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

35th International Scientific Conference

CECIIS

Organized by:
University of Zagreb
Faculty of Organization and Informatics

September 18-20/2024
Varaždin, Croatia

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

UDC 004:007

35th International Scientific Conference

# Central European Conference on Information and Intelligent Systems

CECIIS

Proceedings

September 18th – 20th, 2024
Varaždin, Croatia

Organized by:
University of Zagreb
Faculty of Organization and Informatics

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

**For Publisher**
Marina Klačmer Čalopa, University of Zagreb Faculty of Organization and Informatics, Croatia

**Editors**
Neven Vrček, University of Zagreb Faculty of Organization and Informatics, Croatia
Dina Korent, University of Zagreb Faculty of Organization and Informatics, Croatia

**Contact**
E-mail: ceciis@foi.unizg.hr
Phone: +385 42 390 835
Web: ceciis.foi.hr
Fb: facebook.com/ceciis.foi

**Editorial Board**
Igor Balaban, University of Zagreb, Croatia
Nina Begičević Ređep, University of Zagreb, Croatia
Marko Bohanec, Jožef Stefan Institute, Slovenia
Emer. Leo Budin, University of Zagreb, Croatia
Tonči Carić, University of Zagreb, Croatia
Carlos Carrascosa, Universitat Politècnica de València, Spain
Aruna Chandra, Indiana State University, USA
Luis de Marcos, 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 Gašpar, 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
Marina Klačmer Čalopa, University of Zagreb, Croatia
Michal Kohani, University of Zilina, Slovakia
Mario Konecki, University of Zagreb, Croatia
Vladimir Kovšca, 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
Peter Marton, University of Žilina, Slovakia
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, CAMPUS 02 - University of Applied Sciences, Austria
Emer. 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
Marko Tadić, University of Zagreb, Croatia
Martina Tomičić Furjan, University of Zagreb, Croatia
Katarina Tomičić Pupek, University of Zagreb, Croatia
Jan Tretmans, Radboud University Nijmegen, 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

**Publishing Board**
Marko Matus, University of Zagreb Faculty of Organization and Informatics, Croatia
Jelena Gusić Munđar, University of Zagreb Faculty of Organization and Informatics, Croatia
Lea Masnec, University of Zagreb Faculty of Organization and Informatics, Croatia
Bernarda Kos, University of Zagreb Faculty of Organization and Informatics, Croatia

**The 35th Central Conference on Information and Intelligent Systems is held under the patronage of the President of the Republic of Croatia Zoran Milanović.**

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 | V |
| List of Reviewers | VII |
| Preface | IX |


## Computer Games


| | |
|---|---|
| Bogdan Okreša Đurić, Pasqual Martí, Jaume Jordan, Tomislav Peharda, Vicente Julian<br>**Towards Identifying Social Gamification Techniques Applicable to Groups, Applied to On-Demand Shared Transportation** | 3 |
| Bogdan Okreša Đurić, Ana Kutnjak, Larisa Hrustek, Jože M. Rožanec, Francisco Enguix<br>**Towards a Gamification-Enhanced Agent-Based Simulations for the Digital Transformation Process** | 11 |
| Josmario Albuquerque, Bart Rienties, Sebastian Hähnel, Kristina Cordero<br>**Understanding engagement patterns of game-based learning for children in CatnClever using learning analytics** | 19 |
| Sandra Sačarić, Markus Schatten<br>**Integrating Cognitive Agents into Visual Novels** | 27 |
| Nikola Vlahović, Jasmina Pivar, Ivan Celić<br>**Public Funding Approaches to Video Game Industry Development in EU Member States** | 33 |
| Antonela Devčić, Mihaela Sabol<br>**Exploring Hypersexualization in Video Games: A Literature Review of Female Characters** | 41 |


## Cyber-Physical and Embedded Systems


| | |
|---|---|
| Ivan Sládek, Gabriel Gašpar, Juraj Repka, Roman Budjač<br>**Stereo Vision Based 3D Positioning for Real-Time Control Applications** | 53 |
| Juraj Ďuďák, Jakub Perička, Martin Skovajsa<br>**A Lightweight communication protocol for Body Area Networks** | 61 |
| Roman Budjač, Gabriel Gašpar, Maros Valasek, Martin Barton<br>**Proposal for a weather monitoring system utilizing IoT technologies** | 67 |
| Roman Ruzarovsky, Richard Skýpala, Nikolas Blahušiak, Ján Šido, Martin Csekei, Tibor Horák, Peter Střelec<br>**Virtual Commissioning of Industrial Robot through Software-in-the-Loop Strategies for Cyber-Physical Systems** | 75 |
| Marija Kuštelega, Ahmed Shareef, Renata Mekovec<br>**The Role of Digital Twins in Future Skill Development** | 83 |
| Snježana Križanić, Neven Vrček, Matija Habuš<br>**From Classroom to Real-World:**<br>**University Students' Experience in Learning Solidity and Blockchain** | 91 |
| Matija Habuš, Neven Vrček, Snježana Križanić<br>**Optimization of urban road network and digital twins: a literature review** | 99 |
| Goran FIlinić, Ivan Magdalenić<br>**Agile Software Development Life Cycle Model for Internet of Things** | 107 |

## Digital Transformation and Corporate Culture


Maja Fatiga, Andrej Ignjatić
**Digitalization of Monitoring Side Effects Before, During and After Covid-19 Vaccines – Analysis of the Impact of the Pandemic on Reporting Side Effects in Republic of Croatia** 115

Robertina Zdjelar, Nikolina Žajdela Hrustek
**Public Policymakers' Attitudes on Planning and Implementing Measures for 54+ Population e-Inclusion** 123

Mirona Ana Maria Popescu, Andreea Barbu, Larisa Maria Fecioru, Elena-Teodora Miron
**An Analysis on Digital Public Services in Romania** 133


## Innovation, Entrepreneurship and Economic Growth


Marijana Bubanić, Dina Korent
**Profitability of IT companies in the Republic of Croatia in (non)crisis periods: Skillings-Mack test** 143

Leon Šaško, Ivana Fojs, Kristina Detelj
**The rise of the gaming entrepreneurship – overview of the gaming industry ecosystem, case of Croatia** 157

Zrinka Lacković Vincek, Suzana Keglević Kozjak, Vladimir Kovšca
**More than twenty years of the eXtensible Business Reporting Language** 165

Tamara Šmaguc, Marina Klačmer Čalopa, Mirjana Grčić Fabić
**Phenomenological Inquiry in Entrepreneurship and Small Business with Insights into the Digital Economy Research Context** 175

Iva Gregurec, Magdalena Kuštelega, Goran Vlašić
**Exploring the trend of scientific research of eye-tracking for digital advertising insight – A Bibliometric Analysis** 183


## Intelligent Energy


Zvonimir Šimić, Danijel Topić, Marina Dubravac, Hrvoje Kurkutović
**Using of Battery Storage to Match the Supply and Demand of Electricity in Smart Grids** 193

Jakub Perička, Tomáš Meravý, Juraj Ďuďák
**Sustainable Microgrid Networks: Advances and Applications** 199

## Intelligent Systems


| | |
|---|---|
| Jože M. Rožanec, Dunja Mladenić<br>**WaAUROCC: measuring how steep the ROC AUC curve is** | 209 |
| Barbara Kovačić, Anja Brnjaković, Gaurish Thakkar<br>**Beyond Words: Sentiment Analysis of Croatian Language Attitudes** | 217 |
| Gaurish Thakkar, Vanja Štefanec, Daša Farkaš, Marko Tadić<br>**Building a Large Language Model for Moderately Resourced Language: A Case of Croatian** | 225 |
| Oliver Lohaj, Ján Paralič, Ruslana Buinytska<br>**Exploring usability in combining data sources for detecting toxic behavior in social media** | 233 |
| Ana Meštrović<br>**A Review of Methods for Analyzing Online Media Texts Using Large Language Models** | 241 |
| Lucia Načinović Prskalo<br>**Text Analysis Methods for Identifying and Recognizing Gender Bias, Inequality, and Violence in Language: An Overview**<br>**Metode analize teksta za identifikaciju i prepoznavanje rodne pristranosti, nejednakosti i nasilja u jeziku: Pregled** | 257 |
| Dino Vlahek, Dijana Oreški<br>**Data Spaces in the Republic of Croatia: Potential and Opportunities** | 277 |
| Ivana Ružić, Igor Balaban<br>**Social Robots: Applications, Challenges, and Future Directions - literature review** | 283 |


## Learning Analytics, AI and Technologies in Education


| | |
|---|---|
| Martina Holenko Dlab, Gordan Đurović, Marko Horvat<br>**Perception and Use of GenAI Chatbots in Education by Students at the University of Rijeka** | 293 |
| Patrik Komljenović, Mia Čarapina<br>**Tablet Math Application for Supporting Independent Simultaneous Interaction of Colocated Users** | 303 |
| Sonja Nidogon Višnjić<br>**The relationship between student work and academic success: a case study** | 311 |
| Marko Matus, Igor Balaban<br>**Relationship between factors that contribute to student success in the e-learning environment** | 319 |
| Barbi Svetec, Blaženka Divjak, Damir Horvat, Katarina Pažur Aničić<br>**Predicting the Outcomes of PBL: Comparison of Two Methods** | 325 |
| Blaženka Divjak, Barbi Svetec, Bart Rienties, Paul Astles, Catriona Matthews, Alexander Tillmann, Michael Eichhorn, Julia Schmitt<br>**Prioritizing Learning Outcomes in Different Learning Design Contexts** | 333 |
| Bojan Radišić, Ivan Dunđer, Sanja Seljan<br>**Predictive Factors for Completion or Dropping Out of Professional Studies** | 341 |
| Matija Novak, Marija Maček<br>**Student perspectives on using Artificial Intelligence tools for solving programming assignments**<br>**Perspektiva studenata o korištenju alata umjetne inteligencije za rješavanje programskih zadataka** | 351 |

## Strategic Planning and Decision Making


| | |
|---|---|
| Nikola Kadoić<br>**Understanding the Impact of the Voting Model at the Eurovision Song Contest: Why Croatian Baby Lasagna did not win the ESC 2024?** | 367 |
| Jelena Gusić Munđar, Tena Jagačić, Nikola Kadoić<br>**Bridging Theory and Practice: Enhancing Business Decision-Making Course through Project-Based and Work-Based Learning** | 377 |
| Maja Pušnik, Katja Kous, Boštjan Šumak<br>**Enhancing Strategic planning course with operational research methods: a case study for IT students** | 387 |


## Trends in Software Engineering


| | |
|---|---|
| Víctor-Pablo Prado-Sánchez, Adrián Domínguez-Díaz, Daniel Rodríguez, José-Javier Martínez-Herráiz, Luis de-Marcos<br>**How can ChatGPT help humans in Dark Web content classification? Assessing GPT models reliability and effects of explanations on human decisions** | 395 |
| Juraj Dončević, Kristina Ćaleta, Marko Horvat, Ivica Botički<br>**Mine Action Data Observatory: Tackling Uncertainty and System Longevity with Evolutionary Architecture** | 403 |
| Miro Zdilar<br>**Attribute Based Access Control Metamodel for Spreadsheet Programs** | 409 |
| Sara Kijanović, Darko Stefanović, Dušanka Dakić, Teodora Vučković<br>**Using AI on Open Data: A Systematic Literature Review** | 417 |
| Michaela Pauriková, Michal Varga, Norbert Adamko<br>**Reevaluation of intentions of dynamic agent utilizing BDI reasoning in ABAsim architecture** | 425 |
| Lea Masnec, Zlatko Stapic, Maja Benkus<br>**Bridging Academia and Industry: A Collaborative Approach to Software Development Education** | 433 |

| | |
|---|---|
| Publication Ethics And Malpractice Statement | 441 |
| Author Guidelines | 445 |
| CECIIS 2024 Sponsors | 459 |
| CECIIS 2024 Program and Organizing Committees' Members, Research, Professional Tracks' Chairs and Student Poster and Presentation Section Chair | 465 |

# Index of Authors


Adamko, Norbert 419
Albuquerque, Josmario 19
Astles, Paul 333
Balaban, Igor 283, 319
Barbu, Andreea 133
Bartoň, Martin 67
Benkus, Maja 433
Blahušiak, Nikolas 75
Botički, Ivica 403
Brnjaković, Anja 217
Bubanić, Marijana 143
Budjač, Roman 53, 67
Buinytska, Ruslana 233
Celić, Ivan 33
Cordero, Kristina 19
Csekei, Martin 75
Čarapina, Mia 303
Ćaleta, Kristina 403
Dakić, Dušanka 417
de Marcos, Luis 395
Detelj, Kristina 157
Devčić, Antonela 41
Divjak, Blaženka 325, 333
Domínguez-Díaz, Adrián 395
Dončević, Juraj 403
Dubravac, Marina 193
Ďuďák, Juraj 61, 199
Dunđer, Ivan 341
Đurović, Gordan 293
Eichhorn, Michael 333
Enguix, Francisco 11
Farkaš, Daša 225
Fatiga, Maja 115
Fecioru, Larisa Maria 133
Filinić, Goran 107
Fojs, Ivana 157
Gašpar, Gabriel 53, 67
Grčić Fabić, Mirjana 175
Gregurec, Iva 183
Gusić Munđar, Jelena 377
Habuš, Matija 91, 99
Hähnel, Sebastian 19
Holenko Dlab, Martina 293
Horák, Tibor 75
Horvat, Damir 325
Horvat, Marko 293, 403
Hrustek, Larisa 11
Ignjatić, Andrej 115
Jagačić, Tena 377
Jordan, Jaume 3
Julian, Vicente 3
Kadoić, Nikola 367, 377
Keglević Kozjak, Suzana 165
Kijanović, Sara 417
Klačmer Čalopa, Marina 175
Komljenović, Patrik 303
Korent, Dina 143
Kous, Katja 387
Kovačić, Barbara 217
Kovšca, Vladimir 165
Križanić, Snježana 91, 99
Kurkutović, Hrvoje 193
Kuštelega, Magdalena 183
Kuštelega, Marija 83
Kutnjak, Ana 11
Lacković Vincek, Zrinka 165
Lohaj, Oliver 233
Maček, Marija 351
Magdalenić, Ivan 107
Martí, Pasqual 3
Martínez-Herráiz, José-Javier 395
Masnec, Lea 433
Matthews, Catriona 333
Matus, Marko 319
Mekovec, Renata 83
Meravý, Tomáš 199
Meštrović, Ana 241
Miron, Elena-Teodora 133
Mladenić, Dunja 209
Načinović Prskalo, Lucia 257
Nidogon Višnjić, Sonja 311
Novak, Matija 351
Okreša Đurić, Bogdan 3, 11
Oreški, Dijana 277
Paralič, Ján 233
Pauriková, Michaela 425
Pažur Aničić, Katarina 325
Peharda, Tomislav 3
Perička, Jakub 61, 199
Pivar, Jasmina 33
Popescu, Mirona Ana Maria 133
Prado Sánchez, Víctor-Pablo 395
Pušnik, Maja 387
Radišić, Bojan 341
Repka, Juraj 53
Rienties, Bart 19, 333
Rodríguez, Daniel 395
Rožanec, Jože M. 11, 209
Ružarovský, Roman 75
Ružić, Ivana 283
Sabol, Mihaela 41
Sačarić, Sandra 27
Schatten, Markus 3, 27
Schmitt, Julia 333
Seljan, Sanja 341
Shareef, Ahmed 83
Skovajsa, Martin 61
Skýpala, Richard 75
Sládek, Ivan 53
Stapić, Zlatko 433
Stefanović, Darko 417
Střelec, Peter 75
Svetec, Barbi 325, 333
Šaško, Leon 157
Šido, Ján 75
Šimić, Zvonimir 193
Šmaguc, Tamara 17 5
Štefanec, Vanja 22 5
Šumak, Boštjan 387
Tadić, Marko 225
Thakkar, Gaurish 217, 225
Tillmann, Alexander 333
Topić, Danijel 193
Valášek, Maroš 67
Varga, Michal 425
Vlahek, Dino 277
Vlahović, Nikola 33
Vlašić, Goran 183
Vrček, Neven 91, 99
Vučković, Teodora 417
Zdilar, Miro 409
Zdjelar, Robertina 123
Žajdela Hrustek, Nikolina 123

