computed with the data measured on specific inclusiveness workloads identified in our system as POEMA inclusiveness functions. Our first use case is the French National Education infrastructure, and we are analyzing public datasets of this use case available on the French National Education open data repository (*Explore — Éducation Nationale - Accueil* n.d.). This repository provides open data sets with general information about the schooled population and digital inclusion statistics like the ETIC survey. The ETIC survey collected yearly data from all public schools in France regarding digital equipment, infrastructures, and services that permit us to compute indirect soft partial inclusiveness indicators between different contexts, such as rural or urban. Additionally, the survey provides inclusion data, for example, if a school has pedagogical material available to students supplied for ASH (Adaptation, Schooling, Handicap) and the number of them. These data permit us to compute directly soft partial to build a heuristic to calculate the initial weight of the inclusiveness workload functions in a given school or region. Thus, we have to define a global inclusiveness KVI composed of partials inclusiveness KVIs as cited before and other 5G/6G KPIs as area traffic capacity or energy efficiency.

5 Conclusions and Future Work

Nowadays, millions of learners study online from their homes. During the COVID-19 pandemic, education was strongly impacted, shifting from full-time to online learning when possible, increasing the digital divide. The success factors of the abrupt commutation from face-to-face education systems to remote education systems on a planetary scale were the individual efforts of digitalizing their infrastructures from some countries or schools before the pandemic, hence the help from leading tech companies. This health crisis has revealed difficulties in accessing the Internet and the lack of digital skills in the educational community. Education ICT infrastructures for remote or blended learning are based on central cloud computing paradigms that present limitations regarding data sovereignty, sustainability, or vendor-locking constraints. The edge continuum infrastructures provide partial responses for these issues. However, the education vertical need more flexible, mobile, extensible, and resilient infrastructures to guarantee the continuity of education in scenarios like the COVID-19 pandemic and implement future education scenarios that will emerge with 5G/6G technologies.

To address inclusion challenges, this paper presents the new concept of POEMA, based on a FaaS model that proposes a new tier for a flexible cloud continuum infrastructure for inclusive education with explicit consideration of inclusiveness workloads by design. This architecture would be an extension of the OER conceptual framework for inclusive education infrastructures (OERII), including a new Inclusiveness Education KVI compatible with the 6G KVI notion described by Hexa-X.

In future work, we plan to implement a simulator based on existing cloud continuum simulators like CloudSim Plus in order to validate our system. We will implement an inclusive algorithm to schedule the educational workloads to optimize the inclusiveness of our OERII architecture, maximizing our Inclusiveness Education KVI. We would like to create many simulations crossing basic education use cases (face-to-face learning, blended and full-remote learning) with several specific contexts such as rural versus urban, pre-COVID-19 versus COVID-19 context, post-COVID-19 versus Metaverse and Omniverse contexts. Finally, we will also contemplate creating end-to-end testbed from the public cloud to user devices implementing POEMA level with low-cost devices to validate partial aspects of this POEMA architecture in an almost-real context; for instance, we may deploy some workloads on Wasm runtimes.

References

Akcin, M. (2023). *Council Post: Edge Of Sovereignty: Navigating Data Security And Compliance In Latin America's Evolving Tech Landscape*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2023/04/26/edge-of-sovereignty-navigating-data-security-and-compliance-in-latin-americas-evolving-tech-landscape/

Bernstein, D. (2014). Containers and Cloud: From LXC to Docker to Kubernetes. *IEEE Cloud Computing*, *1*(3), 81–84. https://doi.org/10.1109/MCC.2014.51

Bytecode alliance [Bytecode alliance]. (n.d.). https://bytecodealliance.org/

Ceselli, A., Premoli, M., & Secci, S. (2017). Mobile edge cloud network design optimization. *IEEE/ACM Transactions on Networking*, *PP*. https://doi.org/10.1109/TNET.2017.2652850

Cloud computing | CNIL. (n.d.). https://www.cnil.fr/fr/definition/cloud-computing

CNCF. (n.d.). https://www.cncf.io/

Custers, B. (2022). New digital rights: Imagining additional fundamental rights for the digi tal era. *Computer Law & Security Review*, *44*, 105636. https://doi.org/10.1016/j.clsr.2021.105636

De la Rocha, A. (2020). *@adlrocha - can WASM become the new docker?* https://adlrocha.substack.com/p/adlrocha-can-wasm-become-the-new

Duan, X. (2021). *Online preschool education optimization based on edge computing in the era of covid-19*. Internet Technology Letters, Vol. 4, Issue 5, Wiley, September/October 2021.

Education: An unprecedented crisis | UNESCO. (2022). https://courier.unesco.org/en/articles/education-unprecedented-crisis

Explore — Éducation Nationale - Accueil. (n.d.). https://data.education.gouv.fr/explore/?sort=modified

Fernández-Caramés, T. M., & Fraga-Lamas, P. (2019). Towards Next Generation Teaching, Learning, and Context-Aware Applications for Higher Education: A Review on Blockchain, IoT, Fog and Edge Computing Enabled Smart Campuses and Universities. *Applied Sciences (2076-3417)*, *9*(21), 4479. https://doi.org/10.3390/app9214479

Gackstatter, P., Frangoudis, P. A., & Dustdar, S. (2022). Pushing serverless to the edge with WebAssembly runtimes. *2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 140–149.

Gadepalli, P. K., McBride, S., Peach, G., Cherkasova, L., & Parmer, G. (2020). *Sledge: A serverless-first, light-weight wasm runtime for the edge*. 21st ACM International Middleware Conference (Middleware), Delft, Netherlands, December, 2020.

Green Communications – Internet-of-Edges (IoE). (n.d.). https://www.green-communications.fr/

Guo, Y. (2022). Application of Internet of Things Technology in Mobile Education of Smart Campus Culture and Etiquette. *Journal of Sensors*, 1–10. https://doi.org/10.1155/2022/6321784

Hexa-x h2020 5g-ppp project. deliverable d1.2, expanded 6g vision, use cases, and societal values. (2021). https://hexa-x.eu/wp-content/uploads/2021/05/Hexa-X_D1.2.pdf

Huang, R.H., Liu, D.J., Tlili, A., Yang, J.F., Wang, H.H., et al. (2020). *Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in COVID-19 Outbreak*. Smart Learning Institute of Beijing Normal University, 2020.

K3s. (n.d.). https://k3s-io.github.io/

Khan, W., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). *Edge computing: A survey*. Future Generation Computer Systems, Elsevier, Vol. 97, 2019.

Kubernetes [Kubernetes]. (n.d.). https://kubernetes.io/

LF edge: Building an open source framework for the edge [LF edge]. (n.d.). https://www.lfedge.org/

Linux foundation - decentralized innovation, built with trust. (n.d.). https://www.linuxfoundation.org

Liu, M., Zhou, R., Dai, J., & Feng, X. (2022). Analysis and Practice of Using Modern Information Technology for Classroom Teaching Mode Reform. *Mobile Information Systems*, 1–8. https://doi.org/10.1155/2022/2565735

Long, J., Tai, H.-Y., Hsieh, S.-T., & Yuan, M. J. (2021). *A lightweight design for serverless function as a service*. IEEE Software, Vol. 38, Issue 1, January/February 2021.

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. https://doi.org/10.6028/NIST.SP.800-145

MicroK8s - zero-ops kubernetes for developers, edge and IoT | MicroK8s [Microk8s.io]. (n.d.). http://microk8s.io

National education responses to COVID-19: Summary report of UNESCO's online survey - UNESCO bibliothèque numérique. (2020). United Nations Educational, Scientific; Cultural Organization, Paris, France, April, 2020.

Open source cloud computing infrastructure [OpenStack]. (n.d.). https://www.openstack.org/

Razzaq, S., Shah, B., Iqbal, F., Ilyas, M., Maqbool, F., & Rocha, A. (2023). DeepClassRooms: A deep learning based digital twin framework for on-campus class rooms. *Neural Computing and Applications*, *35*(11), 8017–8026. https://doi.org/10.1007/s00521-021-06754-5

Ressources éducatives libres (REL) accessibles: Document d'information - UNESCO bibliothèque numérique. (n.d.). https://unesdoc.unesco.org/ark:/48223/pf0000380471_fre

Satyanarayanan, M. (2017). The Emergence of Edge Computing. *Computer*, *50*(1), 30–39. https://doi.org/10.1109/MC.2017.9

Siddiqui, S. T., Khan, M. R., Khan, Z., Rana, N., Khan, H., & Alam, M. I. (2023). Significance of Internet-of-Things Edge and Fog Computing in Education Sector. *2023 International Conference on Smart Computing and Application (ICSCA)*, 1–6. https://doi.org/10.1109/ICSCA57840.2023.10087582

Small kubernetes for your local experiments: K0s, MicroK8s, kind, k3s, and minikube – flant blog. (n.d.). https://blog.flant.com/small-local-kubernetes-comparison/

State of the edge report 2021 - state of the edge. (2021). https://stateoftheedge.com/reports/state-of-the-edge-report-2021/

WebAssembly. (n.d.). https://webassembly.org/

Recommendation and Prediction in a Microservice Web Application for Cyber Ranges

Luis de-Marcos, Daniel Rodríguez, José-María Gutiérrez-Martínez, Adrián Domínguez-Díaz, Sergio Caro-Álvaro

Universidad de Alcalá

Computer Science Department

Alcalá de Henares 28805. Madrid, Spain

{luis.demarcos, daniel.rodriguezg, josem.gutierrez, adrian.dominguez, sergio.caro}@uah.es

Abstract*. The EU research project between industry and academia µDevOps is a collaborative research project formed by an international network of organizations including industry and academia that aims to tackle current challenges of microservice development operations. An important case study considered in this project is the Cyber Ranges application a cyber security training and capability development exercises using microservices for the design, delivery, and management of simulation-based, experiences in cyber security developed by Silensec as one of the partners. This work describes the results of analyzing the scenario usage dataset of the Cyber Ranges training platform. This includes the matrix of starts for scenario/user and the attributes of scenarios. The aims are to produce recommendations of scenarios for users based on previous activity and to predict the success of scenarios as measured by the number of starts.*

Keywords. Recommender system, prediction, classification, microservice, web application, cyberrange.

1 Introduction

In online systems users are often overwhelmed by offer and need help finding what they're looking for, so recommendations are essential, since they lead to happier customers and more sales. Recommender systems are like salesmen who know, based on your history and preferences, what users may like. They reduce transaction costs of finding and selecting items in an online environment. The use of efficient and accurate recommendation techniques is very important for a system that will provide good and useful information to its individual users (Enríquez et al, 2019)

Similarly, online businesses can be overwhelmed by the amount of data produced by users when interacting with their platforms. Making sense of all the data can be difficult in a modern environment, where data is highly distributed, diverse, and dynamic. Predicting sales and classifying users are important to businesses because it helps them make better decisions. Sales forecasting helps in overall business planning, budgeting, and risk management. It allows companies to efficiently allocate resources for future growth and manage its cash flow (Atkins et al, 2023). Classification of users is important because it helps businesses understand their customers better and tailor their products or services to meet their needs. This leads to happier customers and more sales (Mahalingam, 2023).

The aim and objectives of this paper are to produce recommendations for new scenarios that a user can take based on the activity of other users in the *Cyber Ranges* training platform. In this preliminary work, we also try to predict number of starts of a given training as a measure of its success based on its descriptive attributes since it determines the incomes produced by the training.

The rest of the paper is structured as follows. Section 2 covers the background. Section 3 presents the state of the art. Section 4 briefly outlines the approach. Section 5 reports findings. The paper concludes with final remarks and future research.

2 Background

This work is part of the µDevOps project[1], which aims to make software better by encouraging testing throughout the software lifecycle. It is a joint research program with academic and industrial partners to share knowledge. The research program has four goals: learning from context, testing continuously in real settings, assessing and managing risks in µDevOps,

[1] https://microdevops.wordpress.com/

and offering μDevOps development and testing as a service. The program is organized into work packages and tries to solve the problems of a fast-changing environment where development and operations are integrated. And as one of the case studies, we use one partner's application, the Cyber Ranges training platform which employs a microservice architectural design. The aim of this research is to provide recommendations for users and predict the success of trainings.

2.1 Microservices

Microservices - also known as the microservice architecture - is an architectural style that structures an application as a collection of services that are independently deployable, loosely coupled, organized around business capabilities, and owned by a small team. Microservices architectures make applications easier to scale and faster to develop, enabling innovation and accelerating time-to-market for new features (Hasselbring and Steinacker, 2017) [4].

Some examples of web applications that use microservices architecture are:

- Amazon: Initially, Amazon was a monolithic application but when microservice came into existence, Amazon was the first platform to break its application into small components, thereby adapting microservice.
- Netflix uses microservices with APIs.
- Uber: When Uber switched from monolithic nature to a microservice, it allowed to scale its services more efficiently contrasting with the bottleneck produced by the previous monolithic architecture.

An architecture of microservices is a software design approach focusing on application creation, formed by a set of small, independent services that communicate with each other through APIs. In the case of online learning platforms, microservices architectures can be beneficial for the development of these platforms as they allow for greater flexibility and scalability. By using this architecture, developers can create small, independent services that communicate with each other through APIs. This allows each service to be scaled and updated independently. Most online learning platforms use microservice architecture including Coursera, edX, Udacity, Udemy and Skillshare (Dreiko, 2021) [5].

2.2 Recommender Systems

Recommender systems are a type of information filtering system that provide suggestions for items that are most relevant to a user. They can help users to choose from many options, such as products, music, news, or movies. Some examples of recommender systems are present in Amazon, IMDb, Facebook, Netflix, and Spotify.

Recommender systems can use different methods to make suggestions, such as collaborative filtering, content-based filtering, or hybrid methods. Content-based filtering uses the features and attributes of items to find similar items. Hybrid methods combine both approaches to improve the quality and diversity of recommendations (Isinkaye et al., 2015; Aymen and Imène, S, 2022).

There are several examples of recommender systems used in online learning platforms. For instance, many online education platforms have built diverse recommender systems that utilize traditional data mining methods, such as Collaborative Filtering (CF) (Li and Kim, 2021). Some researchers have also proposed a deep learning-based course recommender system (DECOR) which captures high-level user behaviours and course attribute features (Liu et al, 2022).