# List of Reviewers

Barać, Dušan, Serbia
Barreto, Nuno Vicente, Great Britain
Beliga, Slobodan, Croatia
Bidlo, Michal, Czech Republic
Bohanec, Marko, Slovenia
Bubanić, Marijana, Croatia
Buchem, Ilone, Germany
Castañeda Quintero, Linda Johanna, Spain
Ćukušić, Maja, Croatia
Cvetokoska, Violeta, Nort Macedonia
Davalas, Athanasios, Greece
de-Marcos, Luis, Spain
Dobrović, Željko, Croatia
Domínguez-Díaz, Adrián, Spain
Dudáš, Adam, Slovakia
Fertalj, Krešimir, Croatia
Fodor, Lidija, Serbia
Grd, Petra, Croatia
Gregurec, Iva, Croatia
Guajardo, Ashley ML, United States of America
Guardia Ortiz, Lourdes, Spain
Hajdin, Goran, Croatia
Hernández, Luis, Spain
Horvat, Marko, Croatia
Hrgarek Lechner, Nadica, Croatia
Ivković, Nikola, Croatia
Jagušt, Tomislav, Croatia
Kadoić, Nikola, Croatia
Keglević Kozjak, Suzana, Croatia
Knežević, Goran, Croatia
Konecki, Mladen, Croatia
Krašna, Marjan, Slovenia
Krhač Andrašec, Eva, Slovenia
Kučina Softić, Sandra, Croatia
Kumar Cherukuri, Aswani , India
Kvet, Michal, Slovakia
Lacković Vincek, Zrinka, Croatia
Maclaren, Iain, Ireland
Magdalenić, Ivan, Croatia
Maina, Marcelo, Spain
Mangiron, Carme, Spain
Martinčić - Ipšić, Sanda, Croatia
Mekovec, Renata, Croatia
Meštrović, Ana, Croatia
Mijač, Tea, Croatia
Milić, Luka, Croatia
Milić, Miloš, Serbia
Nikolić, Dejana, Croatia
Novak, Matija, Croatia
Novotný, Samuel, Slovakia
Nyarko, Emmanuel Karlo, Croatia
Okreša Đurić, Bogdan, Croatia
Oreški, Dijana, Croatia
Paralič, Jan, Slovakia
Pataki, Norbert, Hungary
Pavlin, Samo, Slovenia
Pažur Aničić, Katarina, Croatia
Perković, Sonja, Denmark
Pihir, Igor, Croatia
Poščić, Patrizia, Croatia
Prunera, Jaume Jordan, Spain
Pušnik, Maja, Slovenia
Radaković, Davorka, Serbia
Radošević, Danijel, Croatia
Radovanović, Sandro, Serbia
Rakun, Jurij, Slovenia
Ristić, Sonja, Serbia
Rodič, Blaž, Slovenia
Rodríguez García, Daniel, Spain
Rožanec, Jože M., Slovenia
Schatten, Markus, Croatia
Seme, Sebastijan, Slovenia
Siedlecka-Lamch, Olga, Poland
Sittar, Abdul , Slovenia
Stefanović, Darko, Serbia
Steingartner, William, Slovakia
Straka, Milan, Slovakia
Sturm, Jan, Slovenia
Szymoniak, Sabina, Poland
Šarac, Vasilija, North Macedonia
Tokarčíková, Emese, Slovakia
Tomaš, Boris, Ireland
Tomičić, Igor, Croatia
Topal, Oleksandra, Slovenia
Umek, Lan, Slovenia
Van Petegem, Wim, Belgium
Varga, Michal, Slovakia
Virkus, Sirje, Estonia
Vrček, Neven, Croatia

# Preface

We are pleased to present you with the Proceedings of the 2024 Central European Conference on Information and Intelligent Systems. 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 this year's edition of 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**. For over three decades, CECIIS has successfully convened with the aim of serving as an interface between scholars and professionals in the field of information and intelligent systems and related domains. This year CECIIS took place from **September 18th to 20th, 2024,** in Varaždin, Croatia, and was held under the **high patronage of the President of the Republic of Croatia, Mr. Zoran Milanović**. The official language of the conference is English.

With the thematic focus of "**Cyber Physical and Embedded Systems**" CECIIS 2024 was designed to facilitate scholarly exploration and discourse on the transformative opportunities and challenges these systems present to the economy and the society. Considering the role these systems are playing in reshaping technology and engineering, it is crucial to explore the synergy between hardware and software in this fascinating emerging landscape. At our conference, we discussed how these systems are revolutionizing industries and the society in general. The exploration of cutting-edge topics like AI integration, IoT advancements, and security challenges in embedded systems offered a unique glimpse into the future where the technology is seamlessly integrated with the physical world. A number of workshops was also held on various aspects of the interaction between modern society and technology:

- WORKSHOP: SMART HOME LIVING, Slovak University of Technology, Faculty of Materials Science and Technology in Trnava;
- WORKSHOP: AN EXAMPLE OF BUILDING AND RUNNING THE RAG (RETRIEVAL AUGMENTED GENERATION) APPLICATION WITH ORACLE AI VECTOR SEARCH, Oracle Company;
- HELA WORKSHOP: AGILITY IN PLANNING AND DECISION-MAKING IN HE – Project: Improving HEI Maturity to Implement Learning Analytics;
- CUTIE WORKSHOP: HOW TO IMPROVE TEACHERS' DIGITAL SKILLS IN CO-CREATION WITH STUDENTS? – Project: Competences for Universities using Technology in Education and Institutional Empowerment;
- TRUELA WORKSHOP: RESEARCH WORKSHOP ON TRUSTWORTHY LEARNING ANALYTICS AND AI ALGORITHMS – Project: Trustworthy Learning Analytics and Artificial Intelligence for Sound Learning Design;
- DECIDE WORKSHOP: WORKSHOP ON DECIDE METHODOLOGY – Project: Digital Services for the Circular Economy.

Admittedly, a single conference cannot address all the compelling questions surrounding the use of information technology in modern society. However, we are cultivating a community that is not merely observing the future but actively shaping it through **innovative research, interdisciplinary collaboration, and visionary thinking**. In this way, we remain at the forefront of a technological revolution that is steering the world toward a smarter, more connected future.

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 whose collective insights contribute a substantial value to the fabric of CECIIS.

The Proceedings of CECIIS 2024 encompass 49 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, Cyber-Physical and Embedded Systems, Digital Transformation and Corporate Culture, Innovation, Entrepreneurship and Economic Growth, Intelligent Energy, Intelligent Systems, Learning Analytics, AI and Technologies in Education, Strategic Planning and Decision Making, 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 was conducted with the dedicated efforts of **89 reviewers**, 36 from Croatia and 53 from abroad. Collectively, they completed 108 reviews, leading to the acceptance of 49 papers. These papers, authored by 133 individuals representing **10 nations**, were presented at the conference. Following the event, a selected group of papers will undergo further review, potentially leading to extended publication in the **Journal of Organizational and Information Sciences (JIOS),** the international scientific journal published by the Faculty of Organization and Informatics.

A pivotal hallmark of this year's conference were the distinguished invited **keynote speakers**, individuals whose intellectual contributions have shaped their respective spheres of scientific and professional inquiry. CECIIS 2024 was honored to host the following keynote speakers:

Professor Srđan Čapkun, Ph.D., who addressed the topic "Physical-layer attacks and their impact on wireless networks: two case studies". Dr. Srđan Čapkun is a full professor in the Department of Computer Science, ETH Zurich, and the Chair of the Zurich Information Security and Privacy Center (ZISC).

Professor Hanni Muukkonen, Ph.D., who presented the topic "Stakeholder perspectives of learning analytics and AI development and implementation in education". Dr. Hanni Muukkonen is a Professor in Educational Psychology at the Faculty of Education and Psychology, University of Oulu, Finland. Her research focuses on collaborative learning, knowledge creation, learning analytics, and ethics in AI.

Professor Bart Rienties, Ph.D., who presented the topic "Why culture matters in Information and Intelligent Systems". Dr. Bart Rienties is a Professor of Learning Analytics and Programme Lead of the learning analytics and learning design research programme at the Institute of Educational Technology at the Open University, UK.

In addition to the scholarly sessions, the conference featured two sections dedicated to posters and presentations, which included **2 scientific research posters and 18 student posters and presentations**, collectively involving 39 authors from 7 nations. These sections added depth and diversity to the intel-lectual landscape of the conference. The student section was further enriched by an invited lecture de-livered by Professor Machdel Matthee from the University of Pretoria on the topic of "Digital disruption, digital innovation, and 21st century skills".

A particularly engaging component of the program was the panel discussion entitled "**It's time for digital growth**", moderated by Professor Neven Vrček, Ph.D., from the University of Zagreb Faculty of Organization and Informatics. Esteemed panelists shared their insights on the application of **Industry 4.0 technologies** to stimulate economic growth. The topics covered included **novel economic growth opportunities**, the role of **digital technologies,** rising **security risks in the digital economy, sustainability and environmental challenges**, the importance of **strategic focus at the country level**, and how the **scientific community** can contribute to advancing processes related to digital growth.

We extend our sincere thanks to all contributors to the Conference program and these Proceedings. This encompasses our esteemed keynote speakers, panelists, authors, reviewers, session chairs, workshop organizers 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 conveyed to our respected academic and industrial partners, sponsors and donors, whose unwavering support made CECIIS 2024 a resounding achievement. Collaboration with the University of Zagreb Faculty of Transport and Traffic Sciences, University of Zagreb Faculty of Electrical Engineering and Computing, Josip Juraj Strossmayer University of Osijek Faculty of Electrical Engineering, Computer Science and Information Technology, and University of Rijeka Faculty of Informatics and Digital Technologies was central to the realization of the conference. The financial support from sponsors and donors including Oracle d.o.o., Infodom d.o.o., Grad Novi Marof, Hrvatska banka za obnovu i razvitak, Hrvatska elektroprivreda d.d., Grad Varaždin, Varaždinska županija, AGGERO d.o.o., Factory X d.o.o., Jadranski naftovod d.d., Termoplin d.d. Varaždin, Grad Čakovec, Stolarija i pilana Ratković, and Hidroing d.o.o. was also invaluable.

In conclusion, it is our sincerest hope that CECIIS 2024 provided the participants with enriching and productive experience, and that they will find relevant and interesting papers encompassed in these Proceedings. 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 2025.

September 2024

On the behalf of the Program and Organizing Committees

Neven Vrček and Dina Korent

Editors

# Computer Games

Bogdan Okreša Đurić, Pasqual Martí, Jaume Jordan, Tomislav Peharda, Vicente Julian
**Towards Identifying Social Gamification Techniques Applicable to Groups, Applied to On-Demand Shared Transportation**

Bogdan Okreša Đurić, Ana Kutnjak, Larisa Hrustek, Jože M. Rožanec, Francisco Enguix
**Towards a Gamification-Enhanced Agent-Based Simulations for the Digital Transformation Process**

Josmario Albuquerque, Bart Rienties, Sebastian Hähnel, Kristina Cordero
**Understanding engagement patterns of game-based learning for children in CatnClever using learning analytics**

Sandra Sačarić, Markus Schatten
**Integrating Cognitive Agents into Visual Novels**

Nikola Vlahović, Jasmina Pivar, Ivan Celić
**Public Funding Approaches to Video Game Industry Development in EU Member States**

Antonela Devčić, Mihaela Sabol
**Exploring Hypersexualization in Video Games: A Literature Review of Female Characters**

# Towards Identifying Social Gamification Techniques Applicable to Groups, Applied to On-Demand Shared Transportation

**Bogdan Okreša Đurić, Markus Schatten, Tomislav Peharda**
Artificial Intelligence Laboratory
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{dokresa,markus.schatten,tomislav.peharda}@foi.unizg.hr

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

**Vicente Julian**
Valencian Graduate School and Research Network of Artificial Intelligence
Universitat Politècnica de València
Camino de Vera, s/n 46022 Valencia. Edificio 1F
vjulian@upv.es

**Abstract.** *The paper explores applying gamification techniques to improve on-demand shared transportation systems by influencing group decision-making processes. Gamification, typically applied to individual users, is extended to scenarios involving multiple users to optimise resource usage in transportation. The study identifies key gamification elements such as avatars, badges, leaderboards, and challenges and examines their effectiveness in fostering user cooperation and competition. The research suggests combining cooperative and competitive gamification strategies can enhance user engagement, motivation, and overall system efficiency.*

**Keywords.** gamification, group dynamics, transportation, cooperation, competition

## 1 Introduction

Gamification is described as applying game-specific techniques and methods to non-game contexts to increase participants' engagement and motivation levels.

One of the fields of gamification applications that could benefit considerably from gamification is the transport of people. The transport systems offer travel to groups of users according to their needs. Those that are more flexible, such as a taxi or transport service that can be adapted to demand, adapt their operation to the travel requests of users. The introduction of gamification techniques in this area seeks to motivate users to make 'better use' of the transport system. This best use refers to a travel demand that is properly adapted to the available resources of the transport system at any given time. Consequently, this is achieved by motivating users to change the location at which their journey begins or ends and the time at which they wish to travel whenever these attributes represent conflicts with system resources.

Gamification techniques applied to shared transportation present a complex challenge. The scroll request attributes that should be modified now depend on a set of users instead of a single one. For the modification to be successful, all users in the group must reach the same decision. In this paper, we explore gamification techniques applied to groups of people based on the example of shared transportation. Next, we explore the different situations that occur during group decision-making and discuss the problems that could entail applying gamification techniques.

## 2 Related Work

This section provides topic progression from general gamification towards social and cooperative gamification, introducing concepts relevant to this paper.

### 2.1 Defining Gamification

As mentioned earlier, the generally accepted definition of gamification concerns using game-specific mechan-

ics, techniques, and methods in non-game contexts. The main goal of this process is usually to boost the users' engagement and motivation levels. Some of the earliest use of the concept of gamification as we know it can be traced to 2008 (Dar et al., 2022; Deterding, Dixon, Khaled and L. E. Nacke, 2011). Since the generally accepted definition mentioned earlier is extremely adaptable, some authors, such as Edwards et al. (2016), constrain their research by providing a more specific definition. For example, Edwards et al. (2016) consider a system to be gamified if it utilises at least one of the techniques in a set of recognised fundamental gamification techniques that comprises the following: avatars, badges, challenges, competitions, leaderboards, prizes, and rewards. The fundamental concepts often related to gamification are points, badges, and leaderboards. More advanced concepts that can be included in a gamified system follow the principles of game design (Manzano-León et al., 2021).

Various authors propose different key building blocks of a game design process. Schell (2019) describes games using the four basic elements of games: 1. mechanics (procedures and rules of a game), 2. story (the sequence of events that progresses the game), 3. aesthetics (the way the game is designed, caters to all of the human senses), 4. technology (how the game is implemented, e.g. pen and paper, computer system, etc.). One of the most studied systems of game design elements, according to Manzano-León et al. (2021), the MDE system (Hunicke et al., 2004; Yamani, 2021), consists of the following elements: 1. mechanics, 2. dynamics, 3. aesthetics. Mechanics encompass rules, actions, and control mechanisms made accessible to the player. Dynamics comprises the behaviours that stem from mechanics and depend on them and their execution. Aesthetics is the concept that elicits the wanted emotional response and works on evoking it in the player.

It should be noted here that gamification uses the concepts of games, but the system using such concepts is not necessarily a game. Furthermore, learning supported by gamification does not necessarily entail game-based learning, according to Yamani (2021). The main goal of a game is to elicit a positive emotional response and to have fun within the constraints of a given game's collection of rules (Avedon and Sutton-Smith, 1971; Schell, 2019; Yamani, 2021). Furthermore, taking part in a game is voluntary. On the other hand, gamification focuses on motivating users to engage with the contents of the gamified system without necessarily transforming the content itself, making the interaction easier and less expensive than developing full-fledged educational games. Game-based learning and gamification aim to enhance the experience of interaction, for example, the experience of learning. Ultimately, while game-based learning makes the learning process a core part of the gameplay, gamification enhances existing educational tasks by integrating game mechanics.

Having fun while interacting with different activities and enjoying the process is something people often like (Dindar et al., 2021). Moreover, a number of studies have found that indulging in games and related concepts can positively influence prosocial behaviour in players (Riar, Morschheuser, Hamari and Zarnekow, 2020). Applied gamification mechanics are shown to induce intrinsic and extrinsic motivation in users (Feng et al., 2022). Therefore, it is unsurprising that gamification is becoming increasingly important (Fonseca et al., 2023). However, even though gamification has often been shown to foster motivation, boost engagement, and, for example, improve learning, there are challenges to the use and implementation of gamification techniques in various contexts, such as the long-term viability of the effects of gamification to the users (Sardi et al., 2017). Furthermore, the published research on gamification lacks a generalised approach, as it is often observed in a domain-specific context (Dindar et al., 2021). Another point to emphasise is the lack, and impossible existence, of a definitive list of possible gamification elements because choosing those is a process akin to game design, which is, by definition, experimental (Bakhanova et al., 2023).

### 2.1.1 Gamification Design Frameworks

The above briefly described MDE model can be considered a methodology or a framework that is used in the process of designing video games. The main goal of using that particular methodology is to ease establishing connections between (Manzano-León et al., 2021): rules–mechanics, game–dynamics, and fun–aesthetics. A similar approach can be applied to gamification systems. Mechanics can be translated into gamification concepts such as trading, competing, collecting cards, etc. Dynamics can be observed in socialising and status marking; aesthetics can be recognised in e.g. the narrative, challenges, or entertainment.

The MDE model is not the only gamification framework that has been published recently. In fact, many gamification design frameworks can be found in the published literature, as presented by Mora et al. (2017) in their overview of 40 such frameworks that were published up to October 2015. Two more gamification design frameworks are presented in this paper: 1. the toolkit of gamification elements presented by Werbach and Hunter (2015) and 2. the Octalysis gamification framework by Chou (2015).

The gamification design framework of Werbach and Hunter (2015) is based on the hierarchy of three fundamental game elements: 1. dynamics, 2. mechanics, 3. components. The concepts of dynamics provide motivation to the user and can be observed in the available in-game choices, social interaction, or ways of eliciting specific emotional responses from players. Dynamics can be observed as being a part of one of the follow-

ing categories: constraints, emotions, narrative, progression, and relationships. Akin to the MDE model described above, mechanics are the concepts that include the player and drive their involvement, generate engagement and motivate them to move forward in the game, such as turns, rewards, or types of relationship implementations. Mechanics can be understood as instances of the concepts in dynamics and as 'verbs of a game.' (Werbach and Hunter, 2015) Components comprise the specific actionable elements and objects that implement concepts of dynamics and mechanics, e.g. quests that progress narrative or force social interaction, points that implement one of the aspects of rewards, etc.

The Octalysis gamification design framework, developed and presented by Chou (2015), is a comprehensive model for analysing and designing gamified experiences based on human motivation. This framework's approach to designing gamification differs from the two described above. It identifies the following eight core drives that influence human behaviour and engagement: 1. epic meaning and calling 2. development and accomplishment 3. empowerment of creativity and feedback 4. ownership and possession 5. social influence and relatedness 6. scarcity and impatience 7. unpredictability and curiosity 8. loss and avoidance. By understanding and leveraging these drives, Chou (2015) argues that designers can create experiences that motivate and engage users more effectively. Furthermore, the core drives are divided with more classification bases in mind. For example, some of the core drives are more associated with creativity, and some are more focused on analytic thought and ownership. Furthermore, Chou (2015) classifies some of the core drives as positive and negative regarding how manipulative they are or how good the user feels when they encounter the applicable core drive. The empowerment of creativity and feedback core drive encourages users to express their creativity and solve problems while receiving immediate feedback. An example of this can be found in sandbox games like Minecraft, where players can build and create within the game world while receiving instant feedback on their creations. By incorporating these and other core drives from the Octalysis Framework, designers can craft engaging and motivating experiences tailored to their audience's intrinsic and extrinsic motivations.

## 2.2 Select Application Domains

Gamification has been applied across various domains (Schatten et al., 2017), each leveraging game mechanics to improve user engagement, motivation, and performance. In the requirements elicitation field, gamification involves and engages users, thereby enhancing the process of gathering requirements (Dar et al., 2022). In education, gamification methods aim to foster active learning, which motivates students and strengthens their connection to the lessons (Chi and Wylie, 2014; Dereli and Kahraman, 2024). Specifically, gamified flipped classrooms have been shown to enhance nursing students' skill knowledge and self-confidence during clinical practice. Research consistently indicates that educational gamification positively impacts student motivation, engagement, and academic performance across various educational levels, suggesting it is an effective strategy for improving academic outcomes (Manzano-León et al., 2021; Mora et al., 2017).

In health-related applications, incorporating game design elements like earning points, collecting badges, and interacting with virtual agents has significantly enhanced user experience and motivation, encouraging sustained usage (Priesterroth et al., 2019). The positive effects of gamification also extend to e-learning, where elements such as points, leaderboards, rewards, and challenges enhance student motivation and engagement with classroom content (Saleem et al., 2022). Gamification also plays a crucial role in e-health solutions by ensuring regular user engagement and increasing immersion (Sardi et al., 2017). In collaborative knowledge crowdsourcing, gamification mechanics related to immersion, social interaction, and achievement foster a sense of self-worth and competence among solvers, boosting their participation and contribution (Feng et al., 2022). Additionally, gamification strategies incorporating competition, cooperation, and team dynamics have positively affected students' learning curves (Mohammadi et al., 2023). In the university context, gamification-based cooperative play strategies can enhance students' emotional intelligence, life goals, and learning strategies, thereby increasing academic motivation and reducing dropout rates (Redondo-Rodríguez et al., 2023). Gamification also supports zero-waste practices by encouraging cooperative efforts towards reducing waste (Imani et al., 2021).

## 2.3 Frequently Used Gamification Mechanics

As extensively discussed in the available literature, gamification involves various concepts that enhance user engagement and motivation across different domains. Common gamification forms include badges and leadership boards, frequently used to foster a sense of achievement and competition among users (Dindar et al., 2021; Looyestyn et al., 2017). Moreover, Dar et al. (2022) highlight ten essential ingredients of great games (Deterding, Dixon, Khaled and L. Nacke, 2011), such as avatars, 3D environments, ranks, levels, reputation, feedback, economies, rule-based competition, teams, configurable communication systems, and time pressure, along with other elements like scores, points, badges, awards, and quests. These elements, particularly points, leaderboards, and badges, are frequently

used in requirements elicitation, illustrating their effectiveness in engaging users.

The application of gamification in education and health also showcases the versatility of game mechanics. For instance, Manzano-León et al. (2021) uses the MDE model presented above and describes game mechanics as actions and controls within a game, such as drawing cards or competing, while dynamics refer to behaviours like socialising or reflecting during gameplay. Aesthetics evoke emotional responses in players, enhancing their engagement. Feng et al. (2022) classifies gamification elements into three categories: immersion (e.g., storytelling and avatars), social (e.g., chats and peer ratings), and achievement (e.g., points and badges). These classifications are evident in the diverse applications of gamification methods, including educational escape rooms, board games, and online quizzes in physiotherapy and rehabilitation education (Dereli and Kahraman, 2024). Virtual reality games and gamified websites further illustrate the innovative use of gamification techniques in educational contexts, with leaderboards being particularly motivating as they visualise progress.

The analysed studies consistently report various gamification elements contributing to enhanced user experiences. Feedback, rewards, progress bars, social connections, challenges, and quests are commonly used across applications (Saleem et al., 2022; Sardi et al., 2017). Feedback, in particular, stands out as a widely used technique, often accompanied by parallel communication systems and digital rewards, which can include virtual coins, badges, or other in-game items (Priesterroth et al., 2019). Manzano-León et al. (2021) enumerate a comprehensive list of gamification elements, including challenge, reward, narrative, achievements, points, levels, badges, ranking, mastery, and feedback, underscoring the multifaceted nature of gamification strategies. These elements enhance motivation and engagement and facilitate active learning and behavioural change, proving the efficacy of gamification in various fields.

## 2.4 Social Gamification, Cooperation, and Competition

Social gamification involves transforming individual-based tasks into social events by incorporating elements of interaction and cooperation, thereby creating more immersive experiences for users. Gamification elements are classified into four categories based on social dependence: 1. individual 2. competitive 3. cooperative 4. and competitive-cooperative. Cooperative game elements often include design patterns that encourage teamwork, such as complementarity, synergies between roles, shared goals, and special rules for team members (Imani et al., 2021; Morschheuser et al., 2017). Research indicates that cooperation in gamification fosters better socialisation and psychological health compared to competition. Cooperative settings lead to higher group participation and greater intimacy among group members, making gamified cooperation more effective in developing feelings of belonging and attachment among learners. Both gamified competition and cooperation can sustain task effort and improve learning outcomes, but they engage different psychological processes (Dindar et al., 2021).

Social gamification plays a significant role in creating a sense of belonging by enhancing individual tasks and enabling users to engage with others. It transforms activities such as exercise into social events, fostering a sense of immersion and interaction that is distinct from individualistic gamification. Interaction with cooperative game design features can promote altruism, as players find intrinsic fulfilment and enjoyment in contributing to the welfare of others. Altruistic behaviours are especially prominent within groups where individuals identify with each other, creating a "we"-perspective (Riar, Morschheuser, Hamari and Zarnekow, 2020; Zhang et al., 2023). Cooperative learning further emphasises the importance of working towards shared goals, facilitating participants' support and reinforcing gamification's social aspects (Fonseca et al., 2023).

Designing systems to achieve full cooperation involves addressing individuals' altruistic motivations and group intentions rather than relying solely on individual motives. Social interdependence theory (SIT) explains how social interactions influence people's goals, highlighting positive interdependence in cooperative environments where mutual goals are shared and negative interdependence in competitive settings where one's success depends on others' failure. Individualistic gamification lacks these elements but can be transformed into social gamification by integrating competitive or cooperative mechanisms. Cooperation and interpersonal competition can coexist, but intergroup competition inherently involves cooperation among team members (Zhang et al., 2023).

The concept of coopetition, which combines competition and cooperation, offers a promising approach to designing social gamification (Liu et al., 2017). It opens up new avenues for enhancing engagement and motivation by leveraging the strengths of both competitive and cooperative dynamics (Zhang et al., 2023). Competition is categorised by Zhang et al. (2023) into interpersonal (among individuals) and intergroup (among groups), leading to various forms of coopetition. These include interpersonal coopetition, intergroup coopetition, and hybrid coopetition, each involving different combinations of cooperation and competition. Studies suggest that coopetition can enhance engagement and motivation, offering a fresh perspective on the design of social gamification systems.

The meta-analysis of 148 studies on SIT performed by Roseth et al. (2008) indicates that cooperative goal structures generally outperform competitive ones in

offline tasks (Zhang et al., 2023). This finding supports the effectiveness of cooperative environments in promoting better outcomes. Intergroup coopetition, which integrates cooperation with intergroup competition, has increased game engagement and physical activity, such as daily steps. This approach induces group cohesiveness, leading to more constructive cooperation among teammates and overall enhanced engagement.

Social gamification leverages cooperative and competitive elements to enhance user engagement and motivation. Fostering social interactions and creating a sense of belonging transforms individual tasks into collective activities, promoting psychological well-being and effective learning outcomes. Cooperative game elements, social interdependence, and coopetition strategies are central to designing effective social gamification systems, ultimately leading to higher engagement and better performance (Riar, Morschheuser, Zarnekow and Hamari, 2022). Furthermore, Riar, Morschheuser, Zarnekow and Hamari (2022) argue that designing cooperative gamification brings several advantages, including fostering social identity, team cohesion, and interpersonal ties.

# 3 Example Use Case

This section describes the on-demand shared transportation example used here to showcase how gamification can be used to improve the use of a transportation system and then lists considerations for applying gamification techniques to groups of people.

## 3.1 System Description

We consider the user of the transport system $i$, represented by his travel request $p_i(O_i, D_i, T_I)$. The request describes the origin of the trip $O_i$, the destination $D_i$, and the time the user wants to travel $T_i$. The request will be fulfilled once the user who issued it is picked up by a transport at $O_i$ at a time later than or equal to $T_i$ and finally moved to $D_i$. If the transport operator detects that this request is too expensive for the system to fulfil or simply impossible, the gamification system is activated to motivate the user to modify it. This modification may include an alternative location of origin or destination and a travel time before or after the one entered by the user. The system will calculate the possible modifications and present them to the user together with a reward that motivates them to accept a suggested modification. The reward can represent various benefits depending on the gamification system being implemented. For example, it could be a score on a public scoreboard or a discount on the price of the user's trip. The user will, therefore, have the freedom to accept or reject the modifications offered to them and will be rewarded whenever their decision cuts the costs of the transport system.

## 3.2 Group Decision Making

In the previous section, we described how gamification techniques can be used to improve a transport system. The dynamic described, however, has considered the interaction with a single user. After all, a real system has several users requesting trips simultaneously, so the gamification system must adapt to this situation. Below, we list different considerations that must be considered to make this adaptation.

**Grouping of the decision.** In certain cases, the system may require that each of the group's users accept the same modification on their journey since if they do not all, the benefit for the system would be non-existent. In these cases, a partial commitment process is carried out. Users are informed of the number of participants in the process and are given a maximum time to make their individual decisions. If all participants' decisions are not unanimous when the given time ends, the trip modification is considered rejected, and no user is rewarded.

**Rejection penalty.** Given a modification to the journey of a group of users, the system may partially benefit if most users accept it. The benefit, however, will be less than if the acceptance were unanimous. To motivate this unanimous acceptance, the minority of users who have chosen to reject the modification may be penalised. However, this type of more aggressive technique must be applied very carefully to avoid the appearance of negative feelings in the user regarding the transport system.

**Negotiation between users.** The last case we consider contemplates the negotiation between the group's users, who must make the decision. Given a specific travel modification, each user can propose changes to the modification, thus making it more attractive to them. This negotiation process takes place in turns, during which users make proposals. Once the final round is reached, the group's decision must be unanimous, either accepting or rejecting the negotiated modification.

## 3.3 Proposed Group Gamification Approach

The approach described above, in Sec. 3, is geared towards the individual and their role in a group. Due to the constraints of the used example, this group consists of other commuters, i.e. other users of the on-demand shared transportation system. It may be observed that all the people in the same group share a common goal – the aim of their interaction is to lower their fare as much as possible by utilising the system and the rewards it offers them. The concept of a reward is used very freely in this particular example, with no specified features at the moment other than bringing benefit to the users if they all agree to reach the same decision.

Following the content presented in Sec. 2.4, the gamified system is expected to be more effective in a

cooperative context. Even so, to further enhance the users' motivation and eagerness to interact with the system and each other, we propose the combined approach of cooperation and competition here.

One suggestion following the recent guidelines on social and cooperative gamification outlined in Sec. 2.4 combines cooperation and competition on different levels of social interaction. We propose a mechanic of individual players grouped into long-term groups (named herein guilds). When faced with a challenge or a task, such individual players are, due to the described nature of the example on-demand shared transportation system in Sec. 3, grouped into small short-term temporary groups (named herein parties). In such a setting: i) The individual players can behave following the interpersonal competition, at the very least collecting green points for helping the transportation system spend fewer resources, showcase those on a leaderboard, customise their avatars, socialise with other users, etc. ii) Guilds can act according to intergroup competition, for example, visually spreading their influence showcased as a real-world map overlay, based on the number of members or the number of members following the rules of the group and successfully achieving individual or group objectives and accepting modifications proposed by the system. iii) Individual players within their chosen guilds can behave cooperatively, driven by the common group goals and the aim of strengthening their guild. Furthermore, the social benefits of being a part of the group may include ways of providing a helping hand to fellow guild members. iv) Individual players within the temporary grouping (a party) would also be advised to act cooperatively. All of the members of a party receive the greatest benefits only if they cooperate.

The described system utilises various gamification elements that cater to both the individual and the group levels. The latter being the more important for this paper, it should be noted here that providing challenges on a group level is argued here to be one of the ways of influencing the behaviour of the group members. Providing visual feedback on the strength of the group, as opposed to providing feedback only on the strength of the individual user, is argued here to be beneficial for fostering group behaviour. Further discussion on the effect of the proposed system and further possible gamification elements and their effects in this particular application domain is presented in the following section.

# 4 Discussion

The suggestion of individual competition combined with intra-guild and -party cooperation, and inter-guild competition, might seem to oppose the intergroup competition on the guild level, yet this dichotomy of the source of individual's power and competitive advantage over other users creates a complex social dynamic that may mitigate the possible stagnancy of motivation to follow the gamified system's rules, suggestions, and modification requests.

The proposed approach to interpersonal competition ties in with the described example system's goal of minimising the number of users who are ready to reject the suggested journey modification. Since one of the users' goals is to accumulate as many rewards and benefits as possible, thus winning over the competition posed by other users, it is expected to be natural for users to avoid receiving penalties or not receiving rewards. Further motivation and engagement enhancement is expected to be driven by other available gamification techniques, such as attractive narrative and graphically customised visual representation of the real world, providing emphasis on the effect generated by single user's decisions to accept the system-provided journey modifications in terms of environmental and similar changes, individual and group-based challenges, time-limited journey modification offers, customised reward system that chooses the best award from a predefined set based on the user's profile thus enticing curiosity and unpredictability.

The proposed temporary grouping mechanic described in Sec. 3.3 provides a gamified approach to the decision grouping and interpersonal negotiation stages of the dynamic described in Sec. 3.2. Since the users compete with each other, they are expected to aim to maximise their rewards and benefits. On the other hand, the largest benefit from an intra-party interaction can be achieved by cooperation. Therefore, users in a party are competitors by default, who are put into a situation that demands cooperation. To further foster the efficiency of intra-party cooperation, party actions could have greater weight on the user's individual contribution to their guild's standings.

Finally, the long-term grouping mechanic proposed in Sec. 3.3 is argued here to be able to foster the engagement of users in general. Behaving socially benefits the user's engagement with the system and the game, according to many authors in 2. Cooperative inter-guild behaviour incentivises users' inclusion and sense of belonging, for every journey modification they accept provides benefits for their guild. On the other hand, individual users are expected to accept journey modification more eagerly since it would reward them (benefit for the interpersonal competition) and their guild (benefit for the intra-guild cooperation and inter-guild competition).

# 5 Conclusions

This paper has opened a line of research based on applying gamification techniques to improve transport, specifically on-demand shared transportation. As the text demonstrates, applying these techniques to influence the decision of groups of people is not trivial, and,

therefore, the case described must be carefully considered before the approach of a real system.

Regarding future work, we are currently developing the implementation of a transport simulator that allows us to run different experiments with various gamification techniques. The results of these tests will provide relevant information so that both transport operators and researchers can correctly choose the gamification techniques to apply to specific types of transport.

## Acknowledgments

This work is partially supported by grant PID2021-123673 and PDC2022-133161-C32 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". Pasqual Martí is supported by grant ACIF/2021/259 funded by the "Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital de la Generalitat Valenciana". Jaume Jordán is supported by grant IJC2020-045683-I funded by MCIN/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR". Bogdan Okreša Đurić is supported by the project MOBODL-2023-08-5618 funded by the European Union and the Croatian Science Foundation. Markus Schatten and Bogdan Okreša Đurić are supported by the Croatian Science Foundation under the project number IP-2019-04-5824.