3 Case Study: The CyberRanges Web Application

Cyber Ranges consist of virtual environments that provide secure and legal environment for cybersecurity education, practice, and training. They are designed to isolate trainees from threats by allowing them to recognize and respond to real-world challenges in a controlled environment. This approach ensures that client infrastructure and data are never at risk due to cybersecurity training (Aries Security, 2020). The main commercial cyber range solutions available in the market for are:

- IBM X-Force Command Cyber Range: This is a cloud-based platform that offers immersive and gamified cyberattack simulations, as well as customized scenarios and training modules. It also provides access to IBM's security experts and tools, such as IBM Security QRadar and IBM Resilient.
- Palo Alto Networks Cyber Range: This is a physical facility that hosts hands-on workshops and exercises for cybersecurity professionals, students, and enthusiasts. It also leverages Palo Alto Networks' products and services, such as Cortex XDR and Prisma Cloud, to demonstrate best practices and techniques for threat prevention and detection.
- Fortinet NSE Cybersecurity Training Institute: This is an online platform that offers self-paced courses and certifications for various levels of cybersecurity skills and roles. It also includes access to Fortinet's cyber range labs, where learners can practice their skills in realistic scenarios using Fortinet's security solutions, such as FortiGate and FortiSandbox.

CyberRanges (see Figure 1) is a microservice web application developed by Silensec that provides cybersecurity training and capability development

exercises using technology and services for the design, delivery, and management of simulation-based, deep-dive experiences. CyberRanges can be used to learn, train, test, measure, and improve the digital dexterity and cyber resilience of professionals, teams, and organizations by using a platform and technology that resembles a real-world scenario. A scenario is a training session.

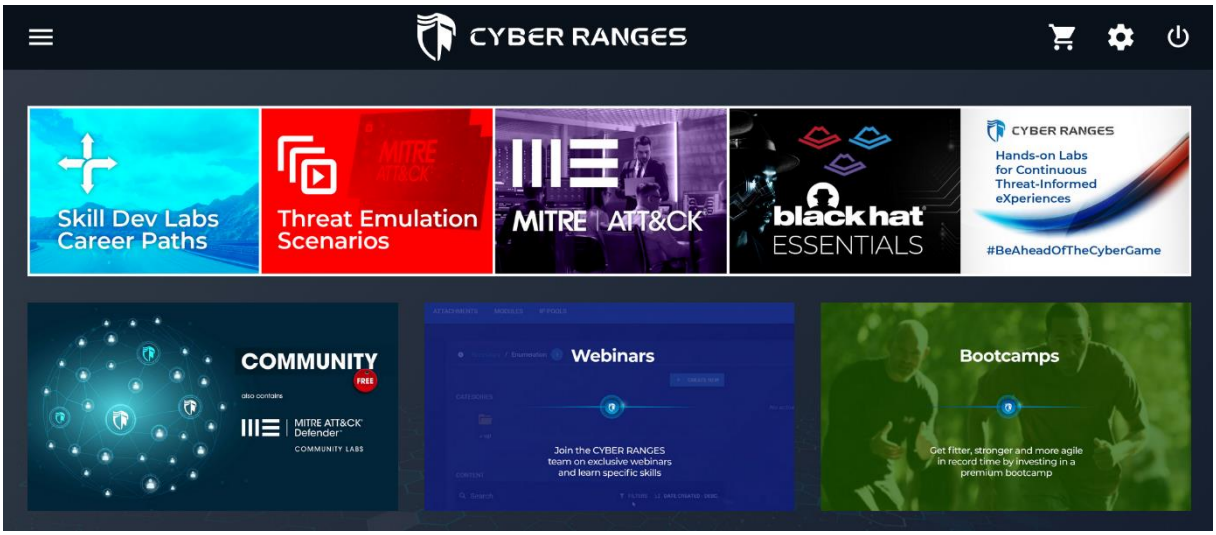


Figure 1. CyberRanges Web Application

As there are several competing platforms offering cyber range solutions. To be competitive it is not only required to provide the better offering in terms of platform functionality, library of scenarios and user base, but it is also necessary include recommender systems that suggests trainings based on to the capabilities of the users. Furthermore, as the incomes of the platform are based on the number of starts of trainings, providers of cyber ranges are usually interested in knowing the features that determine training success from users' perspective. Data analytics can also provide further insights, like the categories that are underrepresented in the library but are particularly successful or appealing to users. All this information can help providers to identify the most meaningful ways to expand their library of scenarios. The present research deals with all these issues by analyzing the data of user activity of the CyberRanges application.

The data used for the recommendation system is composed of two datasets. The first dataset (Dataset 1) is a matrix of IDs of users and IDs of scenarios. Each cell contains a 1 if user started the scenario, 0 it not. There are 11,093 users and 1,534 scenarios.

The second dataset (Dataset 2) is described in Table 1. It comprises multiple attributes and the target attribute (Sum) as the number of times a user starts a scenario (computed for this analysis as an aggregation from Dataset 1)

Table 1. Dataset with scenario information

ID	String
Name	String
Owner	String
Author	String
Status	Categorical. PUBLISHED, DRAFT, PENDING_APPROVAL
Categories	Categorical (multiple values as CSV. E.g. "Web, System Exploitation".
Tags	Categorical (multiple values as CSV describing the scenario). E.g., "SQLi, File Upload, PHP" Arbitrary.
Visible:	Boolean
Type:	Categorical. CCL, CPL, CTF, MEL, TTX
Mode	Categorical. SINGLEPLAYER, MULTIPLAYE
Difficulty	Categorical. EASY, INTERMEDIATE, HARD, IMPOSSIBLE
Technical requirements	Numerical (integer) attributes describing the technical requirements of the scenario: Power, Duration, CPU, RAM (MB), Disk (GB), #Servers, #Containers, #Networks, #Subnets, #Ports, #CloudConfig, #Routers, #Windows, #Linux, #Unknown, #ZoneNova, #ZoneWindows, #Zone BSD, #Zone Passthrough
Sum.	Numerical: Integer. Number of times a user starts a scenario (Computed for this analysis as an aggregation from Dataset 1)

4 Experimental Work

4.1 Data Analysis Methods

In this work, we tackle data analysis through correlations and machine learning. First, we check the correlations of each tuple of scenarios based on the users that started both. From the machine learning point of view, apply: (i) Regression of the number of starts of scenarios (Sum), and (ii) Classification of scenarios (based on the discretization of Sum attribute). Before data analysis, we use descriptive statistics to identify possible anomalies and to describe the main attributes.

The choice of data analysis methods depends on the research questions, the data characteristics, and the available resources. In this project, we decided to use correlations and machine learning as our main tools for exploring the relationships between variables and predicting outcomes. Correlations are useful for measuring the strength and direction of linear associations between two or more variables and can help us identify potential causal factors or confounders. Machine learning is a branch of artificial intelligence that uses algorithms to learn from data and make predictions or classifications. Machine learning can handle complex and nonlinear patterns, deal with missing or noisy data, and discover hidden features or structures in the data. We did not use other methods, such as clustering, or hypothesis testing, because they either did not fit our research goals, required more assumptions or data preprocessing, or were less efficient or accurate than machine learning.

4.2 Exploratory Analysis

Descriptive statistics do not show any significant anomaly. The target attribute Sum (number of starts) seems to follow a power-law, long tail distribution (N=1534, Mean: 25.9. SD: 68.87, SE Mean) as shown in Figure 2, where we can see a few instances with many starts and many instances with a limited number of starts. This makes the analysis more challenging since several methods rely on specific distributions.

It is worth noting that there are 187 instances are missing the values for the following attributes: Power, Duration, CPU, RAM (MB), Disk (GB). Also 215 instances are missing all other numerical attributes (the ones that start with # in Table 1).

Kruskal-Wallis non-parametric tests return statistical differences for all categorical attributes when response is Sum: Status (H=193.4), Visible (H=195.5), Type (H=312.6), Mode (H=37.7) & Difficulty (H=69.8), with $p<0.01$ in all cases. Although in several cases the factors have a reduced number of instances. This is shown in Figure 3. This means that for all this attributes there is a statistical difference for at least one of the groups which can be visually examined in the figure. For instance, when it comes to Difficulty, scenarios rated as 'Impossible' have significantly higher number of starts. Exploratory analysis also shows very high values for several instances of the numerical attributes (particularly Power, Duration, CPU, RAM (MB), Disk (GB)) which may be outliers or bias several results.

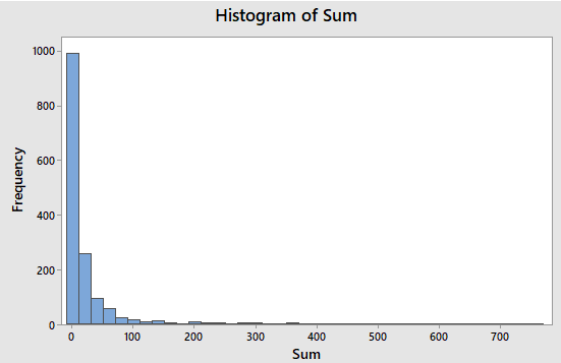


Figure 2. Histogram of the Sum attribute

5 Experimental Results and Discussion

This section provides the results of the experimental work to produce recommendations based on user activities and to predict the number of starts of courses. It also discusses results briefly discussions.

5.1 Correlation-based Recommendations

We computed the correlation between each pair of scenarios based on users' starting them. The matrix was inputted for the correlation as provided (Dataset 1 as described in Section 3). Only valid pairs of correlations were included. These results in a set of 583740 correlations. It takes 3-5 minutes with the dataset provided. Correlations are then filtered by value selecting only those that are larger than 0.6 (moderate correlation) and which have a minimum number of instances in which the correlation is based (initially 10). This rules out highly correlated scenarios which are only based in a very low number of users taking both. Parameters of both filters can later be tuned if needed.

The correlation value can be roughly interpreted as the percentage of number of users who started one of scenarios and the other based on a minimum sample on the dataset. For instance, for a given tuple of courses that were started 238 times we found a correlation value of 0.94 indicating that a significant amount of the users who started the first scenario also started the second. Correlation value is a measure of the strength of the recommendation. Higher means stronger. The minimum number of instances is a measure of number of instances on which the recommendation is based. Higher means better.

5.2 Performance of Regression

To analyze the performance of regression, we firstly performed the following preprocessing:

1. Transpose the matrix (Dataset 1) to compute the Sum of the number of starts for all users.
2. Join Sum to Dataset 2 to use it as the target attribute.
3. Filter out the scenarios with Sum = 0 (never started by any user).
4. Remove all instances with any missing value in any of the numeric attributes.
5. Partition of the dataset for training and testing: 70/30 (draw randomly)

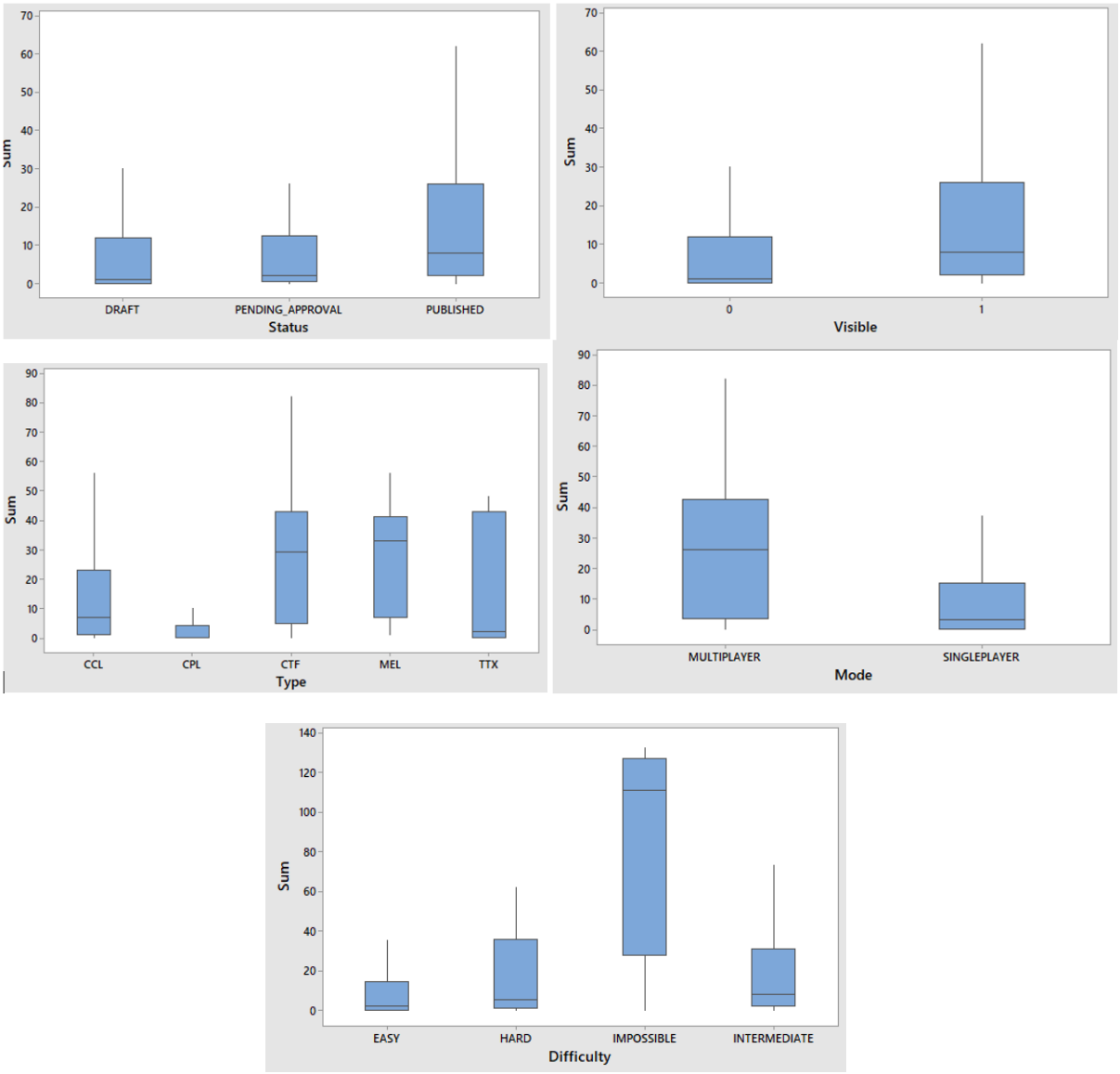


Figure 3. Box plots of the categorical attributes (Status, Visible, Type, Mode & Difficulty)

With the resulting dataset, we trained several machine learning regressors to predict Sum (number of starts of a given scenario) as a function of the rest of the attributes of the scenario. Regressors used and results are shown in Table 2 .