## References

Avedon, E. M., & Sutton-Smith, B. (1971). *The Study Of Games*. John Wiley & Sons.

Bakhanova, E., Garcia, J. A., Raffe, W. L., & Voinov, A. (2023). Gamification Framework for Participatory Modeling: A Proposal. *Group Decision and Negotiation*, *32*(5), 1167–1182. https://doi.org/10.1007/s10726-023-09838-w

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chou, Y.-K. (2015). *Actionable Gamification: Beyond Points, Badges, and Leaderboards* (1st ed.). CreateSpace Independent Publishing Platform.

Dar, H. S., Imtiaz, S., & Lali, M. I. (2022). Reducing Requirements Ambiguity via Gamification: Comparison with Traditional Techniques (D. Zhang, Ed.). *Computational Intelligence and Neuroscience*, *2022*, 1–12. https://doi.org/10.1155/2022/3183411

Dereli, M., & Kahraman, T. (2024). Gamification in physiotherapy and rehabilitation education: A narrative review. *Bulletin of Faculty of Physical Therapy*, *29*(1), 4. https://doi.org/10.1186/s43161-023-00168-1

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 on Envisioning Future Media Environments*, 9–15. https://doi.org/10.1145/2181037.2181040

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. E. (2011). Gamification: Toward a Definition. *Proceedings of the Gamification Workshop Proceedings*.

Dindar, M., Ren, L., & Järvenoja, H. (2021). An Experimental Study on the Effects of Gamified Cooperation and Competition on English Vocabulary Learning. *British Journal of Educational Technology*, *52*(1), 142–159. https://doi.org/10.1111/bjet.12977

Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., Sohanpal, R., Caton, H., Griffiths, C. J., Munafò, M. R., Taylor, S., & Walton, R. T. (2016). Gamification for Health Promotion: Systematic Review of Behaviour Change Techniques in Smartphone Apps. *BMJ Open*, *6*(10), e012447. https://doi.org/10.1136/bmjopen-2016-012447

Feng, Y., Yi, Z., Yang, C., Chen, R., & Feng, Y. (2022). How Do Gamification Mechanics Drive Solvers' Knowledge Contribution? A Study of Collaborative Knowledge Crowdsourcing. *Technological Forecasting and Social Change*, *177*, 121520. https://doi.org/10.1016/j.techfore.2022.121520

Fonseca, I., Caviedes, M., Chantré, J., & Bernate, J. (2023). Gamification and Game-Based Learning as Cooperative Learning Tools: A Systematic Review. *International Journal of Emerging Technologies in Learning*, *18*(21), 4–23.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A Formal Approach to Game Design and Game Research. *Game Design and Tuning Workshop*.

Imani, S., Widyani, Y., & Rusmawati, Y. (2021). Designing Interaction for Zero-Waste Application Using Cooperative Gamification. *2021 International Conference on Data and Software Engineering (ICoDSE)*, 1–6. https://doi.org/10.1109/ICoDSE53690.2021.9648501

Liu, D., Santhanam, R., & Webster, J. (2017). Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Quarterly: Management Information Systems*, *41*(4), 1011–1034. https://doi.org/10.25300/MISQ/2017/41.4.01

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does Gamification Increase Engagement with Online Programs? A Systematic Review. *PloS One*, *12*(3), e0173403. https://doi.org/10.1371/journal.pone.0173403

Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M. A., Guerrero-Puerta, L., Aguilar-Parra, J. M., Trigueros, R., & Alias, A. (2021). Between Level Up and Game Over: A Systematic Literature Review of Gamification in Education. *Sustainability*, *13*(4), 2247. https://doi.org/10.3390/su13042247

Mohammadi, M., Mohammadhasani, N., & Khoshneshin, Z. (2023). The Effect of Competition and Cooperation Gamification Elements On Learning and Motivation of Fourth-Grade Math. *Research in Teaching*. https://doi.org/10.22034/trj.2023.62834

Mora, A., Riera, D., González, C., & Arnedo-Moreno, J. (2017). Gamification: A systematic review of design frameworks. *Journal of Computing in Higher Education*, *29*(3), 516–548. https://doi.org/10.1007/s12528-017-9150-4

Morschheuser, B., Maedche, A., & Walter, D. (2017). Designing Cooperative Gamification: Conceptualization and Prototypical Implementation. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2410–2421. https://doi.org/10.1145/2998181.2998272

Priesterroth, L., Grammes, J., Holtz, K., Reinwarth, A., & Kubiak, T. (2019). Gamification and Behavior Change Techniques in Diabetes Self-Management Apps. *Journal of Diabetes Science and Technology*, *13*(5), 954–958. https://doi.org/10.1177/1932296818822998

Redondo-Rodríguez, C., Becerra-Mejías, J. A., Gil-Fernández, G., & Rodríguez-Velasco, F. J. (2023). Influence of Gamification and Cooperative Work in Peer, Mixed and Interdisciplinary Teams on Emotional Intelligence, Learning Strategies and Life Goals That Motivate University Students to Study. *International Journal of Environmental Research and Public Health*, *20*(1), 547. https://doi.org/10.3390/ijerph20010547

Riar, M., Morschheuser, B., Hamari, J., & Zarnekow, R. (2020). How Game Features Give Rise to Altruism and Collective Action? Implications for Cultivating Cooperation by Gamification. *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Riar, M., Morschheuser, B., Zarnekow, R., & Hamari, J. (2022). Gamification of Cooperation: A Framework, Literature Review and Future Research Agenda. *International Journal of Information Management*, *67*, 102549. https://doi.org/10.1016/j.ijinfomgt.2022.102549

Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting Early Adolescents' Achievement and Peer Relationships: The Effects of Cooperative, Competitive, and Individualistic Goal Structures. *Psychological Bulletin*, *134*(2), 223–246. https://doi.org/10.1037/0033-2909.134.2.223

Saleem, A. N., Noori, N. M., & Ozdamli, F. (2022). Gamification Applications in E-learning: A Literature Review. *Technology, Knowledge and Learning*, *27*(1), 139–159. https://doi.org/10.1007/s10758-020-09487-x

Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-Health. *Journal of Biomedical Informatics*, *71*, 31–48. https://doi.org/10.1016/j.jbi.2017.05.011

Schatten, M., Tomičić, I., & Okreša Đurić, B. (2017). A Review on Application Domains of Large-Scale Multiagent Systems. In V. Strahonja & V. Kirinić (Eds.), *Central European Conference on Information and Intelligent Systems* (pp. 201–206). Faculty of Organization and Informatics, University of Zagreb.

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

Werbach, K., & Hunter, D. (2015). *The Gamification Toolkit: Dynamics, Mechanics, and Components for the Win*. Wharton Digital Press.

Yamani, H. A. (2021). A Conceptual Framework for Integrating Gamification in eLearning Systems Based on Instructional Design Model. *International Journal of Emerging Technologies in Learning*, *16*(4), 14–33. https://doi.org/10.3991/ijet.v16i04.15693

Zhang, J., Jiang, Q., Zhang, W., Kang, L., Lowry, P. B., & Zhang, X. (2023). Explaining the Outcomes of Social Gamification: A Longitudinal Field Experiment. *Journal of Management Information Systems*, *40*(2), 401–439. https://doi.org/10.1080/07421222.2023.2196776

# Towards a Gamification-Enhanced Agent-Based Simulations for the Digital Transformation Process

**Bogdan Okreša Đurić**
Artificial Intelligence Laboratory
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{dokresa}@foi.unizg.hr

**Ana Kutnjak, Larisa Hrustek**
Business Process Management and Digital Transformation Laboratory
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{akutnjak,lhrustek}@foi.unizg.hr

**Jože M. Rožanec**
Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
joze.rozanec@ijs.si

**Francisco Enguix**
Valencian Research Institute for Artificial Intelligence
Universitat Politècnica de València
Camino de Vera s/n, 46022 Valencia, Spain
fraenan@upv.es

**Abstract.** *This paper presents a novel approach to digital transformation simulation using agent-based modelling within intelligent virtual environments. We propose a three-phase architecture integrating ontology-based modelling, customised implementation, and optionally-gamification-enhanced, more nuanced and interactive simulation. Unlike traditional token-based simulations, our approach leverages artificial agents' social and interactive nature to provide richer feedback on planned business process transformations. We demonstrate the application of our architecture using two scenarios from a production planning and scheduling use case. This method offers digital transformation experts a powerful tool for evaluating transformation strategies, predicting resource requirements, and observing emergent behaviours in digitised business processes.*

**Keywords.** digital transformation, artificial intelligence, artificial agents, multiagent systems, gamification, digital twins, simulation

## 1 Introduction

Digital transformation is a "thorough transformation of business that includes innovative changes in key organizational activities, processes, products/services, and business models, with the aim of utilizing the potential provided by digital technologies" (Pihir et al., 2019). It involves innovative changes in organisational activities, processes, and business models through the use of digital technologies. A critical early step is accurately describing the business process to be digitised. Digital twins, representing real-life entities, serve as intermediate products in this transformation, providing a test bed for proposed modifications.

Artificial agents, entities that can act upon and receive input from their environment (Russell and Norvig, 2022), can be considered digital twins in multiagent systems (MASs). These systems offer practical testing grounds for simulating business processes before real-world implementation.

The authors argue for using ontologies to describe business processes in terms translatable to artificial agents to facilitate this process. This enables the generation of a simulation-ready MAS blueprint, specifically an intelligent virtual environment (IVE) comprising agents and interactive objects.

This paper proposes introducing artificial agents within simulation-ready IVEs can enhance the digital transformation process. This approach allows for more nuanced simulations with agents behaving and interacting in a dynamic environment, possibly influenced by various gamification-based reward systems, unlike the more abstract currently often used token-based simulations.

This paper describes how formal business process descriptions, ontologies, and agent-run ontology-based simulation environments can be combined to provide an active, interactive, and reactive feedback loop in the digital transformation process.

The rest of the paper is organised as follows. Section

2 provides an overview of recent research and necessary definitions in the related domains. The example use case laid out in Sec. 3 is used to illustrate the proposed agent-based simulation architecture (presented in Sec. 4) using two scenarios described in Sec. 5. The approach proposed in this paper is discussed in Sec. 6, followed by plans for future research laid out in Sec. 7.

# 2 Related Work

## 2.1 Intelligent Agent

Artificial intelligent agents are systems capable of environmental perception and action to achieve specified objectives (Russell and Norvig, 2022; Wooldridge, 2009; Wooldridge and Jennings, 1995). These agents are designed for inter-agent cooperation and information exchange to enhance functionality. The key characteristics of intelligent agents include social ability, autonomy, reactivity, adaptability, proactiveness, and goal-orientated behaviour. They are frequently employed in implementing distributed systems, social systems, and systems with autonomous components (Schatten et al., 2017). Modelling social dynamics and phenomena using artificial agents provides a safe methodology for generating valuable data without human risk (Chopra and Singh, 2018).

Multi-agent systems (MAS), consisting of multiple intelligent agents, are gaining prominence with ongoing advancements in smart cities, smart home simulations and large-scale artificial intelligence models. In MAS, intelligent agents can be organized into groups, coordinating and delegating tasks to achieve goals. Figure 1 depicts the definition of an intelligent agent (i) organized into specialized MAS groups (ii) that communicate to achieve common goals in an intelligent virtual environment (IVE), which requires different abilities to achieve the goals (iii).

## 2.2 Intelligent Virtual Environments

Intelligent Virtual Environments (IVEs) represent a sophisticated fusion of high-fidelity environmental simulations and artificial intelligence (Luck and Aylett, 2000a). Central to the functionality of IVEs is the deployment of intelligent agents equipped with cognitive abilities designed to operate and interact within virtual spaces. This combination of virtual landscapes with intelligent agents enables the simulation of lifelike scenarios, enhancing the capability for complex decision-making processes.

By incorporating intelligent agents into these virtual settings, a new layer of dynamic interaction is achieved that transcends the limitations of static simulations. This dynamic allows for a deeper exploration and understanding of interactions between systems, entities, and users in a controlled digital context. Such technology is increasingly being applied in a wide range

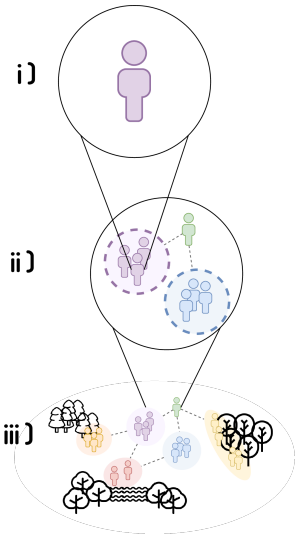



**Figure 1:** Artificial agents on different grouping levels: i) an intelligent agent; ii) a multiagent system; iii) an intelligent virtual environment

of fields, from agricultural advancements (Gutiérrez Cejudo et al., 2024) to sectors including robotics, the Internet of Things (IoT), and urban development (Luck and Aylett, 2000b).

The Flexible IVE Designer (FIVE) is an example of a framework built on Unity and SPADE agents that streamlines the creation of custom IVEs by facilitating the import of 3D models (Carrascosa et al., 2023). Users can construct IVEs through modifications to text files. This platform is equipped with a helper tool that automatically generates 3D virtual environments by allowing users to select real-world locations from an interactive satellite map.

## 2.3 Digital Twin Definition and Background

A digital twin (DT), first proposed in 2003, is defined as making full use of the data such as physical model, sensor update, operation history, etc., and integrating the simulation process with multi-disciplinary, multi-physical quantities, multi-scale, and multi-probability to realise the mapping that can be completed in virtual space, to use it to reflect the corresponding entity equipment life cycle process (Wang and Wu, 2020). The concept is credited to Michael Grieves (Grieves, 2014), who pioneered its development along with John Vickers. The digital twin is meant as the virtual and computerised counterpart of a physical system that can be used to simulate it for various purposes, exploiting a real-time synchronisation of the sensed data coming from the field; such synchronisation is possible thanks to the enabling technologies of Industry 4.0 and, as such, the DT is deeply linked with it (Negri, Fumagalli, and Macchi, 2017).

## 2.4 Digital Transformation Journey: Towards a Process Digital Twin

Technological changes affect the transformation of traditional production processes and their management (Jiang et al., 2024; Liu et al., 2020; Negri, Fumagalli, and Macchi, 2017; Sakr et al., 2021). The era of Industry 4.0 has led to the incorporation of collaborative communication and virtual work in production engineering, and one of the significant paradigms in this domain is Cyber-Physical Systems (CPS) (Negri and Abdel-Aty, 2023; Uhlemann et al., 2017). Production equipment and systems are no longer just physical elements; they have become digital elements that are just as important as representations of the physical system. The digital transformation of the production system is based on drivers such as cloud services and resource virtualisation, virtualisation of shop floor devices, high availability and policy-based security in production (Borangiu et al., 2020). In the context of transformation, data recording, storage, and analysis are increasingly important within the enterprise production system, and embedded systems at the plant level, operations, production execution and resource planning form the basis for a digital twin in a smart factory (Uhlemann et al., 2017). Industry 4.0 and technologies have enabled a vision for the industry of the future based on cyber-physical production systems that provide flexible, dynamically reconfigurable real-time control of strongly connected processes. The information and operational technologies of production systems are merged, and the physical reality of the factory has become a mirror to virtual counterparts, digital twins that represent abstract entities specific to the production domain, taking into account products, orders, and resources (Borangiu et al., 2020).

The application of DT in production refers to CPS whose constant interaction and communication should enable production flexibility (Barthelmey et al., 2019; Jiang et al., 2024; Negri, Fumagalli, and Macchi, 2017; Sakr et al., 2021). For CPS to fulfil its purpose, it is necessary to merge the physical object with its representation in the digital world (Albo et al., 2021; Jiang et al., 2024; Rocha and Barata, 2021; Szabo et al., 2019), that is, the previous creation of a production DT (Barthelmey et al., 2019; Sakr et al., 2021) using different technologies such as Industrial Internet of Things (IIoT) (Sakr et al., 2021), Big Data (Albo et al., 2021), Cloud Computing (Liu et al., 2020), etc., where simulation is an indispensable part of the twins' functioning (Liu et al., 2020; Rocha and Barata, 2021). Simulation enables offline testing, planning and experiments (Albo et al., 2021), while digital twins support the synchronous performance of activities (Liu et al., 2020) and processes and their modification (Rocha and Barata, 2021; Sakr et al., 2021). The aforementioned leads to complementation, optimization, monitoring and control of the production process through the digital world (Jiang et al., 2024; Liu et al., 2020; Rocha and Barata, 2021; Sakr et al., 2021; Szabo et al., 2019), whereby the features of the digital can be used to enrich the physical production counterpart in parallel (Albo et al., 2021; Liu et al., 2020). The concept of DT is based on the two-way exchange of large amounts of production data (Negri, Fumagalli, and Macchi, 2017; Szabo et al., 2019) between a physical and a digital object, whereby changes in one affect changes in the other (Liu et al., 2020) enabling two-way synchronization and real-time updates (Negri, Fumagalli, and Macchi, 2017). This emphasizes the ability of digital twins "to collect and manage information about the system operations, its history, its behaviour and its current state" (Liu et al., 2020).