Table 2. Results of the regression techniques

Regressor	**Adjusted R^2**	**Mean absolute error**
Linear (only with numeric attributes)	0.024	72.10
Simple regression tree	-0.193	41.03
Tree ensemble	0.047	43.66
Gradient boosted trees	-0.087	41.32
Random forest	0.056	42.94

As shown on the table, regression results show that the regression models do not fit the data. In the best case, only 5% of the variability is explained by the best model. Negative R^2 in several cases means that the model fits worse than a linear model. Results then suggest that Sum (number of starts) of a scenario cannot be explained as a measure of the other attributes of the scenario given (in the dataset)

5.3 Classification

To analyse the performance of classification, we firstly pre-processed the dataset using the same procedure described in the previous subsection. We also discretized based on the mean of Sum (number of starts) which is around 25. So, all instances with less than 25 starts where classified as 'Low', and all others (Sum >=25) as 'High'. Partition used stratified sampling. We then trained several classifiers of different types to predict the discretized Sum as a function of the rest of the attributes of the scenario. Results are summarized in table 3.

Table 3. Results of the classifiers

Classifier	Accuracy	Precision (High)	Precision (Low)	AUC
KNN	.77	.46	.83	.63
Decision Tree (C4.5)	.81	.57	.86	.70
Gradient Boosted Tree	.80	.59	.83	.76
Random Forest	*.81*	*.61*	*.83*	*.79*
Tree Ensemble	.80	.59	.83	.78
Naïve Bayes	.76	.39	.81	.68
Fuzzy Rule	.76	.36	.84	.61

Analysis from table 3 shows that Random Forest outperforms other classifiers, although only by a small margin when compared to other ensemble methods (Gradient Boosted Tree, Tree Ensemble), and simple classification tree (C4.5). Differences are significant when compared with instance-based methods (KNN), probabilistic (Bayes) and fuzzy rules. The ROC curve of the Random Forest classifier is presented in Figure 4. It plots two parameters: True Positive Rate (TPR) and False Positive Rate (FPR) at different classification thresholds. Classifiers that give curves closer to the top-left corner indicate a better performance. The red line represents a random classifier. The area under the ROC curve (AUC) is a measure of how well a model can distinguish between two classes.

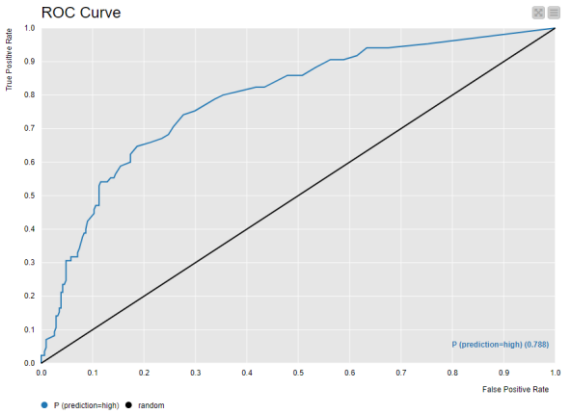


Figure 4. ROC Curve of the best classifier (Random Forest)

Results show that ensemble methods, with Random Forest in particular. produce a moderate successful method to predict the number of starts of a given scenario. In all cases, we can observe a low precision rate for 'High' instances as compares to 'Low'. This is probably due to the imbalance of the resulting discretized dataset.

6 Conclusions and Future Work

Results show that a simple correlation of scenarios from a matrix representing the starts of scenarios and users can be used to produce recommendations of new scenarios to users by applying a set of filters with a minimum threshold for the correlation and a minimum number of instances in which the correlation is based. Its computation, though, can scale poorly as the number of scenarios increases.

Regression results are quite poor. The number of starts of a scenario cannot be explained by other attributes. We argue that labeling of scenarios (attributes Categories and Tags) and the quality of the numerical attributes need to be checked. Further, since the provided attributes are not descriptive for the target metric (number of starts) other attributes, like learning performance, should be considered for future analysis.

Classification models return moderate good results possibly as measured by the metrics used in this study (accuracy, precision, and AUC) because of the unbalance of the dataset (instances with a high number of starts vs low).

As for future work, we suggest the following lines for each analysis of this paper. Concerning correlation, it is necessary to assess whether results are meaningful for the developers and users of the Cyber Ranges application. Next steps also include the integration of correlation results as a recommender system in the learning platform. The correlation process should be integrated in the architecture considering its scalability limitations (limited scalability as a function of the number of scenarios). Concerning regression, we suggest checking the integrity of numeric data and

providing new attributes before any further development. As for classification, we suggest increasing the class underrepresented, which means having more scenarios with a relevant number of starts. Additionally, it is also possible to try other discretization methods.

We also want to mention that although our method and results focus on a microservice web application, it is also possible to apply it to monolithic systems which are scalable vertically and horizontally.

Comparing results with other methods/state of the art is needed to assess the success of our approach. Given the limited number of existing platforms providing cyber ranges training, we did not find any previous studies (research papers or reports).

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 871342

References

Aries Security. (2020). What is a Cyber Range? A Definitive Guide and Definition. https://www.ariessecurity.com/what-is-a-cyber-range-a-definitive-guide-and-definition/ [Accessed on 16th May 2023]

Atkins, C., Uster, M.V.D., Mahdavian, M., and Yee, L. Unlocking the power of data in sales. 2016 [cited 16th May 2023; Available from: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/unlocking-the-power-of-data-in-sales

Aymen, A.T.M., and Imène, S.: 'Scientific Paper Recommender Systems: A Review'. Springer International Publishing, 2022, pp. 896-906

Dreiko, J. How to Plan Software Architecture For eLearning Platforms. 2021 [Last accessed 7th Sept 2023]; Available from: https://polcode.com/resources/blog/how-to-plan-software-architecture-for-e-learning-platforms/

Enríquez, J.G., Morales-Trujillo, L., Calle-Alonso, F., Domínguez-Mayo, F.J., and Lucas-Rodríguez, J.M., 'Recommendation and Classification Systems: A Systematic Mapping Study', Scientific Programming, 2019, pp. 1-18

Hasselbring, W. and Steinacker, G.. Microservice architectures for scalability, agility and reliability in E-commerce. Software Architecture Workshops (ICSAW), 2017 IEEE International Conference on. IEEE, 2017.

Isinkaye, F.O., Folajimi, Y.O., and Ojokoh, B.A.: 'Recommendation systems: Principles, methods and evaluation', Egyptian Informatics Journal, 2015, 16, (3), pp. 261-273

Li, Q., and Kim, J.: 'A Deep Learning-Based Course Recommender System for Sustainable Development in Education', Applied Sciences, 2021, 11, (19), pp. 8993

Liu, T., Wu, Q., Chang, L., and Gu, T.: 'A review of deep learning-based recommender system in e-learning environments', Artificial Intelligence Review, 2022, 55, (8), pp. 5953-5980

Mahalingam, K. The Importance of Sales Forecasting. 2022 15th May 2023]; Available from: https://www.chargebee.com/blog/importance-of-sales-forecasting/.

Semantic methods in our research and teaching

William Steingartner, Valerie Novitzká
Technical University of Košice
Faculty of Electrical Engineering and Informatics
Letná 1/9, 04200 Košice, Slovakia
{william.steingartner, valerie.novitzka}@tuke.sk

Abstract. *The formal semantics of programming languages forms the basis for many other formal methods used in software engineering. Its role is primarily in finding the meanings of programs and language constructs and in verifying the correctness of the design and implementation by means of formal procedures. Formal semantics has an irreplaceable place in the curricula of computer science and related fields of most universities. Its study requires a certain knowledge of mathematics and formal principles, therefore one of the results of our research is the visualization of semantic methods, primarily for educational purposes. In this article, we present an overview of our results in the field of definition and extension of semantic methods and their visualization.*

Keywords. correctness verification, curricula, formal methods, programming languages, semantic methods, software engineering, visualization, university didactic.

1 Introduction

Semantic methods belong to the frequently used formal methods for determining the meaning of programs written in programming languages. Individual methods are formulated to serve some specific purposes.

The first semantic method is denotational semantics, which defines the meaning of programs in terms of sets and functions. This method is not interesting in the process of program execution; it provides only the results with respect to the given input data. Therefore, this method is appropriate in designing programming languages. The second, very popular method is operational semantics. We differentiate two operational semantic methods. Natural semantics, which provides the meaning of the whole statements, and structural operational semantics, which analyzes every detail in program execution, follows the observable behavior of programs. The important part of structural operational semantics is abstract implementation on an abstract machine. Structural operational semantics is useful in the process of program implementation. There exist other semantic methods, for instance, axiomatic semantics, which adds preconditions and postconditions to each statement. These formulas serve to prove partial correctness of programs. Algebraic semantics specifies programs as abstract data types and models them as heterogeneous algebras. All these methods are included in the course Semantics of Programming Languages for graduate students.

In the following text, we present the related research and resources and we sketch the structure of the paper. Categories are very useful mathematical structures for modeling the programs and defining their categorical semantics. Categorical semantics of programs is in the literature oriented mostly on functional programming paradigm, for instance, (Hyland and Ong, 2000), but our approach is for the imperative paradigm modeled by category of states. We defined denotational categorical semantics for imperative languages and we extended it also for procedures (Steingartner et al., 2017). Another categorical structure we defined for structural operational semantics is coalgebra (Steingartner et al., 2020). We followed the ideas published in the book (Jacobs, 2016). The results of both models we explained in Section 2 (denotational semantics) and in Section 3 (operational semantics).

Because graphical illustration of semantics is easy to understand for students, our further work was to prepare visualization software appropriate for the teaching process. We were inspired by the work (Hannan and Miller, 1990) and we used the tool ANTLR 4 (Parr, 2013) as a very strong and useful software for our purposes. Our results (Steingartner et al., 2019; Steingartner, 2020; Steingartner and Sivý, 2023) are shortly characterized in Section 4.

Our results in semantic methods we applied in other languages as concatenative/compositional languages and domain-specific languages. Using (Herzberg and Reichert, 2009) we defined semantics for a simple compositional language KKJ (published in (Mihelič et al., 2021)). During the last decades, a new kind of programming languages, domain-specific languages were developed. We included in our course a simple domain-specific language Robot (Horpácsi and Kőszegi, 2015) and we showed how the semantics of such languages can be defined (Steingartner and Novitzká, 2021; Steingartner et al., 2022). Our results we sum up in Section 5. Section 6 concludes our paper.

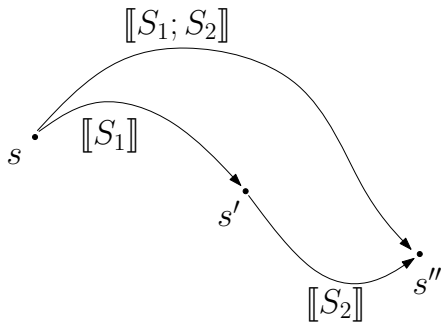


Figure 1: A compound morphism for a sequence of commands

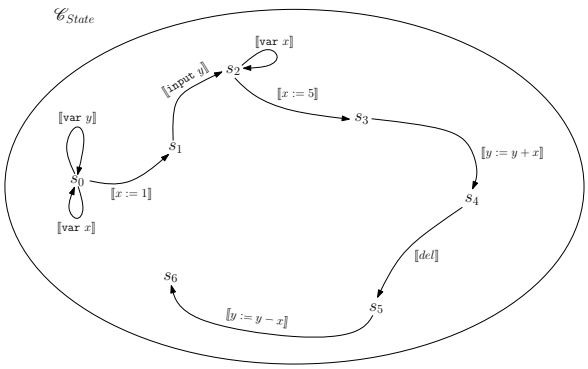


Figure 2: The path of execution P_1

2 Categorical semantics

A very useful mathematical structure for modeling programs and defining their semantics is a category. Therefore, our research focused on the use of categories for defining semantics.

We worked with imperative and procedural languages, where the essential concept is a state. The state is considered as an abstract model of computer memory which is independent from the real architecture and allows to view the changes in computer memory in a formal way. We define a state s as a function

$$s : \mathbf{Var} \times \mathbf{Level} \rightharpoonup \mathbf{Value},$$

where **Var** is a set of variables, **Level** expresses nesting level and **Value** is the set of integers together with undefined value $\perp$. Each state is then a list of triples, where x_i are variables, l is the nesting level and v is a value of a variable of the given level:

$$s = \langle ((x_1, 1), v_1), \ldots, ((x_n, l), v_n) \rangle .$$

States are the objects of our model, the category of states. Each variable of a program needs to be declared. A declaration `var` x is modeled as an endomorfism on a given state s and it adds a new triple to a given state:

$$[\![\texttt{var } x]\!] : s \to s.$$

The statements are modeled as category morphisms – for a statement S, we have

$$[\![S]\!] : s \to s'.$$

It expresses a change of state induced by execution of S. In such way, we defined denotational semantics (Steingartner et al., 2017), where we constructed a category of states and defined commands of a simple imperative language using morphisms.

For illustration, in Fig. 1, we show the semantics of compound statement $S_1; S_2$. We also consider block structure of our language. A block can contain declarations and statements. Here, the nesting level is incremented within entry to a block and the locally declared variables are forgotten at the end of a block:

$$[\![\texttt{begin } D; S \texttt{ end}]\!]s = \\ [\![del]\!] \circ [\![S]\!] \circ [\![D]\!](s \diamond \langle ((\texttt{begin}, l+1), \perp) \rangle),$$

where the keywords `begin` and `end` serve as fictive variables and del is a function releasing all declaration on the level $l + 1$, for $l \in \mathbf{N}_0$.

The semantics of a program is then a path in the category of states, as it is illustrated in Fig. 2 (for the Listing 1).

Listing 1: Program P_1 with nested block

```
var x;
var y;
x := 1;
input y;
begin
    var x;
    x:=5;
    y:=y+x;
end;
y:=y-x
```

We extended the simple procedural language with procedures to the procedural language and constructed a collection of state categories that are linked using functors for calling the procedure and returning to the calling subroutine. The advantage of our approach is the possibility of multiple procedure calls and the use of recursive calls (Figure 3).

Categories allow a graphical representation of program execution, which inspired us to introduce graphical visualization of semantics into the teaching process. Categorical structures are quite difficult for students studying technical sciences because they require quite deep knowledge of mathematics, so we tried to visualize the definition of the traditional approach to defining and visualizing semantics.