The production process is subject to uncertainty factors (Jiang et al., 2024; Rocha and Barata, 2021; Wang and Wu, 2022) such as occasional material shortages, changes in customer orders, equipment failures, or deviations in final product quality. Such factors will make it impossible to execute the production process on time. To prevent the same, digital twins can provide a kind of technical support for production (Liu et al., 2020). In the process of production and deployment, digital twins have a two-sided focus – they look at what actually happens in the process (Albo et al., 2021; Rocha and Barata, 2021), but also enable understanding of what may happen in the future (Rocha and Barata, 2021; Sakr et al., 2021).

Despite the many positive aspects of DT, the authors warn of increasing cybersecurity vulnerabilities such as physical tampering, supply chain attacks and sabotage (Jiang et al., 2024), which requires special attention. For this reason, digital twins can be used to find vulnerabilities in the production system (Rocha and Barata, 2021) and simulate attacks to strengthen the cyber security of the production sector (Jiang et al., 2024).

## 2.5 Digital Twin Classification

Since digital twins are virtual synchronised representations of physical entities that aim to mirror their real-world counterparts as closely as possible, they become novel tools for industries (Purcell and Neubauer, 2023), promising real-time insights and improved decision-making for physical assets.

According to Kritzinger et al. (2018), there are three distinct levels of integration in the implementation of DT, each characterised by the depth and directionality of data flow: digital model (DM), digital shadow (DS), and digital twin (DT). Figure 2 illustrates the difference between them.

A **digital model** exhibits no automated data transfer between the physical entity and its virtual model. Interaction is manual, limiting the potential for real-time analytics and adaptation. Data of a **digital shadow** flow unidirectionally and automatically from the physical entity to the virtual model at this stage. This allows for some level of monitoring and simulation based on

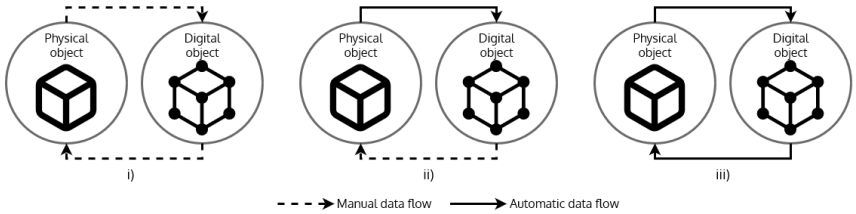



**Figure 2:** i) digital model; ii) digital shadow; iii) digital twin

real-time data, but without reciprocal interaction from the virtual to the physical. A **digital twin** represents the most integrated level, where data exchange is bidirectional. Changes in the virtual model are immediately reflected in the physical entity and vice versa. This dual interaction facilitates enhanced simulation, real-time feedback, and preemptive decision-making.

# 3 Example Use Case

The process of production planning and production capacity scheduling in typical production organisations includes the activities of the production, procurement, and storage departments. Process is intertwined with a series of activities in the above mentioned departments. Roughly presented by the authors, production planning and capacity scheduling can be divided into four fundamental phases that include multiple activities and, in a digital context, the creation of algorithms (Figure 3):

Phase 1: **Planning** involves scheduling production processes, raw materials, and resources to produce products for consumers within predetermined time frames. It determines what needs to be produced and how much work should be done. The goal is to ensure that products are processed efficiently and on time. Effective production planning impacts critical parts of a manufacturing organisation, such as capacity planning, supply chain management, production lead time, and material requirements planning. This phase includes processing input data (data store) based on the planned needs of the established demand, information on the product's technical characteristics, and the basis for setting up algorithms for planning needs for purchase and production. Phase input is a data store (inventory status, final product inventory status, planned incoming inputs, bill of materials, etc.) consisting of: 1. Standardisation of material factors: defining and establishing all the settings of the products that are planned to be produced, respecting the quality standards of the resources that are obtained for production purposes; 2. Technical documentation preparation: checking and determining the correctness of the technical documentation of the products that are planned to be produced, taking into account documents such as bills of material for products; 3. Orders generation: material requirement planning algorithm. Phase outputs are documents (work orders, purchase orders, recommendations for replanning purchase or production).

Phase 2: **Preparation** includes collecting all necessary documents for production planning and determining the current state of the production process. Phase input: Documents (work orders, purchase orders, recommendations for replanning purchase or production) containing: 1. Draft production planning: defining the time period for which the plan is made, collecting work orders, determining the approximate needs for materials; 2. Final production planning: defining the final production needs in the planned periods, defining the type, quantity, and terms of purchase of materials, and creating the final production plan. Phase output: Documents (production plan, purchase plan).

Phase 3: **Scheduling** of production focuses on planning how many capabilities (production resources) are needed and when they must be ready for production. It specifies who will perform the operations and when those operations will take place. By combining production needs with available resources cost-effectively, production scheduling ensures that the manufacturing process runs as smoothly as possible. Phase input: Documents (production plan, purchase plan) contains data on: 1. Checking the status of production capacities: within this activity, the available resources (such as machinery, labour, and materials) are allocated to meet production demands. Capacity planning ensures the right resources are in place to handle the workload. It considers factors like production capacity, lead times, and resource availability; 2. Planning production capacities occupation: Production schedules are created once the resources are allocated. This involves determining the sequence of production orders, assigning time slots for each operation, and coordinating activities across different departments. The goal is to optimise production flow and minimise idle time. Phase output: Document (occupation plan on capacities).

Phase 4: **Implementation** of production activities is carried out according to the schedules during the implementation phase. Capacities are occupied with work orders, and production activities have started. Managers monitor progress, track performance metrics, and make adjustments as needed. Phase input: Document (occupation plan on capacities) detailing: 1. Execution of orders on production capacities: work orders are scheduled on capacities based on the occupation plan, and they are processed during the production time. Phase output: Database updates (inventory status, final product inventory status, etc.).

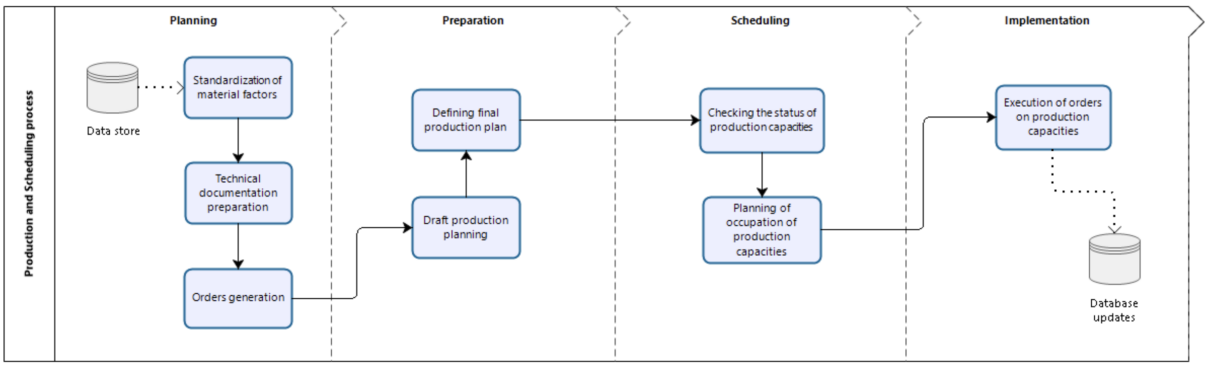



**Figure 3:** Diagram of the four phases in production planning and capacity scheduling.

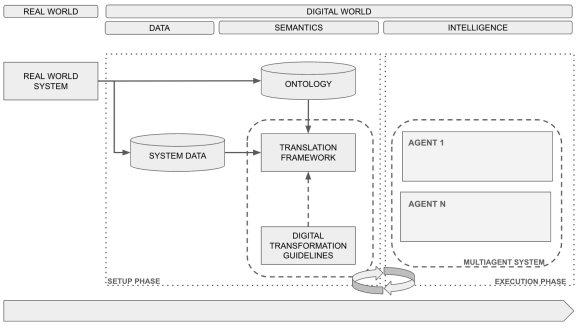



**Figure 4:** Conceptual model of the agent-based simulation architecture for digital transformation, leveraging ontologies and artificial agents to simulate real-world scenarios.

# 4 Proposed Architecture

This paper proposes a conceptual model of the non-specified work-in-progress architecture of an agent-based approach to simulation for the digital transformation process that uses entities in intelligent virtual environments. This approach is divided into the following phases: i) setup phase, ii) customised implementation phase, iii) execution phase (Figure 4).

In the setup phase, real-world elements are matched to intelligent virtual environment components using an ontology being developed based on (Okreša Đurić et al., 2019). This ontology aims for necessary expressiveness without over-constraining applicability. The phase models the real-world scenario using IVE concepts, ensuring initial validity. Validation occurs through the initial simulation run post-implementation, with the setup phase repeated if deemed invalid. The modelled system is valid if the outcomes match the expected behaviour of the real-world system. In the context of digital transformation of business processes, this involves abstracting processes as agents or introducing actor agents that perform the modelled processes. A novel approach of modelling agents as processes for specifying inter-agent communication flow (Peharda et al., 2023) could be applied to convert business processes to artificial agents or facilitate their communication.

The implementation phase consists of translating the ontology-based model of the first phase into an implementation blueprint. This phase uses the in-development framework to translate selected concepts of the chosen ontology into Python implementation blueprints. This blueprint contains the basic implementation of the modelled system, similar to the process described in (Okreša Đurić, 2017). The blueprint is expected to be further specified by a developer based on the domain of the observed real-world system, the goals and guidelines of digital transformation, and the expected comprehensiveness of the simulation, which is the crucial part of phase iii).

Once the IVE and the incorporated agents representing the modelled elements of the world are implemented, system data inputs are used during the simulation phase to describe the real world and to perform the tasks for which the agents were commissioned. Therefore, the agents and objects within the simulation can be observed as digital twins. Furthermore, based on the customised implementation phase, the agents can be observed on any of the three levels of digital twin integration presented in Sec. 2.5. We consider the ever-evolving nature of the digital twin to match and refine the required functionalities, which requires an iterative process where, for each new version, the setup phase may need to be re-run to execute the new environment being deployed.

Gamification techniques are proposed to enhance the natural interaction of simulated agents, complementing their core features. By implementing gamification beyond basic points and badges, agents' behaviour can be subtly modified to reflect nuanced changes in real-world human behaviour. Techniques such as rewarding goal-oriented interaction, encouraging grouping, and incentivising efficiency can be applied depending on simulation requirements. While these essentially form reward systems for agents, framing them as gamification concepts facilitates their application and translation to real-world human agents.

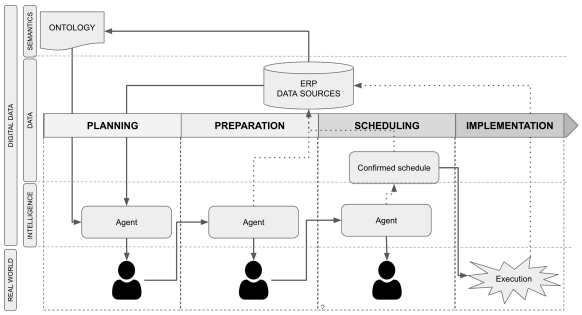



**Figure 5:** Integration of the proposed agent-based simulation architecture to illustrate the application of the digital transformation process.

# 5 Applying the Proposed Architecture to the Example Use Case

Digital transformation of the process described in Sec. 3 is presented here in two distinct possible scenarios: i) digital transformation of Planning (phase 1) and Preparation (phase 2) production process; ii) digital transformation of Scheduling (phase 3) and Implementation (phase 4) production process (Figure 5). Only the simulation stage of the digital transformation process is described in this section, following the three phases of the agent-based approach presented in Sec. 4. The setup phase is the same in both scenarios since both scenarios are observed within the same system.

## 5.1 The First Scenario

In phase 1 (Planning), customer orders are received, which serve as the basis for production planning. Upon the arrival of customer orders, work orders are generated and forwarded in digital form to phase 2 (Preparation). With digital transformation, inputs (material stocks) are synchronised in real time and allocated to work orders. All documents are digital, updated in real-time, and forwarded to subsequent stages. Additionally, due to changes in the inventory status, material requirements are generated in real-time, ensuring constant updates on how much to order from suppliers. The production plan is digitally synchronised in real-time, making production requirements dynamically swiftly to changing market conditions, minimising bottlenecks, and achieving higher efficiency.

The setup phase of this scenario should be customised with additional agents and artefacts related to the observed concepts that play the leading roles of this scenario. For example, should the digital transformation expert wish for a detailed and highly customisable simulation, they may choose to model both customer and work orders as separate agents. Furthermore, an overseer agent can be modelled to oversee the two described phases of the business process. The process of forecasting production needs can be modelled as a separate agent. Inventory status, material stocks, and material requirements can be modelled as simple data inputs or IVE artefacts within the system. Ultimately, suppliers may be modelled as agents, as other entities, or as simple data flows, depending on the desired abstraction and interaction level.

Various behaviours can be implemented in the customised implementation phase, granting different levels of intelligence and autonomy to the included agents; e.g., additional agents may be included that behave as forecasting agents. When the simulation is run, the behaviour of the observed system can be deduced from the individual and emergent interactions of the included agents. Running the simulation is expected to give the digital transformation expert an estimate of, for example, various resource consumption. Such data is predicted to be valuable in the digital transformation decision-making process.

## 5.2 The Second Scenario

In phase 3 (Scheduling), digital work orders are allocated to capacities, i.e., included in the occupation plan of work orders on capacities. The plan is digital and changes in real time, according to the enterprise's defined algorithms (rules). At any moment, the current workload, available capacities, their status, and the capacity requirements for the working day are known. As a result, the need for overtime work and additional human resources is determined. In phase 4 (Implementation), digital work orders are distributed to capacities (machines), and the production preparation time, the production flow, and the completion of the production are determined. In the event of a potential failure or other incidents that affect the normal production flow, the digital process allows feedback on previous stages and enables replanning according to new requirements. The role of the human factor is reduced to a minimum in the entire process, i.e., the human has only a controlling role.

The setup phase of this scenario should include agents that can play the role of work orders and capacities. Based on the desired level of interaction and precision, the amount of work, overtime work, and human resources could be modelled as simple data flows. The forecasting process can be implemented as an individual agent.

Various behaviours can be implemented as available to the simulation agents, e.g., the forecasting agent forecasts daily production capacity, determines the need for extra shifts, and calculates the required workforce for task completion. Upon running the simulation, the results of the agent interaction provide the digital transformation expert with valuable input.

# 6 Discussion

The approach presented in this paper is expected to provide digital transformation experts with valuable feed-

back on the benefits of the planned transformation of the initial business process. These benefits may be observed, e.g., as quantified results, emergent behaviour, or relative change of used resources compared to the initial real-world process. The benefit of utilising artificial agents in a simulation instead of the traditional toke-based simulations is the social and interactive nature of the concept of intelligent artificial agents by default.

Interaction that stems from the fundamental features of artificial agents is argued here to be able to provide different and possibly much more valuable feedback than token-based simulations. Furthermore, implementing intelligent agents is a way of introducing modern methods and algorithms of artificial intelligence into the equation of digital transformation and transformation suggestions or guidelines. Finally, influencing agents' behaviour using gamification techniques can allow digital transformation experts to observe the seemingly natural evolution of the initial model.

# 7 Conclusion & Future Research

The agent-based approach to digital transformation simulation presented in this paper offers several advantages over traditional methods. By utilising intelligent artificial agents within IVEs, our architecture enables more dynamic and interactive simulations that better reflect the complexities of real-world business processes.

Key contributions of this work include 1. a flexible three-phase architecture that combines ontology-based modelling, customised implementation, and gamification-enhanced simulation; 2. integration of artificial intelligence and multiagent systems into the digital transformation process, allowing for more sophisticated modelling of business processes and human factors; 3. the use of gamification techniques to influence agent behaviour, providing a novel way to simulate subtle changes in human behaviour within transformed business processes; 4. a demonstration of the architecture's applicability through two scenarios in a production planning and scheduling use case.

Our approach addresses the limitations of token-based simulations by incorporating the social and interactive nature of intelligent agents. The stated allows digital transformation experts to observe emergent behaviours, quantify resource requirements, and comprehensively evaluate planned transformations' potential benefits.

Future research directions include, but are not limited to, further development and refinement of the ontology for describing business processes and IVE components, expansion of the gamification techniques applied to agent behaviour to model more complex human factors, integration of machine learning algorithms to enhance agent decision-making and adaptability within simulations, empirical validation of the proposed architecture through real-world case studies across various industries, development of visualisation tools to better represent simulation outcomes and aid in decision-making for digital transformation strategies.

By providing a more nuanced and interactive simulation environment, our agent-based approach has the potential to significantly improve the planning and implementation of digital transformation initiatives across diverse business domains.

# Acknowledgement

Bogdan Okreša Đurić is supported by the MOBODL-2023-08-5618 project funded by the European Union and the Croatian Science Foundation.