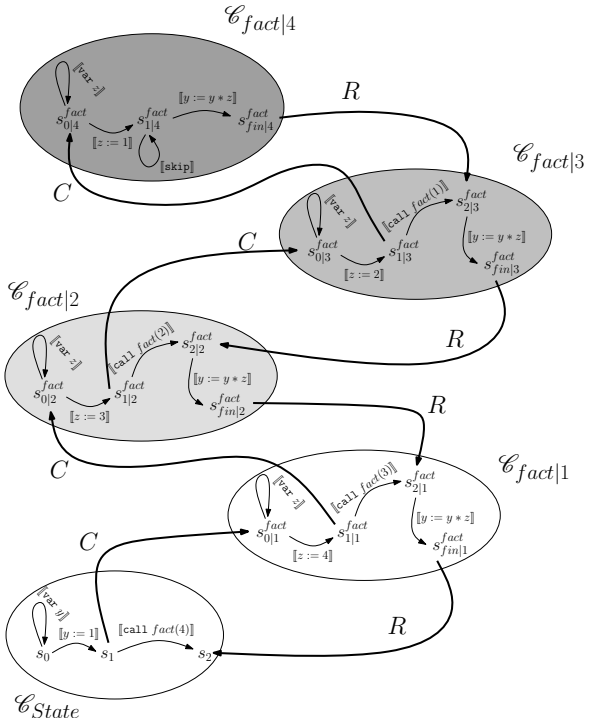


Figure 3: Categorical interpretation of recursive calls

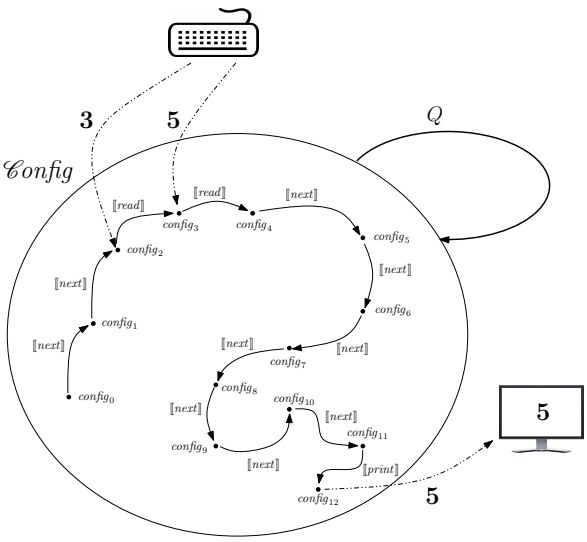


Figure 4: Semantics of program P_2 in coalgebra

3 Categorical operational semantics

The second semantic method, which we modeled with the help of categories, was structural operational semantics, which provides an overview of the detailed steps of the executed program, that is, it describes the behavior of the program. We extended the simple programming language with input and output blocks and commands, which required the construction of another categorical structure, a category of configurations whose objects are lists of commands, states, inputs and outputs (Steingartner et al., 2020). The commands of the language are modeled by morphisms (Figure 4).

Operational semantics provides behavior of programs, that is, it considers every detail in program execution. The best categorical structure, which is appropriate for this purpose, is a coalgebra. Coalgebra is defined as a polynomial endofunctor over state space, here over a base category. Every application of this functor expresses one step of program execution. To construct coalgebra for our purposes, we need to construct base category as a category of configurations. A configuration is a tuple

$$config = (\llbracket D^*; S^* \rrbracket, m, i^*, o^*),$$

where m is actual memory, i^* and o^* are lists of input and output values, respectively. The set **Config** of configuration is our state space. The morphisms of the base category are the transition operations $\llbracket next \rrbracket$ for statements, $\llbracket read \rrbracket$ and $\llbracket print \rrbracket$ for input and output statements, resp. Now we have the base category that is a model of our simple language *Jane*; thus we can proceed to construct a coalgebra modeling the behavior of programs written in *Jane*. The objects of our category form the state space of a coalgebra and the morphisms are the transition mappings. Hence, we construct the polynomial endofunctor for this kind of systems

$$Q(\mathbf{Config}) = 1 + \mathbf{Config} + O \times \mathbf{Config} + \mathbf{Config}^I.$$

Here $I \subseteq \mathbf{Value}$ denotes the domain of input values and $O \subseteq \mathbf{Value}$ denotes the domain of output values of the program execution. The operation $+$ in the definition of the functor Q expresses distinct, mutual exclusive results of the functor.

In Fig. 4, we show the semantics of the following program (Listing 2):

Listing 2: Program P_2

```
var x; var y;
read x; read y;
if x <= y then begin
        var z;
        z:=x; x:=y; y:=z
    end
else skip;
print x
```

4 Visualization of traditional semantic methods

The mentioned modules were solved within the KEGA educational projects and are gradually being introduced into the teaching process. The advantage is that they can serve as an illustrative aid for the teacher, either in lectures or exercises. In addition, they allow students to

individually practice the relevant issue and alert them to the mistakes they make when defining semantics. The modules work in interactive mode, that means that the student defines the semantics step by step, or enters a program and the module returns the semantics of the program. The response from students is very positive. This process can be considered as some pre-evaluation of our software. We observe how the students understand the principles of individual semantic methods faster and understand their principles more deeply than is the case with the classic definition of semantics on the blackboard or in a notebook.

4.1 Natural semantics

Natural semantics is one of the methods of operational semantics. This method is used quite often in practice when deriving the properties of languages and programs written in specific languages. It is also a useful method for designing compilers. Its principle is that the meaning of the program is constructed using derivation trees of natural semantics, where in individual nodes there are transition relations expressing changes in memory states (abstract expression of computer memory) as follows:

$$\langle S, s\rangle \rightarrow s'$$

where S represents a statement and s, s' stand for memory states.

The software module for the visualization of natural semantics (Figure 5, presented in (Steingartner et al., 2019)) works on the principle of translating the input program in the *Jane* language into tokens, with the help of which a derivation tree of natural semantics is subsequently generated. In the core of the application is a standard simple compiler working with the grammar of the *Jane* language. The compiler includes standard phases (lexical analysis, syntactic analysis, semantic analysis and target code generation – here in the form of tokens for derivation tree rendering). The module allows to input the program manually or load it from a file. Then the next step is setting the values of variables in the program. After compilation program performs the visualization of semantic evaluation applying the particular rules. The result provided (when the input program was correct) is then a derivation tree for natural semantics and the table with all states during the program execution. User can save the results for later use in graphical and textual form: program provides a cropped image containing the drawn derivation tree and LaTeX source for this tree.

4.2 Structural operational semantics

Structural operational semantics is a very popular semantic method that models each step of program execution based on transitive relations. Transition sessions can take place in two forms (depending on the result of the calculation) – either the result is a simple state or the next step (next configuration) in the computation:

$$\langle S, s\rangle \Rightarrow s', \quad \text{or} \quad \langle S, s\rangle \Rightarrow \langle S', s'\rangle.$$

For defining the grammar and parser construction, the ANTLR tool (Parr, 2013) with a great success was used. After providing the input sequence of statements and the initial state of variables, a parsing of the input code starts. Compilation provides errors if the input code is not correct. Otherwise, a parse tree is created. After the correct input, a new instance of *SOSGrammmarVisitor* class and the *MainSequence* class are created. So the first step is to obtain a sequence of statements and process them using the appropriate compilation (and semantic) rules. After obtaining the necessary objects, a derivation sequence is created based on the rules of structural operational semantics and the visualization is ready to be performed.

When a user launches the emulation (visualization) process, lexical and syntax analysis are performed. The result of a computation is provided and displayed as the whole computational sequence. User then can choose a mode of visualization: either the whole computation at once with final results or the step-wise emulation with showing all changes of states (and variables). At the end, user is allowed to store the work done: the whole output as a complete report (in one document), or only the state table, or the output computational sequence as a source code for LaTeX. Software module is depicted in Figure 6.

4.3 Abstract machine

Obviously, operational semantics is used to specify the meaning of programs and abstract machines to provide an intermediate representation of the language's implementation. Operational semantics can be presented as inference rules or, equivalently, as formulas in a weak meta-logic permitting quantification at first-order and second-order types. Abstract machines can be presented as rewrite rules describing single-step operations on the state of a computation. Such specifications provide an intermediate level of representation for many practical implementations of programming languages (Hannan and Miller, 1990).

For the complete visualization of particular processes of the abstract machine for the operational semantics, an emulator of abstract machine was developed (Steingartner and Sivý, 2023). Our visualization tool implementation fulfills several tasks. In its final form, it forms a tool accessible through a web browser, which is capable of processing various types of input, compilation (translation into tokens), subsequent visualization of program execution and exporting visualization results in various formats (Fig. 7). The implementation also includes a separate compiler executable via the command line, which is capable of output in various formats and is also capable of interpreting input

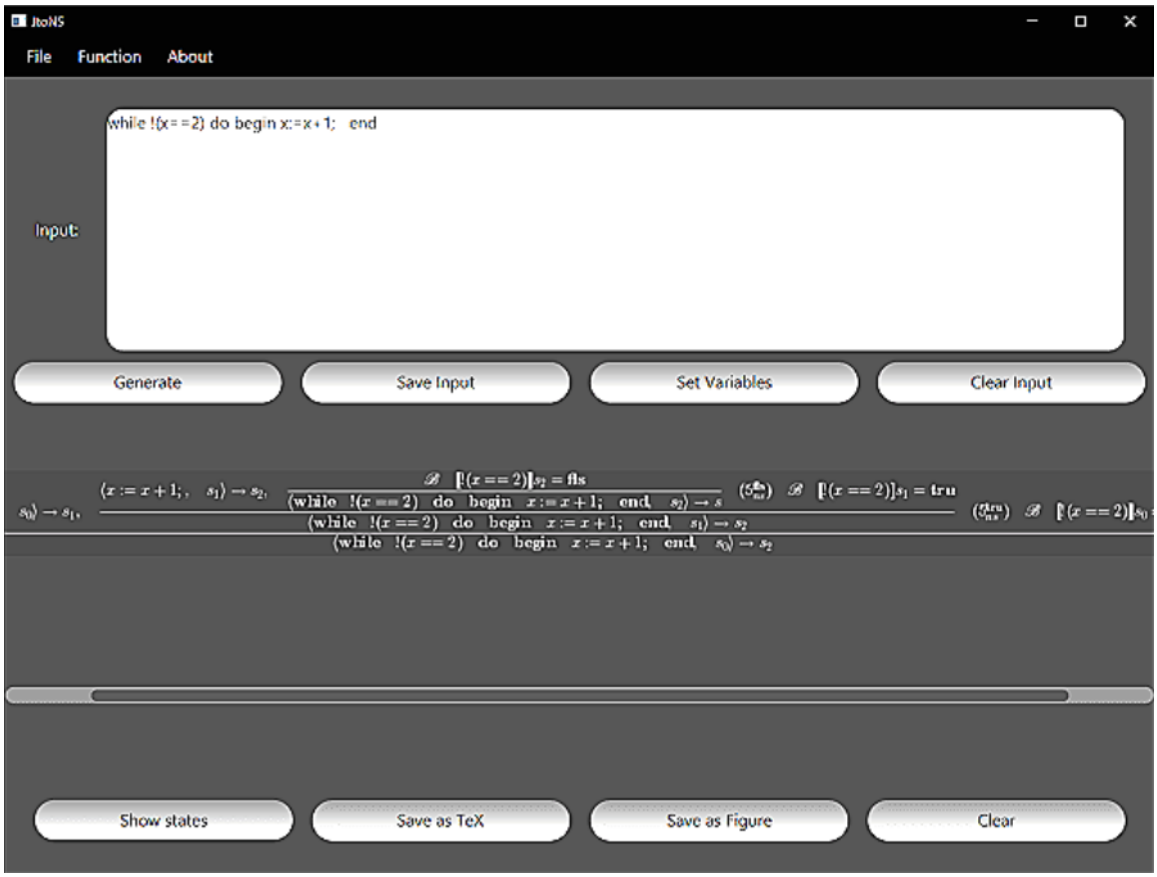


Figure 5: A look at the Natural Semantics Visualizer

programs. The entire tool is conceptually divided into three packages: the editor, which serves to process user input, the compiler, and the visualization tool itself.

4.4 Denotational semantics

The tool for denotational semantics is currently under development. The specificity of the design and development of the tool for denotational semantics results from the nature of the given method. Denotational, also called mathematical semantics, is a semantics that expresses the meaning of programs using mathematical objects (numbers, expressions and functions) and their composition. The main idea is that for a given program we construct a mathematical function that gives the same results as the given program (regardless of the order of individual steps in the program; of course, it is a function on states).

For simple statements and for the conditional statement pattern, the implementation is fairly straightforward. However, the challenge is the implementation of the denotation of the loop statement, for which in practice we construct a functional (a higher-order function) and then search for this functional's least fixed point, which expresses the meaning (denotation) of the loop statement.

It is obvious that for educational purposes automation of selected patterns of solving a fixed point of the functional appears to be a suitable solution, of course in interactive mode.

The development and methodology itself differ only little from the development of previous modules. Also in this case, thanks to the ANTLR tool, a grammar for input processing (*Jane* programs) is constructed. The compiler reads and translates the input program and decomposes it into a sequence of tokens, thanks to which it is possible to visualize calculations in denotational semantics. After completion, we plan to deploy the tool itself in the same way as the previously mentioned programs.