# References

Albo, A., Svedlund, L., & Falkman, P. (2021). Modular Virtual Preparation method of production systems using a Digital Twin architecture. *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA )*, 1–8. https://doi.org/10.1109/ETFA45728.2021.9613654

Barthelmey, A., Lee, E., Hana, R., & Deuse, J. (2019). Dynamic digital twin for predictive maintenance in flexible production systems. *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, *1*, 4209–4214. https://doi.org/10.1109/IECON.2019.8927397

Borangiu, T., Morariu, O., Răileanu, S., Trentesaux, D., Leitão, P., & Barata, J. (2020). Digital Transformation of Manufacturing. Industry of the Future with Cyber-Physical Production Systems. *Romanian Journal of Information Science and Technology*, *23*(1), 3–37
Accepted: 2023-02-06T22:16:37Z.

Carrascosa, C., Enguix, F., Rebollo, M., & Rincon, J. (2023). Consensus-based learning for mas: Definition, implementation and integration in ives. *International Journal of Interactive Multimedia and Artificial Intelligence*, *8*, 21–32. https://doi.org/10.9781/ijimai.2023.08.004

Chopra, A. K., & Singh, M. P. (2018). Sociotechnical Systems and Ethics in the Large. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 48–53. https://doi.org/10.1145/3278721.3278740

Grieves, M. (2014). *Digital Twin: Manufacturing Excellence through Virtual Factory Replication* (Whitepaper).

Gutiérrez Cejudo, J., Enguix Andrés, F., Lujak, M., Carrascosa Casamayor, C., Fernandez, A., & Hernández López, L. (2024). Towards agrirobot digital twins: Agri-ro5 –a multi-agent architecture for dynamic fleet simulation. *Electronics*, *13*(1). https://doi.org/10.3390/electronics13010080

Jiang, Y., Wang, W., Ding, J., Lu, X., & Jing, Y. (2024). Leveraging Digital Twin Technology for Enhanced Cybersecurity in Cyber–Physical Production Systems. *Future Internet*, *16*(4), 134. https://doi.org/10.3390/fi16040134

Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, *51*(11), 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474

Liu, C., Jiang, P., & Jiang, W. (2020). Web-based digital twin modeling and remote control of cyber-physical production systems. *Robotics and Computer-Integrated Manufacturing*, *64*, 101956. https://doi.org/10.1016/j.rcim.2020.101956

Luck, M., & Aylett, R. (2000a). Applying artificial intelligence to virtual reality: Intelligent virtual environments. *Applied Artificial Intelligence*, *14*, 3–32. https://doi.org/10.1080/088395100117142

Luck, M., & Aylett, R. (2000b). Guest editorial: Intelligent virtual environments. *Applied Artificial Intelligence*, *14*, 1–2. https://doi.org/10.1080/088395100117133

Negri, E., & Abdel-Aty, T. A. (2023). Clarifying concepts of Metaverse, Digital Twin, Digital Thread and AAS for CPS-based production systems. *IFAC-PapersOnLine*, *56*(2), 6351–6357. https://doi.org/10.1016/j.ifacol.2023.10.818

Negri, E., Fumagalli, L., & Macchi, M. (2017). A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing*, *11*, 939–948. https://doi.org/10.1016/j.promfg.2017.07.198

Okreša Đurić, B. (2017). Organisational Metamodel for Large-Scale Multi-Agent Systems: First Steps Towards Modelling Organisation Dynamics (S. Omatu & J. M. Corchado, Eds.). *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *6*(3), 17–27. https://doi.org/10.14201/ADCAIJ2017631727

Okreša Đurić, B., Rincon, J., Carrascosa, C., Schatten, M., & Julian, V. (2019). MAMbO5: A new Ontology Approach for Modelling and Managing Intelligent Virtual Environments Based on Multi-Agent Systems. *Journal of Ambient Intelligence and Humanized Computing*, *10*(9), 3629–3641. https://doi.org/10.1007/s12652-018-1089-4

Peharda, T., Okreša Đurić, B., & Tomičić, I. (2023). Towards Application of Programming Language for Communication Flows Specification in Multiagent Systems on Real-World Use Cases. In N. Vrček, L. de Marcos Ortega, & P. Grd (Eds.), *Proceedings of the Central European Conference on Information and Intelligent Systems* (pp. 17–22). Faculty of Organization and Informatics.

Pihir, I., Tomičić-Pupek, K., & Tomičić Furjan, M. (2019). Digital transformation playground-literature review and framework of concepts. *Journal of information and organizational sciences*, *43*(1), 33–48.

Purcell, W., & Neubauer, T. (2023). Digital twins in agriculture: A state-of-the-art review. *Smart Agricultural Technology*, *3*, 100094. https://doi.org/https://doi.org/10.1016/j.atech.2022.100094

Rocha, A. D., & Barata, J. (2021). Digital twin-based optimiser for self-organised collaborative cyber-physical production systems. *Manufacturing Letters*, *29*, 79–83. https://doi.org/10.1016/j.mfglet.2021.07.007

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

Sakr, A. H., Aboelhassan, A., Yacout, S., & Bassetto, S. (2021). Building Discrete-Event Simulation for Digital Twin Applications in Production Systems. *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA )*, 01–08. https://doi.org/10.1109/ETFA45728.2021.9613425

Schatten, M., Tomičić, I., & Okreša Đurić, B. (2017). A Review on Application Domains of Large-Scale Multiagent Systems. In V. Strahonja & V. Kirinić (Eds.), *Central European Conference on Information and Intelligent Systems* (pp. 201–206). Faculty of Organization and Informatics, University of Zagreb.

Szabo, G., Racz, S., Reider, N., Munz, H. A., & Peto, J. (2019). Digital Twin: Network Provisioning of Mission Critical Communication in Cyber Physical Production Systems. *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 37–43. https://doi.org/10.1109/ICIAICT.2019.8784852

Uhlemann, T. H.-J., Schock, C., Lehmann, C., Freiberger, S., & Steinhilper, R. (2017). The Digital Twin: Demonstrating the Potential of Real Time Data Acquisition in Production Systems. *Procedia Manufacturing*, *9*, 113–120. https://doi.org/10.1016/j.promfg.2017.04.043

Wang, Y., & Wu, Z. (2020). Model Construction of Planning and Scheduling System Based on Digital Twin. *The International Journal of Advanced Manufacturing Technology*, *109*(7-8), 2189–2203. https://doi.org/10.1007/s00170-020-05779-9

Wang, Y., & Wu, Z. (2022). Digital Twin-Based Production Scheduling System for Heavy Truck Frame Shop. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *236*(4), 1931–1942. https://doi.org/10.1177/0954406220913306

Wooldridge, M. (2009). *An Introduction to MultiAgent Systems* (2nd ed.). John Wiley & Sons Ltd.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The knowledge engineering review*, *10*(2), 115–152.

# Understanding engagement patterns of game-based learning for children in CatnClever using learning analytics

**Josmario Albuquerque, Bart Rienties**
The Open University
Institute of Educational Technology
Walton Hall, Kents Hill, Milton Keynes MK7 6AA
{josmario.albuquerque, bart.rienties}@open.ac.uk

**Sebastian Hähnel**
Clever Forever Education AG
Lerchenfeldstrasse 3 9014 St. Gallen. Switzerland
sebastian@catnclever.com

**Kristina Cordero**
Brooklyn College, City University of New York
2900 Bedford Avenue Brooklyn, NY 11210
kristina.cordero@brooklyn.cuny.edu

***Abstract***. *Despite the potential of game-based learning (GBL) for young children, its effectiveness remains under-researched, particularly in preschool education. This study examines engagement data from 8,365 German preschoolers across 60,279 activities within the GBL app CatnClever. Using learning analytics, we aim to assess activity difficulty, learner effort, and performance without relying on formal assessment data. Findings suggest that difficulty aligns well with effort, suggesting optimal challenge levels. Key engagement moments were also identified, potentially informing further interventions. Overall, we stress the potential of learning analytics to deepen our understanding of young learners' interactions in GBL, paving the way for tailored educational strategies.*

**Keywords.** Game-Based Learning, Learning Analytics, Preschool Education, CatnClever

## 1 Introduction

It is widely acknowledged that games and game-based learning (GBL) opportunities can spark engaged learning opportunities (Sun et al., 2023). Well-designed games can lead to enjoyable and engaging experiences, in particular for young people and children (Maureen et al., 2022; Plass et al., 2015). Furthermore, games can be fun, interesting, motivating, and playful.

While a lot of literature has focussed on primary and secondary school children (Guan et al., 2024) and adults (Banihashem et al., 2023; Battistella & von Wangenheim, 2016; Gee, 2004; Mahat et al., 2022), there is mixed evidence regarding whether GBL has positive (e.g., knowledge, skills, attitudes) or negative impacts on young children's learning experience (Guan et al., 2024). In particular, there is a paucity of research in how young learners engage in GBL.

As argued in a recent systematic literature review by Guan et al. (2024), play is essential for child development. In this explorative study we aimed to investigate how learners engaged in one specific game-based mobile app called CatnClever that is designed specifically for children aged between 3-6. Using principles of learning analytics and artificial intelligence (Banihashem et al., 2023), we are specifically interested in exploring whether we could use engagement data on children's progress over time in CatnClever to predict learning performance without formal assessment data. As extensive testing of young children might not be appropriate in terms of gameplay, motivation, and data collection (Kucirkova et al., 2024), in this study we sought to explore whether we could estimate the learning performance and activity difficulty of 170 CatnClever activities in four subjects (i.e., mathematics, language, social and emotional learning, and sport) purely from children's engagement data.

## 2 Background

In a recent systematic literature review of 35 experimental studies of GBL amongst primary school children by Guan et al. (2024) found that 54% of studies reported positive outcomes when using GBL (e.g., academic success, creativity, knowledge, motivation), while 43% reported mixed findings. However, most identified studies used relatively basic descriptives of outcomes rather than actual engagement data, and mostly reported on relatively small samples of learners (< 200 learners).

There is ample consensus on the effectiveness of storytelling as a pedagogical tool, particularly for helping children develop literacy skills in their first language (Koehnecke, 2000), as well as for second language learning (Ellis & Brewster, 2002). But stories

can also support learning in subjects like history, science, philosophy and even math (Eades, 2005) and can help learners develop twenty-first century skills like critical thinking (Roche, 2014) and social and emotional learning. Through vivid, dramatic sequences of events, stories create drama, elicit emotion, and convey content in a way that is memorable (Egan, 2005). Stories are powerful, transformative tools that connect even the youngest of learners to the meaning of being humans. As Bishop (1990) noted, stories can be mirrors of our own experiences, windows into other worlds, and sliding doors connecting the two.

Advancements in learning analytics (Banihashem et al., 2023; Chen et al., 2022; Emerson et al., 2020) and artificial intelligence can help to shed light on the complex engagement patterns of learners in GBL. For example, Banihashem et al. (2023, p. 2) argued that "[o]ne promising opportunity in digital GBL is harvesting and analysing learners' data generated in GBL environments for the purpose of understanding how learning happens in such learning environments to better support the learning process".

By looking at detailed logfiles and engagement patterns of learners' progress within a game over time, educators and researchers alike may be able to understand how learning takes place within a unit of a game, across several units or stories, and even across the game over time (Banihashem et al., 2023; Emerson et al., 2020). As GBL environments are often data rich and highly interactive, GBL has the potential to provide a lot of valuable data that could be tracked, analysed, and possibly visualised back to educators, researchers, and even end-users (Kaliisa et al., 2024).

Indeed in their recent systematic literature review of GBL of 20 studies in higher education using learning analytics Banihashem et al. (2023, p. 2) found that all studies used gameplay data, while 30% used player data about the actual gamer, and 20% reported on the game metadata. While these data might be useful for understanding the complexities of game-play and learning, in particular amongst young children, there are obvious concerns in terms and how this could be done in an ethically appropriate manner (Kucirkova et al., 2024).

In this paper we aim to report on how we used learning analytics data to understand the engagement patterns in a GBL environment called CatnClever. CatnClever is a play-based app for children aged 3-6 that supports early language, mathematics, social and emotional learning (SEL), and physical training skills through original, illustrated stories with appealing animal characters and music. Based on the Swiss-German curriculum and the International Baccalaureate Early Years Programme, CatnClever integrates skills (counting and spatial imagination in mathematics; vocabulary and phonological awareness in literacy; empathy and emotional identification in SEL) into playful stories revolving around themes relevant to young children. This story-based learning approach (McQuiggan et al., 2008) is essential in the design of CatnClever. Multimodal features allow learners to actively engage in the stories — in a given activity, they may, for example, count the number of suns (as illustrated in Fig. 1), listen to a word and then tap on its image, and exercise with animal characters working out in real-life landscapes. For example, in Fig. 1 the learners are asked to pinpoint four suns, which exercises their counting and fine motor skills.

CatnClever integrates several strands of research in early childhood learning, namely that play (Scarlett et al., 2004) and storytelling (Casey et al., 2008; Maureen et al., 2022) which are transformative tools for consolidating early literacy, mathematics and other skills. Music (Samsudin et al., 2019) and physical activities also support early learning by giving kids the chance to listen and move their bodies. More recent research has confirmed the value of SEL in the early years (Berlinski et al., 2009; McCormick et al., 2021) to support student well-being and mental health. It also aligns strongly with learning science-based principles of high-quality EdTech products, as app users learn actively, engaged, meaningfully, iteratively, and joyfully (Hirsh-Pasek et al., 2015).

**Figure 1**. Example of CatnClever mathematics activity

Early years are crucial for children's basic skill development (Britto et al., 2017). CatnClever's theory of change rests on the power of stories and regular practice: by offering learning experiences through stories that are authentic, interactive, and culturally and linguistically relevant. CatnClever helps young learners acquire foundational skills that will allow them to thrive in school and throughout life. It follows a comprehensive lesson production process which involves various teams such as educators, content writers, software engineers and arts professionals (actors, visual designers, musicians). Currently the app is tailored to English- & German-speaking markets in a European cultural and visual setting.

With a standards-based competency framework for each subject, the app uses adaptive learning to track progress and personalise lessons to the user's pace, strengths and weaknesses, engaging playfully through encouraging direct and indirect feedback (Kapur & Bielaczyc, 2012; Plass & Pawar, 2020). A data dashboard, launched in the period of data collection, allows parents and care-givers to understand their children's progress and engage actively in their learning (Kent et al., 2022).

In order to minimise game-play interference and maximise play and flow, the game lets users play with the tool without a pre-test or knowledge assessment, and uses scaffolding to situate the learner (Cai et al., 2022). Each activity is a learning opportunity for a learner, whereby they will get feedback on how well they have navigated a particular activity. One challenge is to determine the level of difficulty of each activity. Perhaps some activities in one subject might be harder than others, or perhaps some activities take children a bit more time to answer appropriately. In this study we want to explore whether we could estimate activity difficulty of 170 CatnClever activities in four subjects (i.e., mathematics, language, social and emotional learning, and sport) purely from children engagement data.

# 3 Methods

## 3.1 Setting and context

At present CatnClever has 160K downloads of the app, and 10K active users per month. In this study we specifically focussed on one group of 8,369 learners from Germany who used the app to complete at least five activities in the period between March and October 2023. We specifically chose this group of learners as these were the largest group of learners of the app in that period. In total these learners completed 63.014 activities on their mobile device (Android or iPhone/iPad), and CatnClever provided them with access to 170 unique activities. As mentioned before, CatnClever uses story-based learning approaches, and while each of these activities is unique, they all do follow a particular story line.

## 3.2 Activity Difficulty

In order to assess the activity difficulty of the 170 unique activities available to learners in CatnClever in this respective period when the data was extracted, we used two straightforward indicators: (1) the number of ‘Attempts’ and (2) ‘Time to Complete’ each attempt. Given the challenges of conducting detailed assessments with young children, these metrics provide a practical alternative to assessing their knowledge, skills, and competence to complete a particular activity, and can provide the basis for an “activity difficulty” construct and scale that are based on children’s demonstrated abilities rather than external frameworks. 'Attempts' measured how many tries a child needed to complete an activity correctly, reflecting the effort involved. 'Time to complete' tracks how long a child spent on an activity, shedding light on the time commitment required for each respective activity. The activity difficulty score was calculated by aggregating these data across all learners, following these steps:

1. **Data Collection**: Compile 'Attempts' and 'Time to Complete' for each activity from the app’s user data. This step ensured a robust dataset that reflected real-world usage and learning behaviours by participants.
2. **Log Transformation**: To address skewness and manage discrepancies in scale, we apply the logarithmic function log(x+1) to 'Attempts' and 'Time to Complete'. This transformation reduces the impact of outliers, stabilises variance, and improves the normality of the data, ensuring that zero values are handled appropriately. By doing so, it prepares the dataset for more accurate normalisation and analysis, making subsequent calculations reliable and comparable across different activities.
3. **Normalisation**: Apply min-max normalisation to scale the data uniformly. This approach adjusted the data to a common scale, eliminating discrepancies caused by varying ranges in attempts and time, thus ensuring comparability across activities.
4. **Average Calculation**: Determine the difficulty score by averaging the normalised values of attempts and completion time for each activity. This average provided a single, comprehensible metric that encapsulated both the effort and time dimensions of activity difficulty.

In summary, the difficulty of a given activity Aj was computed using the formula:

$$\text{Difficulty}_{A_j} = \frac{\text{Normalized}(\log(\text{Attempts}_{A_j} + 1)) + \text{Normalized}(\log(\text{Completion Time}_{A_j} + 1))}{2} \quad (1)$$

## 3.3 Learner effort

Finally, learner effort was computed using the same fundamental metrics as Activity Difficulty—'Attempts' and 'Time to Complete'—but applying them on an individual level rather than aggregating across all users. For each activity completion by a learner, the following process was utilised:

1. **Normalisation of Individual Metrics**: Each learner’s attempts and time to complete were log-transformed and normalised against the activity-specific aggregated ranges, ensuring that the effort measurement was adjusted for the inherent variability in each activity's difficulty.
2. **Effort Calculation**: Learner Effort was then calculated by averaging the normalised attempts and completion time values. This metric reflected the individual effort expended by the learner to complete the activity, distinct from the collective difficulty faced by all learners. For instance, the following formula was used to compute learner effort:

$$\text{Learner Effort}(L_i, A_j) = \frac{\text{Normalized}(\log(\text{Attempts}_{jL_i,A_j} + 1)) + \text{Normalized}(\log(\text{Completion Time}_{L_i,A_j} + 1))}{2} \quad (2)$$

### 3.4 Learner performance

Learner performance was calculated using the metrics previously discussed for learner effort and activity difficulty. Given the complexity of learning environments and individual differences, our approach to quantifying performance focused on the efficiency with which learners engaged with the activities. Accordingly, learner performance for a specific learner \(L_i\) on a specific activity \(A_j\) was computed using the following formula:

$$\text{Performance}(L_i, A_j) = 1 - (\text{Effort}(L_i, A_j) \times \overline{\text{Difficulty}_{A_j}})$$

(3)

Where:

- Effort (Li, Aj) refers to the learner effort outlined in the previous section (see Section 3.3).

- \overline {Difficulty (Aj)} denotes the average activity difficulty, as detailed in Section 3.2.

This calculation frames performance as inversely related to the product of effort and difficulty, based on the assumption that higher performance is achieved when a learner can complete more challenging activities with less effort. This method allows us to identify potential patterns of learning effectiveness and to tailor further educational content to better suit the evolving needs of learners.

### 3.5 Data Analysis

The data analysis process comprised several key steps, including data extraction, preprocessing, exploratory data analysis, temporal analysis, and visualisations. Below, we provide more details on how these methods were systematically applied to assess learner engagement and activity difficulty.

Our exploratory data analysis began with initial visualisations to identify outliers in the attempts and completion time metrics, crucial for ensuring data integrity. After outlier removal, we applied normalisation and log transformations to these metrics, enabling uniform comparisons across various activities. This preparation facilitated the calculation of a unified activity difficulty score that accurately reflects the combined demands of effort and time for each activity. In line with recommendations by Alcock et al. (2024) and Kaliisa et al. (2024) we shared the initial insights back with the developers and engineers of CatnClever. Over a period of two months bi-weekly online data sessions were held over Zoom to discuss the initial insights and interpret the findings, and where needed further fine-tune the visualisations and learning analytics approaches. Further investigation involved analysing learner effort across normalised time periods using line plots, which highlighted both general trends and subject-specific variations in effort. This detailed examination helped us understand the dynamics of learner engagement throughout their engagement with the app.

To support this analytical process, we exported our datasets into CSV files for further processing. Subsequent analysis was conducted using Python within Jupyter Notebooks (version 8.20.0), providing an interactive environment for exploratory work. We employed several Python libraries: Pandas (version 2.1.4) for data manipulation, Matplotlib (version 3.8.2) and Seaborn (version 0.13.1) for data visualisation, and NumPy (version 1.26.3) for numerical computations.

## 4 Results

Our dataset comprised 60,279 entries across 170 unique activities, capturing metrics on learner performance, effort, and activity difficulty. In terms of time to complete, on average children spent 90.25 seconds per activity (range 10.21-1199.18). Obviously some activities required more time for children to complete due to the design of a particular task (e.g., watching a longer video with follow-up activity) relative to other activities. Furthermore, on average 1.16 (range 0-10) attempts were made by the children.

Table 1. Descriptive statistics of learner performance, effort, and activity difficulty (n = 60,729)

| | Learner Performance | Learner Effort | Activity Difficulty |
|---|---|---|---|
| Mean | 0.86 | 0.37 | 0.37 |
| Std Dev | 0.06 | 0.14 | 0.05 |
| Min | 0.42 | 0.00 | 0.14 |
| 25% | 0.83 | 0.25 | 0.35 |
| 50% | 0.85 | 0.39 | 0.39 |
| 75% | 0.91 | 0.44 | 0.40 |
| Max | 1.00 | 1.00 | 0.64 |

After transforming and normalising the data as described in section 2, the average activity difficulty and learner effort scores both stand at 0.37. This level, on a scale from 0 to 1 – where 0 represents the lowest difficulty and 1 the highest – suggests that the activities are moderately challenging: neither too easy to solve effortlessly, nor too difficult to discourage engagement. The average scores for activity difficulty ranged from 0.14 to 0.64, indicating a broad spectrum of challenges, which can accommodate diverse learner capabilities.

Learner performance showed a high average of 0.86, with a relatively narrow standard deviation of 0.06, pointing to generally high achievement levels among the learners. In essence, a performance score of 0.86 reflects that learners, on average, are able to complete challenging activities while expending

relatively low effort. The minimum and maximum performance scores were 0.42 and 1.00, respectively, highlighting a spectrum of outcomes that, while skewed towards high achievement, also includes underperformers (see Fig. 2).

Figure 2. Distribution of performance, effort and difficulty

## 4.2 Analysis of activity difficulty

As illustrated in Fig. 2, our analysis highlighted significant variability in the difficulty levels across different GBL activities. Fig. 3 illustrates the activities that learners found to be the most challenging as well as those that were relatively easier for learners, followed by a discussion on the average difficulty across different subjects.

**Most challenging activities**: The activities with the highest difficulty scores primarily comprised puzzles and movement and sport tasks, which demand critical thinking and problem-solving skills. The activity labelled **activity_empathy17_9** emerged as the most difficult, with a difficulty score of 0.64. This score, on a scale from 0 to 1 where 1 represents the highest level of difficulty, indicates that this activity requires significant time and/or attempts relative to other activities in the dataset (see Section 3.2 for details). It was closely followed by **activity_language1** and **activity_training1**, both scoring 0.59. Other notably difficult activities included various puzzles such as **activity_puzzle18_6** and **activity_puzzle12_8**, each with difficulty scores exceeding 0.58. These high scores suggested tasks that likely required higher cognitive engagement and sustained concentration from learners.

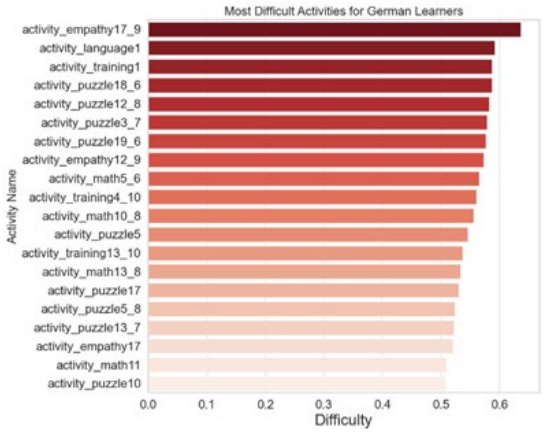



Figure 3. Most challenging activities

Easiest activities: Conversely, as illustrated in Fig. 4 the activities with the lowest difficulty scores were predominantly focused on empathy and language development, suggesting that these might be more accessible or less demanding. The easiest activity identified was activity_empathy3_10, with a difficulty score of only 0.14, significantly lower than the others. This was followed by activities like activity_empathy14 and activity_empathy5, which had difficulty scores of 0.16 and 0.16, respectively. These activities might involve simpler interactions and could be designed to foster basic skills in younger learners or those new to the subject matter.

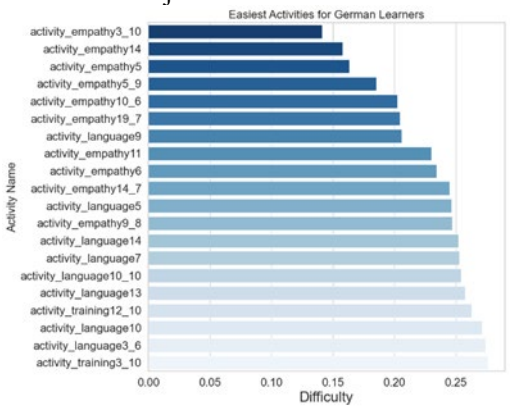



Figure 4. Most easy activities

**Difficulty variation by subject**: When examining the average difficulty of activities by subject, as illustrated in Fig. 5 mathematics was found to be the most challenging, with an average difficulty score of 0.45. This is indicative of the complex problem-solving and logical reasoning required in mathematics. In contrast, Empathy activities were the easiest on average, with a score of 0.34, possibly reflecting their focus on social and emotional learning rather than problem-solving reasoning. Language and Movement and Sport subjects were moderately challenging, scoring 0.36 and 0.37, respectively.

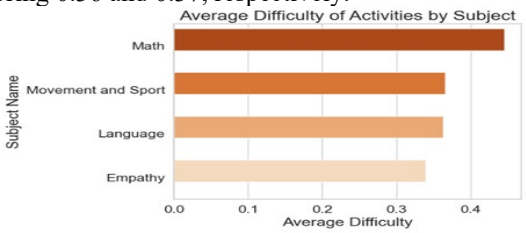



Figure 5. Difficulty by subject

## 4.3 Analysis of learner engagement

Finally, our analysis of learner engagement over the normalised period from initial to final activity presented a detailed view of changes in effort exerted by learners (See Fig. 6). The normalised time scale, ranging from 0 (the start of the engagement period) to 100 (the end), facilitates a direct comparison of effort levels regardless of the absolute length of each learner's engagement period. As some learners only

engaged in a couple of activities, while others completed all 170 tasks this normalised time scale provides an insight into the overall learning effort of learners over time.

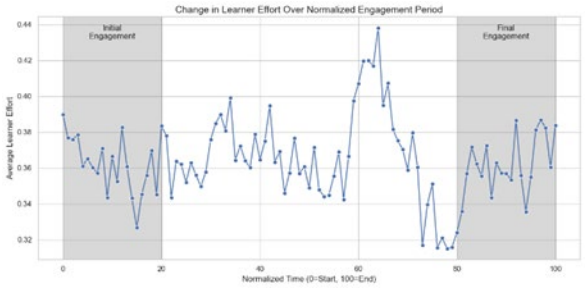



Figure 6. Overall learner effort over time.

At the outset of the engagement period (Normalised Time = 0), the average learner effort was observed at approximately 0.39 on a scale from 0 to 1, where 0 indicates minimal effort and 1 the highest (see Section 3.3). This score reflects moderate initial effort in terms of number of attempts and time needed to complete activities. In addition, this initial effort slightly decreased in the early stages, reaching a lower point around 15% of their engagement time, where it dropped to about 0.34. Interestingly, as learners progressed towards the midpoint of their engagement period, a gradual increase in effort was noticeable, with a peak around 60% of their engagement, recording an effort level above 0.42.

This peak suggests an increase in engagement or a possible response to more challenging or engaging content as the learners approached the middle of their GBL timeline. Following this peak, there was a notable drop in effort, indicating a potential engagement with easier activities. However, as learners neared the end of their engagement period (after instant 80), their effort exhibited slight fluctuations but remained comparatively high, ending with an average effort of about 0.38.

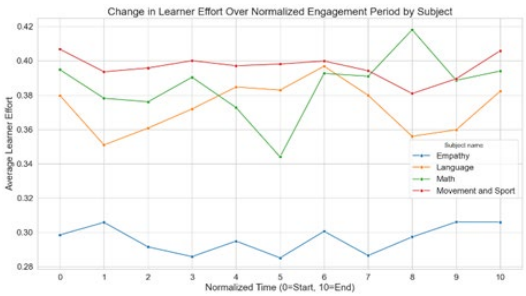



Figure 7. Effort over normalised time by subject

**Learner effort over time by subject:** Fig. 7 illustrates the variation in learner effort across subjects over normalised time. 'Movement and Sport' consistently shows higher effort, likely due to the physical demands of its activities. mathematics peaked notably at instant 80, suggesting engagement with complex or cumulative reviews. Language effort increased in the middle and late stages, potentially explained by the introduction of advanced topics. In contrast, empathy exhibited consistently lower effort throughout, indicating less cognitive or physical demand in its activities.

# 5 Discussion

This learning analytics study delved into a largely uncharted territory to predict learning performance without formal assessment data of game-based learning (GBL). As widely discussed in the literature on GBL with young children (Banihashem et al., 2023; Guan et al., 2024; Kucirkova et al., 2024; Maureen et al., 2022; Sun et al., 2023), there are obvious ethical and methodological challenges of frequent and repeated testing of activities by children. Therefore, we explored whether we could estimate activity difficulty, learner effort, and performance across 170 activities in CatnClever using relatively unintrusive engagement data from log data from 8,365 German preschoolers.

Our analysis indicated significant variability in activity difficulty across different GBL activities that are embedded in CatnClever. Mathematics emerged as the most challenging subject overall, followed by language, movement and sport, and empathy. Learner engagement over time exhibited substantial fluctuations, with a notable increase in effort around the midpoint of the engagement period in CatnClever, possibly in response to more challenging content.

The findings of this study underscore the importance of understanding young learners' engagement patterns in GBL environments. The metrics developed during this exploratory study illuminated significant insights into the engagement and success of learners with the educational content provided. The data revealed a strong alignment between the difficulty of activities and the effort learners put in, with both averages mirroring each other. This synchronisation suggests that the activities might be well-tuned to the learners' capabilities, contributing to the high average performance observed. Further investigation could focus on the outliers to better understand the factors affecting lower performance and engagement levels.

The broad range of activity difficulty may reflect a pedagogical strategy designed to reach learners with varied skill levels. Activities categorised under empathy are intentionally less complex, which seeks to consolidate the development of foundational, and contribute to the children's sense of self-efficacy and self-awareness. In contrast, mathematics activities seem to incorporate more intricate puzzles, potentially aimed at fostering early analytical and problem-solving capabilities. This varied difficulty range can contribute to a balanced curriculum that supports both cognitive and emotional growth. Additionally, language and body movement activities presented moderate difficulty levels, potentially promoting language skills

and physical coordination, which are essential for holistic development in preschoolers.

The alignment between activity difficulty and learner effort suggests the importance of providing optimal challenge levels to maintain engagement and foster learning. The pattern observed suggests that learner engagement did not remain constant but varied, possibly in response to curriculum complexity, personal adaptation to the learning activities, or external factors affecting learner motivation and capacity to engage. Highlighting specific periods such as the initial, middle, and final phases, we identified where potential interventions or motivational strategies might be most effectively applied to sustain or enhance learner engagement. Moreover, the identification of key engagement moments can inform the design of tailored interventions to enhance learning outcomes. By leveraging learning analytics and artificial intelligence, educators and developers can gain valuable insights into young learners' interactions with GBL apps, enabling the refinement of educational strategies to better meet the needs of preschoolers. However, ethical considerations regarding data collection and analysis in GBL environments warrant careful attention.

# 6 Limitations and next steps

Obviously a first limitation of our study is that we did not objectively and independently measure the actual activity difficulty of the 170 activities, the inter-related story-based approaches, and the academic performance of the children. At the moment we are working together with pedagogical and disciplinary experts to align whether the identified difficulty levels based upon engagement data correspond with experts' perspectives. A second limitation is that we did not directly ask the children about their lived experiences, and whether (or not) the measured activity difficulty resonated with them. Finally, as we only focussed on a sub-group of German children, future research should explore whether children from other school systems and cultural contexts experienced similar engagement patterns. In future research we aim to explore the relationships between the various stories embedded in CatnClever and user activity/outcomes. Furthermore, we aim to create and validate a tool integrating GenAI and human-led tasks to adapt our learning app CatnClever culturally and linguistically, for children from underserved regions with French and Spanish as language of instruction. Replicable for other languages, the tool will help scale content production and user base, and collect cross-cultural data on early learning. Alliances with the Open University and the Society of Learning Analytics Research will enable product improvement and data accessibility for external researchers.

## Acknowledgments

We are grateful for the support by the Open University EdTech Foundry grant and CatnClever.