5 Extension of semantic methods for other languages

In addition to the application and extension of semantic methods in the area of the imperative paradigm, we further focused on the use of semantic methods also for languages from other paradigms. We focused on concatenative (or compositional) languages, which have the character of stack-based languages (Sec. 5.1). The very nature of domain-specific languages makes it possible to explore many of their interesting properties, including from the point of view of formal semantics, so we focused our efforts in research further in this direc-

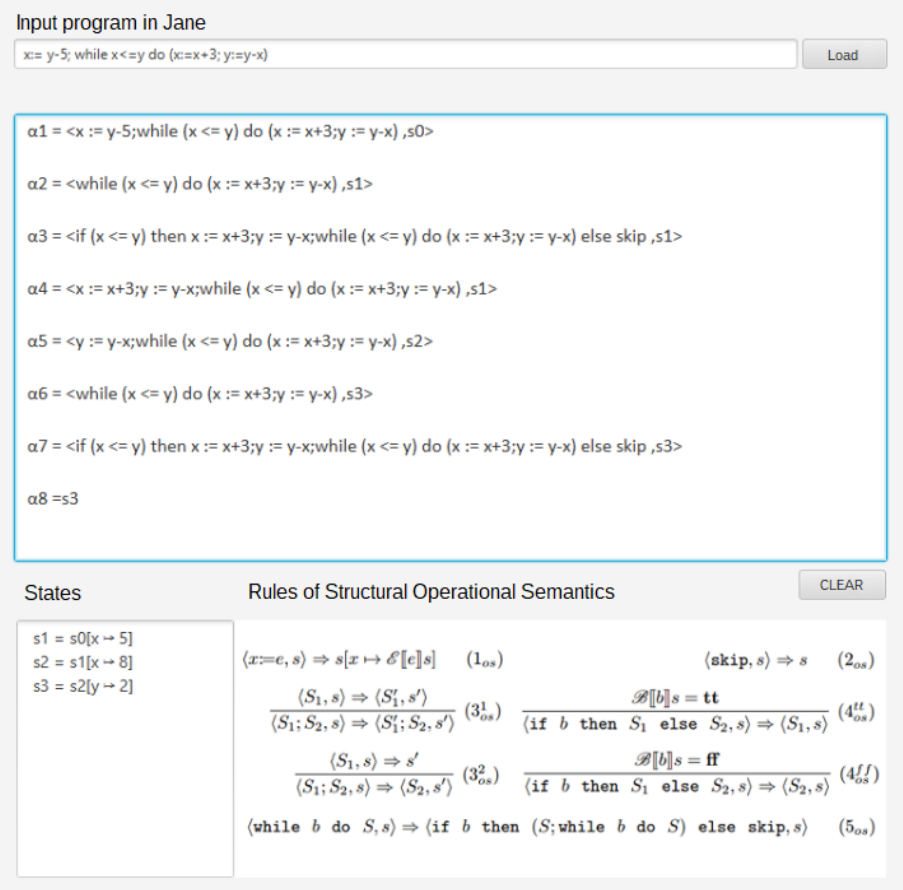


Figure 6: A module for the structural operational semantics

tion (Sec. 5.2).

5.1 Semantics of compositional languages

Another area we focused on was concatenative (stack-based) languages, which demonstrate to students that semantic methods are also suitable for other programming languages.

We have extended our research in the field of formal semantics of programming languages to the field of *concatenative languages*. This group of languages usually belongs to the functional paradigm (Herzberg and Reichert, 2009).

The name of this group of languages is derived from the property that the syntactic chaining (concatenation) of programs corresponds to the semantic composition of functions. The paradigm of concatenative languages is suitable for research into the fundamentals of software engineering and appears to be a suitable foundation for the future of programming.

We present a simple programming language KKJ, which contains a small set of language constructions, but offers a flexible and useful environment for programming. The development of the language and research related to its properties and semantics takes place in cooperation with the University of Ljubljana (Mihelič et al., 2021).

Thanks to its simplicity, the language provides several possibilities, for example, a clear definition of semantics, the possibility of direct development of an interpreter, the design of tools for analyzing programs for the purpose of optimization or verification, etc.

To describe the syntax of expressions $E \in \mathbf{Expr}$ in the programming language KKJ, we use the well-known BNF-notation:

$$E ::= \varepsilon \mid i \mid n \mid \{E\} \mid E\,E.$$

Here, ε stands for an empty expression, a numeral $i \in \mathbf{IntNum}$ and a name $n \in \mathbf{Name}$ are considered as expressions as well. As an example of calculations in this language, we show in Fig. 8 an evaluation of the program

$$[\![\texttt{14 \{dup dup\} \{add add\} compose apply}]\!]\, s$$

in particular steps.

The attentive reader will quickly realize that it is a calculation of the value of the expression 3×14.

5.2 Semantics of domain-specific languages

The latest contribution to the teaching of semantic methods is the definition of the semantics of domain-specific languages, which represent the modern trend

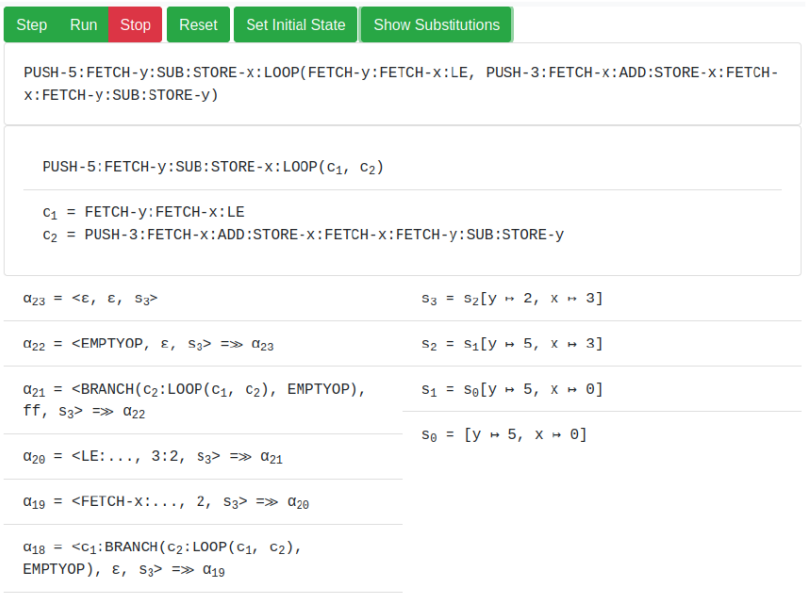


Figure 7: Window containing a visualization of abstract machine

$$
\begin{aligned}
&[14\ \{\text{dup dup}\}\ \{\text{add add}\}\ \text{compose apply}]\ s = \\
&\quad = [\text{apply}] \circ [\text{compose}] \circ [\{\text{add add}\}] \circ [\{\text{dup dup}\}] \circ [14]\ s \\
&\quad = [\text{apply}] \circ [\text{compose}] \circ \text{push}\ [\text{add add}] \circ \text{push}\ [\text{dup dup}] \circ [14]\ s \\
&\quad = [\text{apply}] \circ [\text{compose}]\ s \cdot 14 \cdot [\text{dup dup}] \cdot [\text{add add}] \\
&\quad = [\text{apply}]\ s \cdot 14 \cdot ([\text{add add}] \circ [\text{dup dup}]) \\
&\quad = [\text{add add}] \circ [\text{dup dup}]\ s \cdot 14 = [\text{add}] \circ [\text{add}] \circ [\text{dup}] \circ [\text{dup}]\ s \cdot 14 \\
&\quad = [\text{add}] \circ [\text{add}]\ s \cdot 14 \cdot 14 \cdot 14 = s \cdot 42
\end{aligned}
$$

Figure 8: Step-wise evaluation of the calculation in KKJ language

in the development of programming languages. For teaching, we chose a simple domain-specific language for moving the robot along a rectangular grid (Figure 9).

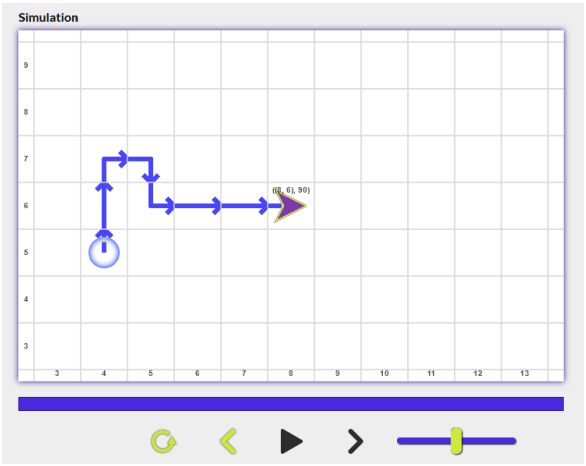


Figure 9: Visualization of robot movement in a grid

The language is quite easy itself. The structure of directions is defined by the production rule:

$$d ::= \textbf{left} \mid \textbf{right} \mid \textbf{up} \mid \textbf{down}$$

and we define the structure of statements as follows:

$$S ::= d \mid d\ n \mid \textbf{reset} \mid \textbf{skip} \mid S; S,$$

where

- $D\ n$ expresses n-steps of movement in given direction;
- **reset** defines movement to starting position;
- **skip** is the empty statement;
- $S; S$ is sequence of statements.

The second version of this language, which is closer to real languages, has the following syntax.

$$\begin{aligned} S ::= \quad & \textbf{turn left} \mid \textbf{turn right} \mid \textbf{forward} \\ & \textbf{forward}\ n \mid \textbf{reset} \mid \textbf{skip} \mid S; S. \end{aligned}$$

We explored and refined the definition of its denotational and structural operational semantics (formulated in (Horpácsi and Kőszegi, 2015)) and defined natural

semantics (Steingartner and Novitzká, 2021) and the first version of abstract machine for structural operational semantics (Steingartner et al., 2022) and thereby again showed students that semantic methods are applicable to such languages. An example of computation on an abstract machine is in Figure 10.

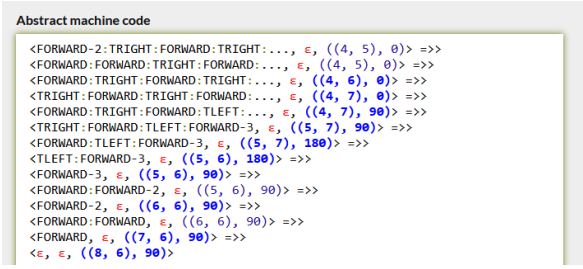


Figure 10: Example of computation on an abstract machine

Currently, we are extending this robot language with additional commands that allow to bypass the obstacle in the path of the robot and determine the size of the grid. Adding this problem to the language requires defining new commands (conditional, loop) so that the robot moves on a grid of a fixed size and knows how to avoid obstacles. A possible promising extension for the future is the addition of a third dimension in an orthogonal grid, allowing for semantic modeling of drone movement in space.

6 Conclusion

In this article, we presented the most important results of our research in the field of semantic methods. In our research, we extended the set of traditional semantic methods for imperative languages by two methods based on category theory, which were very positively received by students. Despite their mathematical nature, the categories are also positively received by students with a technical focus (with less mathematical background) due to their illustrative and expressive power. In our research, we further dealt with the extension of traditional methods for some other paradigms of languages, specifically for selected domain-specific languages and for concatenative (stack-oriented) languages. As a big positive, students receive software support for visualization and interactive processing of calculations using selected semantic methods. Hereby, we have presented a comprehensive view of research in the field of semantic methods and semantic modeling, as well as making them available to students in an interesting and interactive form.

Acknowledgments

This work was supported in the frame of the initiative project “Semantics-Based Rapid Prototyping of Domain-Specific Languages” under the bilateral program “Aktion Österreich – Slowakei, Wissenschafts- und Erziehungskooperation” granted by the Slovak Academic Information Agency and by the project KEGA 030TUKE-4/2023 – “Application of new principles in the education of IT specialists in the field of formal languages and compilers”, granted by the Cultural and Education Grant Agency of the Slovak Ministry of Education.

References

Hannan, J. and Miller, D. (1990). From operational semantics to abstract machines: Preliminary results. In Kahn, G., editor, *Proceedings of the 1990 ACM Conference on LISP and Functional Programming, LFP 1990, Nice, France, 27-29 June 1990*, pages 323–332. ACM.

Herzberg, D. and Reichert, T. (2009). Concatenative programming - an overlooked paradigm in functional programming. In *Proceedings of the 4th International Conference on Software and Data Technologies - Volume 1: ICSOFT,*, pages 257–262. INSTICC, SciTePress.

Horpácsi, D. and Kőszegi, J. (2015). Formal semantics. `https://dtk.tankonyvtar.hu/xmlui/handle/123456789/3736`. Accessed: Dec 14th, 2022.

Hyland, J. and Ong, C.-H. (2000). On Full Abstraction for PCF: I, II, and III. *Information and Computation*, 163(2):285–408.

Jacobs, B. (2016). *Introduction to Coalgebra: Towards Mathematics of States and Observation*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

Mihelič, J., Steingartner, W., and Novitzká, V. (2021). A denotational semantics of a concatenative/compositional programming language. *Acta Polytechnica Hungarica*, 18(4).

Parr, T. (2013). *The Definitive ANTLR 4 Reference*. Pragmatic Programmers, LLC, The, Raleigh.

Steingartner, W. (2020). Support for online teaching of the Semantics of Programming Languages course using interactive software tools. In *2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, pages 665–671.

Steingartner, W., Baraník, R., and Novitzká, V. (2022). Abstract machine for operational semantics of domain-specific language. In Chiusano, S.,

Cerquitelli, T., Wrembel, R., Nørvåg, K., Catania, B., Vargas-Solar, G., and Zumpano, E., editors, *New Trends in Database and Information Systems*, pages 413–424, Cham. Springer International Publishing.

Steingartner, W., Haratim, M., and Dostál, J. (2019). Software visualization of natural semantics of imperative languages - a teaching tool. In *2019 IEEE 15th International Scientific Conference on Informatics*, pages 509–514.

Steingartner, W. and Novitzká, V. (2021). Natural semantics for domain-specific language. In Bellatreche, L., Dumas, M., Karras, P., Matulevičius, R., Awad, A., Weidlich, M., Ivanović, M., and Hartig, O., editors, *New Trends in Database and Information Systems*, pages 181–192, Cham. Springer International Publishing.

Steingartner, W., Novitzká, V., Bačíková, M., and Š. Korečko (2017). New approach to categorical semantics for procedural languages. *Computing and Informatics*, 36(6).

Steingartner, W., Novitzká, V., and Schreiner, W. (2020). Coalgebraic operational semantics for an imperative language. *Computing and Informatics*, 38(5):1181–1209.

Steingartner, W. and Sivý, I. (2023). From high-level language to abstract machine code: An interactive compiler and emulation tool for teaching structural operational semantics. In *New Trends in Database and Information Systems: ADBIS 2023 Short Papers, workshop MADEISD*. forthcomming.

Publication Ethics And Malpractice Statement

Central European Conference on Information and Intelligent Systems – CECIIS is committed to upholding the highest standards of publication ethics and takes all possible measures against any publication malpractices for its double-blind peer-reviewed Conference Proceedings.

1. Duties of Editors and Track Directors

Editors are responsible for maintaining the integrity of the academic record, for having processes in place to assure the quality of the material they publish and for precluding business needs from compromising intellectual and ethical standards.

Track Directors are responsible that appropriate reviewers are selected for submissions (i.e. individuals who are able to judge the work and are free from disqualifying competing interests).

Publication decisions

Editors and track directors are responsible for deciding which of manuscripts submitted to CECIIS will be accepted for presentation and for publication of the final paper in the Conference Proceedings. This decision is based on the recommendation of reviewers. Main selection criteria are the contribution's importance, originality and clarity, and the study's validity and its relevance. Decision will not be influenced by the authors' race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy. Editors and track directors may be guided by the policies of the CECIIS Program Committee and constrained by such legal requirements as shall then be in force regarding libel, copyright infringement and plagiarism. Editors and track directors may confer with reviewers in making this decision.

Confidentiality

Editors and track directors must not disclose any information about a manuscript submitted to anyone other than the corresponding author, reviewers, potential reviewers, other editorial advisers, and the publisher, as appropriate.

Disclosure and conflicts of interest

Unpublished materials disclosed in a submitted manuscript will not be used in an editors´ or track directors' own research without the express written consent of the author.

2. Duties of Reviewers

Contribution to editorial decisions

Papers will be published in the Conference Proceedings after double-blind peer-reviewed process. The reviewers advise track directors and editors. Track directors communicate with authors, as required, and help them in improving quality of their manuscript. The reviewers don't know the author's identity and reviewers' comments to track directors and editors are confidential and before passing on to the author will be made anonymous. The names of the reviewers remain strictly confidential; with their identities known only to editors and corresponding track director.

Promptness

Any selected reviewer who feels unqualified to review the research reported in manuscript or knows that its prompt review will be impossible, should notify the track director and withdraw from the review process.

Confidentiality

Any manuscripts received for review must be treated as confidential documents. They must not be disclosed to or discussed with others except as authorized by editors.

Standards of objectivity

Reviews should be conducted objectively. Personal criticism of the author is inappropriate. Reviewers should express their views clearly with supporting arguments, if necessary with explanation.

Acknowledgement of sources

Reviewers should identify relevant published work that has not been cited by authors. They should point out whether observations or arguments derived from other publications are accompanied by the respective source. Reviewers will notify track directors or editors of any substantial similarity or overlap between the manuscript under consideration and any other published paper of which they have personal knowledge.

Disclosure and conflict of interest

Privileged information or ideas obtained through the review process must be kept confidential and not used for personal advantage. Reviewers should not consider manuscripts in which they have conflicts of interest resulting from competitive, collaborative, or other relationships or connections with any of the authors, companies, or institutions associated with the papers.

3. Duties of Authors

Reporting standards

Authors of manuscripts should present an accurate account of the work performed as well as an objective discussion of its significance. Underlying data should be represented accurately in the manuscript. A manuscript should contain sufficient detail and references to permit others to replicate the work. Fraudulent or knowingly inaccurate statements constitute unethical behavior and are unacceptable.

Data access and retention

Authors could be asked to provide the raw data in connection with a manuscript for review and should be prepared to provide public access to such data, if practicable, and should in any event be prepared to retain such data for a reasonable time after publication.

Originality and plagiarism

Authors should ensure that they have written entirely original works, and if authors have used the work and/or words of others that this has been appropriately cited or quoted. Plagiarism in any form, including the touting of material contained in another paper (of the same authors or some other author) with cosmetic changes as a new paper; copying or paraphrasing substantial parts of another's paper (without attribution), and claiming results from research conducted by others are among the numerous forms of plagiarism. In all its forms plagiarism constitutes unethical publishing behavior and is unacceptable.

Acknowledgement of sources

Proper acknowledgment of the work of others must always be given. Authors should cite publications that have been influential in determining the nature of the reported work.

Multiple, redundant or concurrent publication

An author should not in general publish manuscripts describing essentially the same research in more than one journal or primary publication. Submitting the same manuscript to more than one journal or conference concurrently constitutes unethical publishing behaviour and is unacceptable.

Authorship of the paper

Authorship should be limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study. All those who have made significant contributions should be listed as co-authors. Where there are others who have participated in certain substantive aspects of the manuscript, they should be acknowledged or listed as contributors.

The corresponding author should ensure that all appropriate co-authors and no inappropriate co-authors are included on the manuscript, and that all co-authors have seen and approved the final version of the manuscript and have agreed to its submission for publication.

Fundamental errors in published work

When an author discovers a significant error or inaccuracy in his/her own published work, it is the author’s obligation to promptly notify editors or publisher and cooperate with editors to retract or correct the paper.

If editors or the publisher learns from a third party that a published work contains a significant error, it is the obligation of the author to promptly retract or correct the paper or provide evidence to editrors of the correctness of the original paper.

Disclosure and conflicts of interest

All authors should disclose in their manuscript any financial or other substantive conflict of interest that might be construed to influence the results or interpretation of their manuscript. All sources of financial support for the paper should be disclosed.

Publisher's confirmation

In cases of alleged or proven scientific misconduct, fraudulent publication or plagiarism the publisher, in close collaboration with the editors, will take all appropriate measures to clarify the situation and to amend the paper in question. This includes the prompt publication of an erratum or, in the most severe cases, the complete retraction of the affected work.

Author Guidelines

Microsoft Word Template for the Central European Conference on Information and Intelligent Systems (CECIIS)

Author 1, Author 2
Author's Common Affiliation
Department/Institute
Full Address
{author1, author2}@domain.com

Author(s) Name(s)
Author Affiliation(s)
Department/Institute
Full Address(es)
E-mail(s)

Abstract. *The abstract is to be in fully-justified italicized text, at the top of the left-hand column as presented here, below the author information. Use the word "Abstract." in 10-point Times, boldface type, left positioned, initially capitalized, followed by the abstract in 10-point, single-spaced type, up to 100 words long.*
Leave one blank line after the abstract, and then begin the keywords. Use the word "Keywords." in 10-point Times, boldface type, left positioned, initially capitalized, followed by up to ten keywords in 10-point, separated by comma, as below.
Leave two blank lines after the keywords, and then begin the main text.

Keywords. CECIIS, conference paper, template

1 Introduction

This is an example of a Central European Conference on Information and Intelligent Systems (CECIIS) submission. These guidelines include complete description of the paper style including formatting, fonts, spacing, and related information for producing your proceedings manuscripts.

This template can be used to create your own paper but if you have any troubles or if in doubt do not hesitate to contact us via e-mail: ceciisoo@foi.hr or via the conference web site at http://www.ceciis.foi.hr.

2 Structure and Requirements for Research Paper (Examples)

Central European Conference on Information and Intelligent Systems (CECIIS) welcomes research papers on all topics of interest (Data and Knowledge Bases, Economics and Information Systems, Education for Information Society, Information and Communication Technologies, Information Systems Security, Intelligent Information Systems, Intelligent Transport Systems and Autonomous Vehicles, Software Engineering, Quality of Software and Services) as well as on the Conference theme/special emphasis announced on yearly base, addressing both, research theory and practice.

The following paper structure examples regarding the research type could be used for preparing a CECIIS paper:

1. Quantitative questionnaire-based: background research (state of the art); research problem; hypothesis or research question; description of instrument used; sampling; statistical methods used for analysis; discussion of results with limitations of research.
2. Qualitative research: background research (state of the art); research problem; research question; qualitative method used (case study research, interviews, content analysis etc.); discussion of results with limitations of research.
3. Algorithms and theoretical research: theory description with references; original theoretical contribution (theorem, algorithm etc.) with proofs or confirmations; future research and conclusion
4. Systematic literature review with meta-analysis: scope of research, research questions or research purpose; sources of literature with period covered (data bases, journals etc.), methods of analysis, conclusions with future research

Research paper should be theoretically and evidence based, clearly and consistently as well as ethically prepared and written. Central European Conference on Information and Intelligent Systems (CECIIS) has zero-tolerance policy on any kind of plagiarism. Paper containing text copied and used from another author(s) work without permission or from authors previously published own work, without referencing, will be retracted.

3 Paper Organization and Formatting

All papers should be written in English, up to 8 pages long (approximately 20 000 characters), and arranged in the following order:

- Main title
- Author(s), affiliation(s), full and e-mail addresses
- Abstract
- Keywords
- Body text (Main text)
- Footnotes
- Acknowledgements
- References

All printed material, including text and figures, must be kept within a print area of 16 cm (6.3") wide by 24.7 (9.7") high. Do not write or print anything outside the print area. Paper size is A4: 21x29.7 cm (8.3x11.7"). Margins (top, bottom, left, right) are 2.5 cm (1"). All *text* must be in a two-column format. Columns are to be 7.7 cm (3") wide, with a 0.6 cm (0.24") space between them. Text must be fully justified. Indent each paragraph by 0.5 cm (0.2").

The final submission has to be submitted in a single PDF file with all fonts that are used embedded, and also the source file (Microsoft Word, LaTeX - zip original files and images).

4 Paper Submission in Croatian Language

Authors who also want to publish the translation of their paper in Croatian, need to upload both the English and Croatian version of the paper (in the same document) for review. English and Croatian versions need to be in parallel columns (side by side). Left column is the paper in English and right column is the paper in Croatian. A camera-ready full paper should be prepared according to the Instructions for Authors and submitted following all the stages of the Paper Submission Process.

5 Main Title

The main title (on the first page) should begin 3,5 cm from the top edge of the page, centred, and in Times 18-point, bold face. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave a blank line after the title.

6 Author Name(s) and Affiliation(s)

Author names and affiliations are to be centred beneath the title and printed in Times 10-point type. Author names should be in bold face. Multiple authors shall be grouped by affiliation as shown in the title above. Include also e-mail addresses in 9-point `Courier/Typewriter` face. Follow the author information by two blank lines before main text.

7 Type-style and Fonts

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times that you have access to. Please avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

Also make sure that all fonts in the final PDF file are embedded. Thus we strongly encourage the use of Latex especially if you use non-standard fonts for formulas for your examples. You can make sure that your fonts are embedded by opening the PDF file in some PDF viewing program and taking a look at the properties of the document.

8 First-order Headings (May Extend to the Next Line)

For example, “1 Introduction”, should be Times 14-point boldface, initially capitalized, flush left, with two blank lines before, and one blank line after.

Don't use periods (“.”) after the heading number, use them only in lower order headings to separate them from higher order headings. For *long* headings use *a hanging indent* aligning the text to the right of the heading number as shown above.

All paragraphs in the text should be indented except for the first paragraph in a section.

8.1 Second-order Headings (May Extend to the Next Line)

As in this heading, they should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

8.1.1 Third-order Headings (May Extend to the Next Line)

Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 10-point Times, boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

9 Main Text

Type your main text in 10-point Times, single-spaced. Do not use double-spacing. All paragraphs should be indented 0.5 cm (0.2") except for the first paragraph in a section.

Be sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

9.1 Figures

All figures are to be included within the text. **Figure captions** are to be *below* the figures, in 10-point Times (or a similar serif font), normal face. Initially capitalize only the first word of each figure caption.

Figures are to be numbered consecutively with Arabic numerals throughout the paper, for example: "Figure 1. Database contexts", and are referred to in the text as Fig. 1, Fig. 2, etc.

9.1.1 Illustrations, Photographs and Graphs

Illustrations, photographs and graphs are considered as figures. All graphics should be centred. Your artwork must be in place in the article (preferably printed as part of the text rather than pasted up). If you are using photographs and are able to have halftones made at a print shop, use a 100- or 110-line screen. Supply the best quality photographs and illustrations possible.

Pencilled lines and very fine lines do not reproduce well. Remember, the quality of the book cannot be better than the originals provided. The conference proceedings will be printed in grey scale, but you can provide images in colour since the papers will also be published on the web site of the conference.

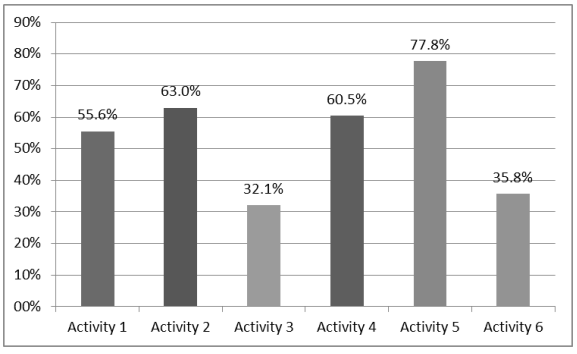


Figure 1. Times, 10 points, normal face

9.2 Tables

All tables are to be included within the text. **Table titles** are to be *above* the tables, in 10-point Times (or a similar serif font), normal face. Initially capitalize only the first word of each table title. Tables are to be numbered consecutively with Arabic numerals throughout the paper, for example: "Table 1. Input data", and are referred to in the text as Table 1, Table 2, etc. Column headings should be as brief as possible.

Table 1. Times, 10 points, normal face

	Column 1	Column 2	Column 3
Row 1	1	2	3
Row 2	4	5	6
Row 3	3	4	7
Row 4	2	6	4
Row 5	2	2	4

9.3 Equations

Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses, and referred to in the text by eq. 1, eq. 2 etc.

$$3x^2 + 3x - 23 = 0 \quad (1)$$

9.4 Footnotes

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced.[1]

10 In-text Citations and Referencing

In-text citation must be used to denote all text used from another author(s) work or from authors previously published own work. When referenced in the text, enclose the author surname(s) and the publication year in brackets, for example (Surname, year) or (Surname, year, p. x) where x is the page number of the cited reference. When referenced paper has more than three authors, cite the paper as (Surname et al., year).

List and number all bibliographical references, alphabetically sorted, in 10-point Times, single-spaced, with a hanging indent, at the end of your paper. Use 6-point paragraph spacing after each reference. References should be distributed evenly in both columns as much as possible, so use a column break where appropriate.

In the References section below you will find examples of common types of bibliographical items: book (Mayer, 2009, p. 52), journal paper (Bule & Peer, 2014), journal paper with more than three authors (Shukor et al., 2014), conference proceedings paper (Steingartner & Novitzká, 2015), report or standard

[1] Use Times 8-point type, single-spaced. Avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

(*ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*, 2011) and web page ("Web Content Accessibility Guidelines (WCAG) 2.0," 2008). For other types of bibliographical items refer the **APA style guidelines**. It should be ensured that every reference cited in the text is also listed in the References section (and vice versa).

Acknowledgments

Acknowledgments, if necessary, should appear in a separate paragraph preceding the references.

References

Bule, J., & Peer, P. (2014). Technical, Legal, Economic and Social Aspects of Biometrics for Cloud Computing. *Journal of Information and Organizational Sciences*, 38(2), 83–95.

ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. (2011).

Mayer, R. E. (2009). *Multimedia Learning* (Second Edi.). New York: Cambridge University Press.

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, 116, 4844–4853. doi:10.1016/j.sbspro.2014.01.1036

Steingartner, W., & Novitzká, V. (2015). A new approach to operational semantics by categories. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (pp. 247–254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

Web Content Accessibility Guidelines (WCAG) 2.0. (2008). Retrieved from https://www.w3.org/TR/WCAG20/

Microsoft Word Template for papers in Croatian Language for the Central European Conference on Information and Intelligent Systems

Microsoft Word predložak za radove na hrvatskome jeziku za Central European Conference on Information and Intelligent Systems

Author 1, Author 2
Author's Common Affiliation
Naziv ustanove
Department/Institute
Naziv odjela/instituta
Full Address / Adresa
{author1, author2}@domain.com

Author(s) Name(s)
Author Affiliation(s)
Naziv ustanove
Department/Institute
Naziv odjela/instituta
Full Address / Adresa
E-mail(s)

Abstract*. The abstract is to be in fully-justified italicized text, at the top of the column as presented here, below the author information. Use the word "Abstract." in 10-point Times, boldface type, left positioned, initially capitalized, followed by the abstract in 10-point, single-spaced type, up to 100 words long.*
Leave one blank line after the abstract, and then begin the keywords. Use the word "Keywords." in 10-point Times, boldface type, left positioned, initially capitalized, followed by up to ten keywords in 10-point, separated by comma, as below.
Leave two blank lines after the keywords, and then begin the main text.

Keywords. CECIIS, conference paper

1 Introduction

This is an example of a Central European Conference on Information and Intelligent Systems (CECIIS) submission. These guidelines include complete description of the paper style including formatting, fonts, spacing, and related information for producing your proceedings manuscripts.

This template can be used to create your own paper but if you have any troubles or if in doubt do not hesitate to contact us via e-mail: ceciis@foi.unizg.hr.

Sažetak*. Tekst sažetka treba biti obostrano poravnat i pisan kurzivom, na vrhu stupca kako je prikazano ovdje ispod podataka o autoru. Upotrijebite riječ "Sažetak." u fontu Times veličine 10, podebljano, poravnato ulijevo, velikim početnim slovom. Nakon riječi "Sažetak." slijedi sažetak rada u fontu veličine 10, jednostruki prored, dužine do 100 riječi.*
Nakon sažetka ostavite jedan redak prazan, a zatim započnite ključne riječi. Upotrijebite riječi "Ključne riječi." u fontu Times veličine 10, podebljano, lijevo poravnanje, s velikim početnim slovom, nakon čega slijedi do deset ključnih riječi u fontu veličine 10, odvojenih zarezom, kao u nastavku.
Ostavite dva prazna retka nakon ključnih riječi, a zatim započnite glavni tekst.

Ključne riječi. CECIIS, konferencijski rad

1 Uvod

Ovo je predložak rada za Central European Conference on Information and Intelligent Systems (CECIIS). Smjernice uključuju cjelovit opis stila rada, uključujući oblikovanje, fontove, razmake i srodne informacije za objavu članaka u zborniku.

Predložak možete koristiti za izradu vlastitog rada, a ako imate bilo kakvih problema ili pitanja, ne ustručavajte se kontaktirati nas putem e-pošte: ceciis@foi.unizg.hr.

2 Structure and Requirements for Research Paper (Examples)

Central European Conference on Information and Intelligent Systems (CECIIS) welcomes research papers on all topics of interest (Computer Games, Data and Knowledge Bases, Education and Learning Analytics, Emerging trends in ICT, ICT Entrepreneurship and Innovation, Intelligent Information Systems, Quality of Software and Services, Software Engineering, Strategic Planning and Decision making) as well as on the Conference theme/special emphasis announced on yearly base, addressing both, research theory and practice.

The following paper structure examples could be used for preparing a CECIIS paper:

1. Quantitative questionnaire-based: background research (state of the art); research problem; hypothesis or research question; description of instrument used; sampling; statistical methods used for analysis; discussion of results with limitations of research.
2. Qualitative research: background research (state of the art); research problem; research question; qualitative method used (case study research, interviews, content analysis etc.); discussion of results with limitations of research.
3. Algorithms and theoretical research: theory description with references; original theoretical contribution (theorem, algorithm etc.) with proofs or confirmations; future research and conclusion
4. Systematic literature review with meta-analysis: scope of research, research questions or research purpose; sources of literature with period covered (data bases, journals etc.), methods of analysis, conclusions with future research

Research paper should be theoretically and evidence based, clearly and consistently as well as ethically prepared and written. Central European Conference on Information and Intelligent Systems (CECIIS) has zero-tolerance policy on any kind of plagiarism. Paper containing text copied and used from another author(s) work without permission or from authors previously published own work, without referencing, will be retracted.

3 Paper Organization and Formatting

All papers should be written in English, up to 8 pages long (approximately 20 000 characters), and arranged in the following order:

- Main title

2 Struktura i zahtjevi za istraživački rad (primjeri)

Central European Conference on Information and Intelligent Systems (CECIIS) očekuje istraživačke radove na teme od interesa (računalne igre, baze podataka i baze znanja, edukacija i analitika učenja, novi trendovi u ICT-u, ICT poduzetništvo i inovacije, inteligentni informacijski sustavi, kvaliteta softvera i usluga, programsko inženjerstvo, strateško planiranje i donošenje odluka), kao i na godišnju temu konferencije, a bave se kako teorijskim, tako i primijenjivim istraživanjem.

Sljedeći primjeri strukture rada mogli bi se koristiti za pripremu rada za CECIIS:

1. Kvantitativni upitnik: pregled dosadašnjih istraživanja; problem istraživanja; hipoteza ili istraživačko pitanje; opis korištenog instrumenta; uzorkovanje; statističke metode korištene za analizu; rasprava o rezultatima s ograničenjima istraživanja.
2. Kvalitativno istraživanje: pregled dosadašnjih istraživanja; problem istraživanja; istraživačko pitanje; korištena kvalitativna metoda (istraživanje slučaja, intervjui, analiza sadržaja itd.); rasprava o rezultatima s ograničenjima istraživanja.
3. Algoritmi i teorijsko istraživanje: opis teorije s referencama; izvorni teorijski doprinos (teorem, algoritam, itd.) s dokazima ili potvrdama; buduće istraživanje i zaključak
4. Sustavni pregled literature s metaanalizom: opseg istraživanja, istraživačka pitanja ili svrha istraživanja; izvori literature s obuhvaćenim razdobljem (baze podataka, časopisi itd.), metode analize, zaključci o budućim istraživanjima

Istraživački rad trebao bi biti teorijski i zasnovan na dokazima, jasno i dosljedno te etički pripremljen i napisan. Central European Conference on Information and Intelligent Systems (CECIIS) ima politiku nulte tolerancije prema bilo kojoj vrsti plagijarizma. Rad koji sadrži tekst kopiran i korišten iz djela drugog autora bez odobrenja ili iz prethodno objavljenog vlastitog djela, bez referenciranja, bit će odbačen.

3 Organizacija i formatiranje rada

Svi radovi trebaju biti predani na engleskom jeziku, dužine do 8 stranica (oko 20 000 znakova) i organizirani prema sljedećem redoslijedu:

- Naslov rada
- Autor(i), pripadnost(i), adrese i e-mail adrese
- Sažetak

- Author(s), affiliation(s), full and e-mail addresses
- Abstract
- Keywords
- Body text (Main text)
- Footnotes
- Acknowledgements
- References

All printed material, including text and figures, must be kept within a print area of 16 cm (6.3") wide by 24.7 (9.7") high. Do not write or print anything outside the print area. Paper size is A4: 21x29.7 cm (8.3x11.7"). Margins (top, bottom, left, right) are 2.5 cm (1"). All *text* must be in a two-column format. Columns are to be 7.7 cm (3") wide, with a 0.6 cm (0.24") space between them. Text must be fully justified. Indent each paragraph by 0.5 cm (0.2").

The final submission has to be submitted in a single PDF file with all fonts that are used embedded, and also the source file (Microsoft Word, LaTeX).

4 Paper Submission in Croatian Language

Authors who also want to publish the translation of their paper in Croatian, need to upload both the English and Croatian version of the paper (in the same document) for review. English and Croatian versions need to be in parallel columns (side by side). Left column is the paper in English and right column is the paper in Croatian. A camera-ready full paper should be prepared according to the Instructions for Authors and submitted following all the stages of the Paper Submission Process.

5 Main Title

The main title should begin 3,5 cm from the top edge of the page, centred, and in Times 18-point, bold face. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave a blank line after the title.

6 Author Name(s) and Affiliation(s)

Author names and affiliations are to be centred beneath the title and printed in Times 10-point type. Author names should be in bold. Multiple authors shall be grouped by affiliation as shown in the title above. Include also e-mail addresses in 9-point

- Ključne riječi
- Tekst rada
- Fusnote
- Zahvale
- Reference

Sav materijal, uključujući tekst i slike, mora biti u području ispisa od 16 cm (6,3") širine i 24,7 (9,7") visine. Ne pišite i ne ispisujte ništa izvan područja ispisa. Veličina papira je A4: 21x29,7 cm (8,3x11,7"). Margine (gore, dolje, lijevo, desno) su 2,5 cm (1"). Sav tekst mora biti u formatu s dva stupca. Stupci trebaju biti široki 7,7 cm (3"), a između njih moraju biti razmaci od 0,6 cm (0,24"). Tekst mora biti obostrano poravnat. Uvucite svaki odlomak za 0,5 cm (0,2").

Konačni rad mora se poslati u jednoj PDF datoteci s ugrađenim svim fontovima koji se koriste, kao i izvornoj datoteci (Microsoft Word, LaTeX).

4 Predaja rada na hrvatskome jeziku

Autori koji žele objaviti i prijevod svog rada na hrvatskome jeziku, trebaju predati i englesku i hrvatsku verziju rada (u istom dokumentu) na recenziju. Engleska i hrvatska verzija trebaju biti u paralelnim stupcima (jedan pored drugog). Lijevi stupac je rad na engleskom jeziku, a desni stupac je rad na hrvatskome jeziku. Finalnu verziju cjelovitog rada treba pripremiti u skladu s Uputama za autore i predati slijedeći sve faze procesa predaje radova.

5 Naslov rada

Naslov rada (na prvoj stranici) trebao bi započeti 3,5 cm od gornjeg ruba stranice, centrirano, u fontu Times veličine 18, podebljano. Naslov započinje velikim početnim slovom. Ostavite jedan prazan redak nakon naslova.

6 Imena autora i pripadnost(i)

Imena autora i pripadnost trebaju biti centrirani ispod naslova i pisani fontom Times veličine 10. Imena autora trebaju biti podebljana. Više autora bit će grupirano prema pripadnosti kao što je prikazano u ovom predlošku. Uključite i adrese e-pošte u fontu `Courier/Typewriter` veličine 9. Nakon podataka o autoru slijede dva prazna retka prije nego počne glavni tekst.

`Courier/Typewriter` face. Follow the author information by two blank lines before main text.

7 Type-style and Fonts

Wherever Times is specified, Times Roman or Times New Roman may be used. Please avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

Also make sure that all fonts in the final PDF file are embedded. We strongly encourage the use of LaTeX, especially if you use non-standard fonts for formulas for your examples.

8 First-order Headings (May Extend to the Next Line)

For example, “1 Introduction”, should be Times 14-point boldface, initially capitalized, flush left, with two blank lines before, and one blank line after.

Don't use periods (“.”) after the heading number, use them only in lower order headings to separate them from higher order headings. For *long* headings use *a hanging indent* aligning the text to the right of the heading number as shown above.

All paragraphs in the text should be indented except for the first paragraph in a section.

8.1 Second-order Headings (May Extend to the Next Line)

As in this heading, they should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

8.1.1 Third-order Headings (May Extend to the Next Line)

Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 10-point Times, boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

9 Main Text

Type your main text in 10-point Times, single-spaced. Do not use double-spacing. All paragraphs should be indented 0.5 cm (0.2") except for the first paragraph in a section.

Be sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

7 Stil i fontovi

Gdje god je naveden font Times, može se koristiti Times Roman ili Times New Roman. Molimo izbjegavajte upotrebu bitmapiranih fontova ako je moguće. Poželjni su fontovi True-Type 1.

Također provjerite jesu li ugrađeni svi fontovi u konačnu PDF datoteku. Pritom, snažno preporučamo upotrebu LaTeX-a, posebno ako za svoje primjere koristite nestandardne fontove za formule.

8 Naslovi prvog reda (mogu se nastaviti u sljedeći redak)

Na primjer, "1 Uvod" treba biti font Times veličine 14, podebljano, s velikim početnim slovom, poravnato ulijevo, s dva prazna retka prije i jednim praznim redom nakon naslova.
Ne upotrebljavajte točke (".") iza broja naslova, koristite ih samo u naslovima nižeg reda kako biste ih odvojili od naslova višeg reda. Za duge naslove upotrijebite “hanging indent” poravnavajući tekst udesno od broja naslova, kao što je prikazano gore.

Svi odlomci u tekstu trebaju biti uvučeni, osim prvog odlomka u odjeljku.

8.1 Naslovi drugog reda (mogu se nastaviti u sljedeći redak)

Kao i u ovom naslovu, trebali bi biti font Times veličine 12, s velikim početnim slovom, poravnati lijevo, s jednim praznim retkom prije i jednim nakon. Koristite “hanging indent” za duge naslove.

8.1.1 Naslovi trećeg reda (mogu se nastaviti u sljedeći redak)

Naslovi trećeg reda, kao u ovom paragrafu, ne preporučuju se. Međutim, ako ih morate koristiti, upotrijebite font Times veličine 10, podebljano, s velikim početnim slovom, poravnati ulijevo, s jednim praznim retkom prije i jednim nakon. Koristite “hanging indent” za duge naslove.

9 Tekst rada

Tekst rada pišite fontom Times veličine 10, s jednostrukim proredom. Ne koristite dvostruki prored. Svi odlomci trebaju biti uvučeni 0,5 cm (0,2"), osim prvog odlomka u odjeljku.

Provjerite je li vaš tekst poravnat obostrano - to jest, i lijevo i desno. Molimo ne stavljajte dodatne prazne retke između odlomaka.

9.1 Figures

All figures are to be included within the text. **Figure captions** are to be *below* the figures, in 10-point Times (or a similar serif font), normal face. Initially capitalize only the first word of each figure caption.

Figures are to be numbered consecutively with Arabic numerals throughout the paper, for example: “Figure 1. Database contexts”, and are referred to in the text as Fig. 1, Fig. 2, etc.

9.1.1 Illustrations, Photographs and Graphs

Illustrations, photographs and graphs are considered as figures. All graphics should be centred. Your artwork must be in place in the article (preferably printed as part of the text rather than pasted up). Supply the best quality photographs and illustrations possible.

Very fine lines do not reproduce well. The conference proceedings will be printed in grey scale, but you can provide images in colour since the papers will also be published on the conference web site.

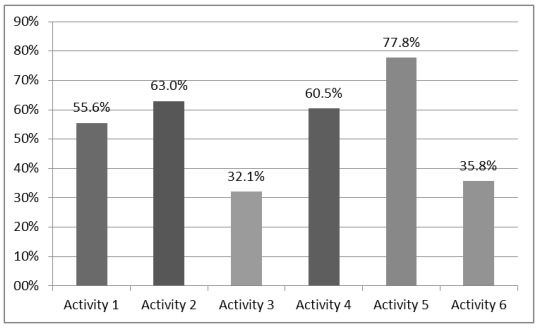


Figure 1. Times, 10 points, normal face

9.2 Tables

All tables are to be included within the text. Table titles are to be *above* the tables, in 10-point Times, normal face. Initially capitalize only the first word of each table title. Tables are to be numbered consecutively with Arabic numerals throughout the paper, for example: “Table 1. Input data”, and is referred to in the text as Table 1.

Table 1. Times, 10 points, normal face

	Column 1	Column 2	Column 3
Row 1	1	2	3
Row 2	4	5	6
Row 3	3	4	7
Row 4	2	6	4
Row 5	2	2	4

9.1 Slike

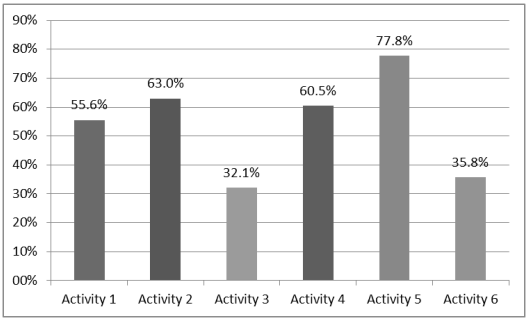
Sve slike trebaju biti uključene u tekst. Nazivi slika trebaju biti ispod slika, u fontu Times veličine 10 (ili sličnom fontu). Samo prvu riječ svakog naziva slike pišite velikim početnim slovom.

Slike se numerira uzastopno arapskim brojevima u cijelom radu, na primjer: "Slika 1. Konteksti baze podataka", a u tekstu se navodi kao Slika 1.

9.1.1 Ilustracije, fotografije i grafovi

Ilustracije, fotografije i grafikoni smatraju se slikama. Sva grafika treba biti centrirana. Svaka slika ima svoje mjesto u članku. Osigurajte najbolju moguću kvalitetu fotografija i ilustracija.

Vrlo tanke linije ne reproduciraju se dobro. Zbornik radova s konferencije bit će tiskan u nijansama sive boje, ali možete osigurati slike u boji jer će radovi biti objavljeni i na mrežnoj stranici konferencije.



Slika 1. Times, veličina 10

9.2 Tablice

Sve tablice trebaju biti uključene u tekst. Naslovi tablica trebaju biti iznad tablica, pisane fontom Times veličine 10. Velikim početnim slovom pišite samo prvu riječ svakog naslova tablice. Tablice se u radu redom numeriraju arapskim brojevima, na primjer: "Tablica 1. Ulazni podaci", a u tekstu se naziva Tablica 1.

Tablica 1. Times, 10 points, normal face

	Column 1	Column 2	Column 3
Row 1	1	2	3
Row 2	4	5	6
Row 3	3	4	7
Row 4	2	6	4
Row 5	2	2	4

9.3 Jednadžbe

9.3 Equations

Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses, and referred to in the text by eq. 1, eq. 2 etc.

$$3x^2 + 3x - 23 = 0 \quad (1)$$

9.4 Footnotes

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced.[1]

10 In-text Citations and Referencing

In-text citation must be used to denote all text used from another author(s) work or from authors previously published own work. When referenced in the text, enclose the author surname(s) and the publication year in brackets, for example (Surname, year) or (Surname, year, p. x) where x is the page number of the cited reference. When referenced paper has more than three authors, cite the paper as (Surname et al., year).

List and number all bibliographical references, alphabetically sorted, in 10-point Times, single-spaced, with a hanging indent, at the end of your paper. Use 6-point paragraph spacing after each reference. References should be distributed evenly in both columns as much as possible, so use a column break where appropriate.

In the References section below you will find examples of common types of bibliographical items: book (Mayer, 2009, p. 52), journal paper (Bule & Peer, 2014), journal paper with more than three authors (Shukor et al., 2014), conference proceedings paper (Steingartner & Novitzká, 2015), report or standard (*ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*, 2011) and web page ("Web Content Accessibility Guidelines (WCAG) 2.0," 2008). For other types of bibliographical items refer the **APA style guidelines**. It should be ensured that every reference cited in the text is also listed in the References section (and vice versa).

Acknowledgments

Jednadžbe se numeriraju slijedno s desne strane arapskim brojevima u zagradama, a u tekstu se na njih poziva tekstom Jednadžba 1, Jednadžba 2 itd.

$$3x^2 + 3x - 23 = 0 \quad (1)$$

9.4 Fusnote

Koristite fusnote štedljivo (ili ih uopće ne koristite!) i stavite ih na dno stupca na stranici na kojoj su navedeni.[2]

10 Navodi u tekstu i reference

Citiranje u tekstu mora se koristiti za označavanje cijelog teksta korištenog iz djela drugog autora ili prethodno objavljenih vlastitih djela. Kad se citira u tekstu, prezimena autora i godinu izdanja priložite u zagrade, na primjer (Prezime, godina) ili (Prezime, godina, str. X) gdje je X broj stranice citirane reference. Kada referentni rad ima više od tri autora, navedite ga kao (Prezime i sur., Godina).

Navedite sve bibliografske reference, poredane po abecedi, u fontu Times veličine 10, s jednostrukim proredom, s "hanging indent", na kraju vašeg rada. Upotrijebite razmak paragrafa veličine 6 nakon svake reference. Reference bi trebale biti raspoređene u oba stupca što ravnomjernije, stoga upotrijebite prijelom stupca tamo gdje je to prikladno.

U odjeljku Reference u nastavku naći ćete primjere uobičajenih vrsta bibliografskih jedinica: knjiga (Mayer, 2009., str. 52.), rad u časopisu (Bule & Peer, 2014.), članak s više od tri autora (Shukor i sur., 2014), članak s konferencije (Steingartner i Novitzká, 2015), izvještaj ili standard (*ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models, 2011*) i mrežna stranica ("Web Content Accessibility Guidelines (WCAG) 2.0," 2008). Za ostale vrste bibliografskih jedinica pogledajte **smjernice za stil APA**. Potrebno je osigurati da svaka referenca navedena u tekstu bude navedena i u odjeljku Reference (i obrnuto).

Zahvale

Zahvale bi se, ako je potrebno, trebale pojaviti u zasebnom odlomku koji prethodi referencama.

[1] Use Times 8-point type, single-spaced. Avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

[2] Koristite font Times veličine 8, s jednostrukim proredom. Izbjegavajte potpuno koristiti fusnote i u tekst uvrstite potrebna periferna zapažanja (u zagrade, ako želite, kao u ovoj rečenici).

Acknowledgments, if necessary, should appear in a separate paragraph preceding the references.

References

Bule, J., & Peer, P. (2014). Technical, Legal, Economic and Social Aspects of Biometrics for Cloud Computing. *Journal of Information and Organizational Sciences*, 38(2), 83–95.

ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. (2011).

Mayer, R. E. (2009). *Multimedia Learning* (Second Edi.). New York: Cambridge University Press.

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, 116, 4844–4853. doi:10.1016/j.sbspro.2014.01.1036

Steingartner, W., & Novitzká, V. (2015). A new approach to operational semantics by categories. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (pp. 247–254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

Web Content Accessibility Guidelines (WCAG) 2.0. (2008). Retrieved from https://www.w3.org/TR/WCAG20/

Reference

Bule, J., & Peer, P. (2014). Technical, Legal, Economic and Social Aspects of Biometrics for Cloud Computing. *Journal of Information and Organizational Sciences*, 38(2), 83–95.

ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. (2011).

Mayer, R. E. (2009). *Multimedia Learning* (Second Edi.). New York: Cambridge University Press.

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, 116, 4844–4853. doi:10.1016/j.sbspro.2014.01.1036

Steingartner, W., & Novitzká, V. (2015). A new approach to operational semantics by categories. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (pp. 247–254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

Web Content Accessibility Guidelines (WCAG) 2.0. (2008). Retrieved from https://www.w3.org/TR/WCAG20/

CECIIS 2023 Sponsors

CECIIS 2023 GOLD SPONSORS

INFODOM

ORACLE

CECIIS 2023 BRONZE SPONSOR

CECIIS 2023 Program Committee

Neven Vrček, Program Committee Chair, University of Zagreb, Croatia

Victoria Appatova, University of Cincinnati, USA
Igor Balaban, University of Zagreb, Croatia
Nina Begičević Ređep, University of Zagreb, Croatia
Marko Bohanec, Jožef Stefan Institute, Slovenia
Leo Budin, University of Zagreb, Croatia
Tonči Carić, University of Zagreb, Croatia
Carlos Carrascosa, Universitat Politècnica de València, Spain
Anna Czarczyńska, Kozminski University, Poland
Luis de Marcos Ortega, University of Alcalá, Spain
Blaženka Divjak, University of Zagreb, Croatia
Jasminka Dobša, University of Zagreb, Croatia
Juraj Ďuďák, Slovak University of Technology in Bratislava, Slovakia
Matjaž Gams, Jožef Stefan Institute, Slovenia
Gabriel Gaspar, Slovak University of Technology in Bratislava, Slovakia
Gordan Gledec, University of Zagreb, Croatia
Andrina Granić, University of Split, Croatia
Lourdes Guàrdia, Open University of Catalonia, Spain
Marjan Heričko, University of Maribor, Slovenia
Dieter Hertweck, Herman Hollerith Research Centre, Germany
Edouard Ivanjko, University of Zagreb, Croatia
Dimitris Karagiannis, University of Vienna, Austria
Valentina Kirinić, University of Zagreb, Croatia
Marina Klačmer Čalopa, University of Zagreb, Croatia
Michal Kohani, University of Zilina, Slovakia
Mario Konecki, University of Zagreb, Croatia
Melita Draganić, University of Zagreb, Croatia
Marjan Krašna, University of Maribor, Slovenia
Alen Lovrenčić, University of Zagreb, Croatia
Sandra Lovrenčić, University of Zagreb, Croatia
Ivan Luković, University of Belgrade, Serbia
Ivan Magdalenić, University of Zagreb, Croatia
Aleksandar Marković, University of Belgrade, Serbia
Marta Mas Machuca, UIC Barcelona, Spain
Dunja Mladenić, Jožef Stefan Institute, Slovenia
Oliver Moravčik, Slovak University of Technology, Slovakia
Enrique Mu, Carlow University, USA
Karlo Nyarko, University of Osijek, Croatia
Jan Paralič, Technical University of Košice, Slovakia
Elisabeth Pergler, University of Applied sciences, Austria
Wolf Rauch, University of Graz, Austria
Sonja Ristić, University of Novi Sad, Serbia
Markus Schatten, University of Zagreb, Croatia
Zlatko Stapić, University of Zagreb, Croatia
William Steingartner, Technical University of Košice, Slovakia
Nebojša Stojčić, University of Dubrovnik, Croatia
Vjeran Strahonja, University of Zagreb, Croatia
Diana Šimić, University of Zagreb, Croatia
Katarina Tomičić-Pupek, University of Zagreb, Croatia
Jan Tretmans, Radboud University Nijmegen, The Netherlands
Wim Van Petegem, KU Lueven, Belgium
Violeta Vidaček Hainš, University of Zagreb, Croatia
Sirje Virkus, Tallinn University, Estonia
Mladen Vouk, North Carolina State University, USA
Ksenija Vuković, University of Zagreb, Croatia

Honorary Members - Former Program Committee Chairs

Boris Aurer, University of Zagreb, Croatia
Tihomir Hunjak, University of Zagreb, Croatia
Mirko Maleković, University of Zagreb, Croatia

CECIIS 2023 Organizing Committee

Petra Grd, Organizing Committee Chair, University of Zagreb, Croatia

Ena Barčić, University of Zagreb, Croatia
Ivana Fojs, University of Zagreb, Croatia
Vika Kerekeš, University of Zagreb, Croatia
Valentina Kirinić, University of Zagreb, Croatia
Marina Klačmer Čalopa, University of Zagreb, Croatia
Amalija Koren Cavaliero, University of Zagreb, Croatia
Bernarda Kos, University of Zagreb, Croatia
Karlo Nadoveza, University of Zagreb, Croatia
Izabela Oletić Tušek, University of Zagreb, Croatia
Katarina Pažur Aničić, University of Zagreb, Croatia
Laura Strupar, University of Zagreb, Croatia
Matija Šajn, University of Zagreb, Croatia

CECIIS 2023 Research Tracks' Chairs

Computer Games: Markus Schatten, University of Zagreb, Croatia

Data and Knowledge Bases: Sandra Lovrenčić, University of Zagreb, Croatia and Patrizia Poščić, University of Rijeka, Croatia

Digital Human: Neven Vrček, University of Zagreb, Croatia

Digital Transformation and Corporate Culture: Anna Czarczyńska, Kozminski University, Poland and Katarina Tomičić Pupek, University of Zagreb, Croatia

Education and Learning Analytics: Blaženka Divjak, University of Zagreb, Croatia, Igor Balaban, University of Zagreb, Croatia

ICT Entrepreneurship and Innovation: Ksenija Vuković, University of Zagreb, Croatia, Marina Klačmer Čalopa University of Zagreb, Croatia and Nebojša Stojčić, University of Dubrovnik, Croatia

Intelligent Energy: Danijel Topić and Marinko Barukčić, University of Osijek, Croatia

Intelligent Systems: Jasminka Dobša, University of Zagreb, Croatia and Dunja Mladenić, Jožef Stefan Institute, Slovenia

Quality in ICT: Valentina Kirinić and Melita Draganić, University of Zagreb, Croatia

Strategic Planning and Decision Making: Nina Begičević Ređep and Tihomir Hunjak, University of Zagreb, Croatia

Systems for Smart Mobility: Tonči Carić, Edouard Ivanjko, University of Zagreb, Croatia

Trends in Software Engineering: Zlatko Stapić, University of Zagreb, Croatia, William Steingartner, Technical University of Košice, Slovakia and Luis de Marcos Ortega, University of Alcalá, Spain

CECIIS 2023 Poster Section Chairs

Neven Vrček, University of Zagreb, Croatia

Petra Grd, University of Zagreb, Croatia

CECIIS 2023 Student Poster and Presentation Section Chairs

Violeta Vidaček Hainš, University of Zagreb, Croatia

Victoria Appatova, University of Cincinnati, USA

ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)