# References

Alcock, S., Rienties, B., Aristeidou, M., & Kouadri Mostéfaoui, S. (2024). *How do visualizations and automated personalized feedback engage professional learners in a Learning Analytics Dashboard?* Proceedings of the 14th Learning Analytics and Knowledge Conference, Kyoto, Japan. https://doi.org/10.1145/3636555.3636886

Banihashem, S. K., Dehghanzadeh, H., Clark, D., Noroozi, O., & Biemans, H. J. A. (2023). Learning analytics for online game-Based learning: a systematic literature review. *Behaviour & Information Technology*, 1-28. https://doi.org/10.1080/0144929X.2023.2255301

Battistella, P., & von Wangenheim, C. G. (2016). Games for teaching computing in higher education–a systematic review. *IEEE Technology and Engineering Education, 9*(1), 8-30.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of public economics, 93*(1), 219-234. https://doi.org/https://doi.org/10.1016/j.jpubeco.2008.09.002

Bishop, R. S. (1990). Windows and mirrors: Children's books and parallel cultures. California State University reading conference: 14th annual conference proceedings,

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., Perez-Escamilla, R., Rao, N., Ip, P., Fernald, L. C. H., MacMillan, H., Hanson, M., Wachs, T. D., Yao, H., Yoshikawa, H., Cerezo, A., Leckman, J. F., & Bhutta, Z. A. (2017). Nurturing care: promoting early childhood development. *The Lancet, 389*(10064), 91-102. https://doi.org/10.1016/S0140-6736(16)31390-3

Cai, Z., Mao, P., Wang, D., He, J., Chen, X., & Fan, X. (2022). Effects of Scaffolding in Digital Game-Based Learning on Student's Achievement: a Three-Level Meta-analysis. *Educational Psychology Review, 34*(2), 537-574. https://doi.org/10.1007/s10648-021-09655-0

Casey, B., Erkut, S., Ceder, I., & Young, J. M. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology, 29*(1), 29-48. https://doi.org/https://doi.org/10.1016/j.appdev.2007.10.005

Chen, X., Zou, D., & Xie, H. (2022). A decade of learning analytics: Structural topic modeling based bibliometric analysis. *Education and Information Technologies, 27*, 10517–10561. https://doi.org/10.1007/s10639-022-11046-z

Eades, J. (2005). *Classroom tales: Using storytelling to build emotional, social and academic skills across the primary curriculum*. Jessica Kingsley Publishers.

Egan, K. (2005). *An imaginative approach to teaching*. John Wiley & Sons.

Ellis, G., & Brewster, J. (2002). *Tell it again!: the new storytelling handbook for primary teachers*. Penguin English.

Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology, 51*(5), 1505-1526. https://doi.org/10.1111/bjet.12992

Gee, J. P. (2004). Learning language as a matter of learning social languages within Discourses. In M. R. Hawkins (Ed.), *Language learning and teacher education: A sociocultural approach* (pp. 11-31). Multilingual Matters.

Guan, X., Sun, C., Hwang, G.-j., Xue, K., & Wang, Z. (2024). Applying game-based learning in primary education: a systematic review of journal publications from 2010 to 2020. *Interactive Learning Environments, 32*(2), 534-556. https://doi.org/10.1080/10494820.2022.2091611

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff , R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological Science, 26*(7), 1071-1083. https://doi.org/10.1177/0956797615581493

Kaliisa, R., Misiejuk, K., López-Pernas, S., Khalil, M., & Saqr, M. (2024). *Have Learning Analytics Dashboards Lived Up to the Hype? A Systematic Review of Impact on Students' Achievement, Motivation, Participation and Attitude* Proceedings of the 14th Learning Analytics and Knowledge Conference, Kyoto, Japan. https://doi.org/10.1145/3636555.3636884

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45-83.

Kent, C., du Boulay, B., & Cukurova, M. (2022). Keeping the Parents outside the School Gate—A Critical Review. *Education Sciences, 12*(10), 683. https://www.mdpi.com/2227-7102/12/10/683

Koehnecke, D. S. (2000, 2000 Winter). Increasing literacy through story telling *Reading Improvement*, *37*(4), 187.

Kucirkova, N., Livingstone, S., & Radesky, J. (2024). Advancing the understanding of children's digital engagement: responsive methodologies and ethical considerations in psychological research [Conceptual Analysis]. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1285302

Mahat, J., Alias, N., & Yusop, F. D. (2022). Systematic literature review on gamified professional training among employees. *Interactive Learning Environments*, 1-21. https://doi.org/10.1080/10494820.2022.2043910

Maureen, I. Y., van der Meij, H., & de Jong, T. (2022). Evaluating storytelling activities for early literacy development. *International Journal of Early Years Education, 30*(4), 679-696.

McCormick, M. P., Neuhaus, R., O'Connor, E. E., White, H. I., Horn, E. P., Harding, S., Cappella, E., & McClowry, S. (2021). Long-term effects of social-emotional learning on academic skills: Evidence from a randomized trial of INSIGHTS. *Journal of Research on Educational Effectiveness, 14*(1), 1-27.

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008, 2008//). Story-Based Learning: The Impact of Narrative on Learning Experiences and Outcomes. Intelligent Tutoring Systems, Berlin, Heidelberg.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of Game-Based Learning. *Educational Psychologist, 50*(4), 258-283. https://doi.org/10.1080/00461520.2015.1122533

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education, 52*(3), 275-300.

Roche, M. (2014). *Developing children's critical thinking through picturebooks: A guide for primary and early years students and teachers*. Routledge.

Samsudin, M. A., Bakar, K. A., & Noor, N. M. (2019). The benefits of music and movement in early mathematics. *Creative Education, 10*(12), 3071-3081.

Scarlett, W. G., Naudeau, S., Salonius-Pasternak, D., & Ponte, I. (2004). *Children's play*. Sage Publications.

Sun, L., Kangas, M., Ruokamo, H., & Siklander, S. (2023). A systematic literature review of teacher scaffolding in game-based learning in primary education. *Educational Research Review, 40*(August ), 100546. https://doi.org/https://doi.org/10.1016/j.edurev.2023.100546

# Integrating Cognitive Agents into Visual Novels

**Sandra Sačarić, Markus Schatten**
Artificial Intelligence Laboratory
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
ssacaric20@student.foi.hr,markus.schatten@foi.unizg.hr

**Abstract.** *In this paper we introduce the integration of cognitive agents in visual novels to enhance player interaction and immersion as part of an experimental student project during a Game Development Platforms course. Traditional visual novels rely on predetermined dialogue options, limiting the player's engagement and creating a static narrative experience, while our approach uses cognitive agents for real-time, dynamic conversations. The agent, trained on an extensive question-answer database, improves the realism of in-game dialogues. Despite challenges in linking the cognitive agent with the visual novel, our work marks a significant step towards more interactive and lifelike gaming experiences.*

**Keywords.** cognitive agents, natural language processing, AI integration, visual novel, game development, interactive storytelling, Monogatari engine, JavaScript, Python, Chatterbot

## 1 Introduction

Cognitive agents (Markus Schatten, Đurić, Peharda, and Tomičić, 2022), as sophisticated software entities capable of simulating human-like cognitive functions, have become a pivotal area of research and development in artificial intelligence (AI). Cognitive agents, as opposed to chatbots, have the ability to understand, learn, and interact within complex environments using various cognitive interfaces possibly including visual as well as audio (especially human language) interfaces based on various text to speech (TTS) and speech to text (STT) technologies. Additionally, they usually have an animated avatar making their presence even more realistic. Such characteristics make them ideal for enhancing user experiences across various domains, including interactive digital entertainment. This paper explores the innovative integration of cognitive agents into visual novels, aiming to transform traditional narrative experiences into dynamic, immersive interactions.

The core concept of cognitive agents lies in their capacity for natural language understanding, decision-making, and adaptive learning. Unlike static, hard-coded systems, cognitive agents can process and respond to user inputs in real-time, offering personalized and engaging interactions. These agents operate on datasets, making use of machine learning algorithms to continuously improve their performance and adapt to new scenarios. This approach not only broadens the scope of player interaction but also sets a standard for the application of cognitive agents in narrative-driven game environments.

Visual novels (Cavallaro, 2009) are interactive, narrative-driven video games that combine written text, static or animated graphics, and sound to tell a story, often with branching plotlines and multiple endings based on player choices. They are a popular genre in Japanese gaming culture, known for their rich storytelling and immersive experiences, spanning various genres such as romance, mystery, and fantasy.

The integration of cognitive agents into visual novels presents unique challenges, particularly in terms of ensuring the agents are embedded properly within the visual novel. However, the potential benefits, including more lifelike character interactions and adaptive storytelling, make this a promising area for future exploration. Through this work, we aim to demonstrate how cognitive agents can elevate the interactivity and immersion of digital narrative experiences, making way for more user-centered game design.

Herein we will showcase an experimental implementation of a visual novel game that includes cognitive agents as actors in-game. The integrated cognitive agents are based on the Beautiful ARtificial Intelligence Cognitive Agent (B.A.R.I.C.A.) architecture that we have developed in previous work (Markus Schatten, Đurić, and Peharda, 2021; Markus Schatten, Đurić, Peharda, and Tomičić, 2022; Markus Schatten, Protrka, et al., 2021; Markus Schatten, Marinela Schatten, and Inkret-Martinčević, 2023; Šokec, 2019). For the implementation we have used the Monogatari visual novel engine[1]. The experimental project has been conducted during a Game Development Platforms undergraduate course at the Faculty of Organization and Informatics University of Zagreb in which a team of 12 students

[1] See https://monogatari.io/

have been tasked to develop a visual novel by integrating cognitive agents technology under the guidance of the course lecturer.

The rest of this paper is organized as follows: firstly in section 2 we provide an overview of related work. Then, in section 3 the methodology of implementation is shown. In section 4 we showcase our results and in section 5 we draw our conclusions and provide guidelines for future research.

## 2 Related Work

While there has been numerous applications of chatbots in various gaming contexts (Khan and Sabahat, 2024) including but not limited to chatbots in online communities (Seering et al., 2020), virtual reality games Ren et al., 2020 as well as serius games (Van Rosmalen et al., 2012) there is, up to our best knowledge, no direct precedent for the specific application of cognitive agents in visual novels. Nevertheless, there are some related works that demonstrate the potential of cognitive systems within various forms of entertainment media. Two notable examples are the multimodal chatbot (Bhushan et al., 2020). and the cognitive modeling approach (Funge, 1999).

Bhushan et al. (2020) present No Body lives here (ODO), a sophisticated multimodal chatbot for interactive theater performances. This chatbot integrates vision and natural language processing capabilities, performing tasks like face detection and crowd movement tracking. While it operates within a live theatrical context, its relevance to our research lies in demonstrating how cognitive agents can manage complex, real-time interactions within a narrative framework. However, ODO's focus on multimodal inputs and its movable LED embodiment significantly differs from our work, which emphasizes integrating cognitive agents for natural dialogue and storytelling in a digital visual novel setting.

John Funge's book, "AI for Games and Animation," discusses cognitive modeling for creating autonomous characters in animated and real-time game contexts. Funge's focus on advanced character animation and automated cinematography underscores the role of cognitive modeling in enhancing digital character realism and autonomy. While relevant to our project, his approach centers on animation and gameplay, differing from our work, which uses cognitive agents to generate dynamic dialogue and respond to player inputs in visual novels.

Both works highlight the versatile application of cognitive systems in entertainment media but do not address integrating cognitive agents in visual novels.

Additionally, in our previous work we have used our B.A.R.I.C.A. cognitive agent architecture (overview shown on Fig.1) for the implementation of various application including for university student support (Markus Schatten, Đurić, and Peharda, 2021; Šokec,

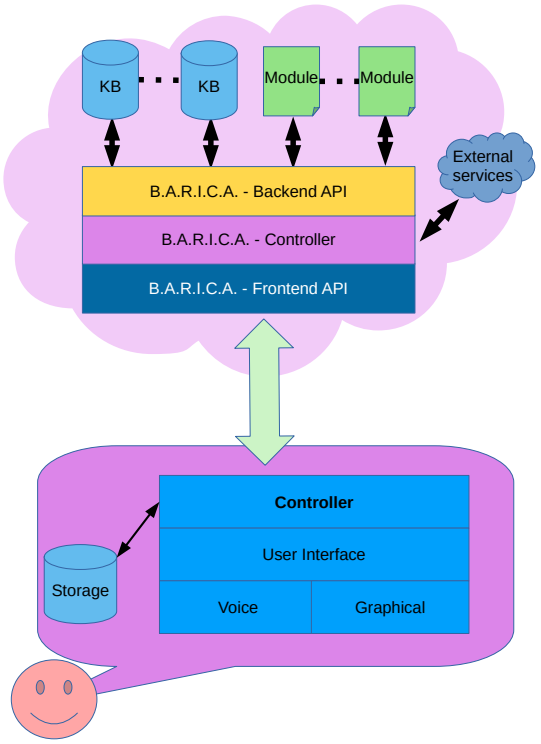



**Figure 1:** The B.A.R.I.C.A. cognitive agent architecture overview (Markus Schatten, Đurić, and Peharda, 2021)

2019), smart mobility (Markus Schatten, Đurić, Peharda, and Tomičić, 2022), telemedicine (Markus Schatten, Protrka, et al., 2021) as well as primary school education (Markus Schatten, Marinela Schatten, and Inkret-Martinčević, 2023). Some of the implementations of these applications are shown on Fig.2

Herein we aim to extend the possible application of cognitive agents to visual novels to transform static dialogue systems in visual novels into dynamic, responsive interactions, introducing unprecedented realism and interactivity in the genre, setting a new standard in interactive storytelling.

## 3 Methodology

At the beginning of a 15 week Game Development Platforms course a team of 12 students had been tasked to develop a visual novel game using the Monogatari engine, incorporating cognitive agents based on the B.A.R.I.C.A. cognitive agents architecture. The project involved creating and preparing a visual novel game where interaction is achieved through speech or menu selections. The game was to be hosted on a server running Debian or Ubuntu OS, and designed in an adventure style. The team needed to create a narrative with alternative endings, design characters (cognitive agents), and implement machine learning mod-

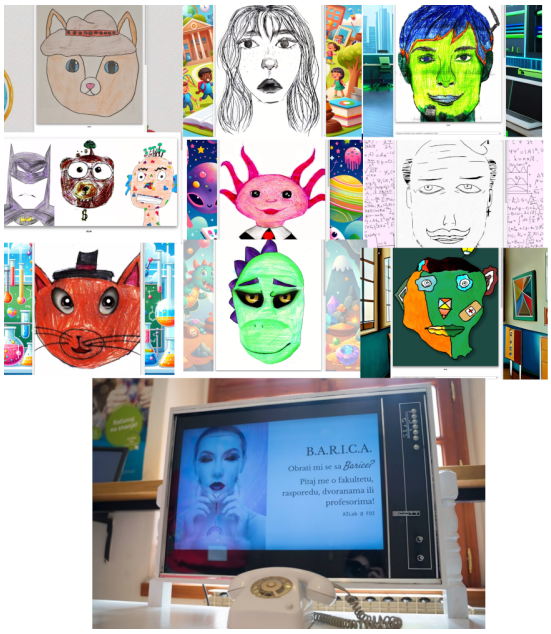



**Figure 2:** Some incarnations of the B.A.R.I.C.A. cognitive agents (Markus Schatten, Marinela Schatten, and Inkret-Martinčević, 2023)

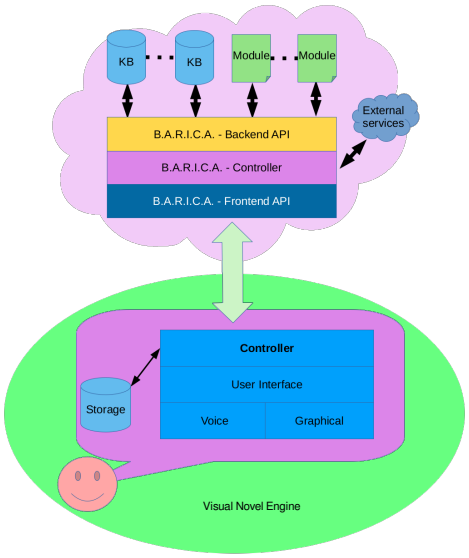



**Figure 3:** Extended B.A.R.I.C.A. architecture

els using the Python Chatterbot[2] module. Character animations should have been created using tools like Crazy Talk[3] or other speech animation applications.

The team's responsibilities included creating a project plan and Game Desing Document (GDD), designing the game's graphical user interface (GUI), developing character animations representing cognitive agents, and ensuring the use of free/open-source assets compliant with licensing. Additionally, the team had to implement, test the game, and develop an update system to pull and install the latest code from the project repository. Comprehensive project documentation detailing all components and usage instructions should have been prepared, with all code and documentation published in the project repository.

Additionally, the team had to elect leaders which were tasked to organize the team into departments, oversee the work on the project as well as to report to "upper management" (the lecturer). In the following we detail their development process of creating and embedding cognitive agents within the visual novel, going over details regarding how the models had been trained, the server communication setup, and integration of the bot with speech recognition capabilities. The newly developed system roughly followed the extended architecture shown on Fig.3.

## 3.1 Game Desing and Storytelling

The storytelling department was instrumental in designing the overall narrative and characters for the visual novel "Betrayed Trust." They crafted an immersive fantasy world where the player, as a guild member, must navigate a complex web of trust and betrayal. The narrative focuses on the player's interactions with their two companions, Bizarr, a mysterious assassin, and Rioth, a fearless warrior. The story has alternative endings, hinging on the player's choices and interactions with these characters. The team emphasized creating a dynamic storyline where the player's decisions influence the narrative's progression, allowing for a personalized experience.

In designing the characters, the team utilized AI-generated visuals to create distinct and memorable protagonists. Bizarr and Rioth (shown on Fig.4), the two main characters, were designed to reflect their complex personalities and roles within the story. Despite initial technological limitations, which prevented more detailed hand-drawn designs, the team managed to create compelling character animations.

## 3.2 Bot Training and Development

Initially, the team utilized the ChatterBot library for training individual cognitive agents. Each agent was tailored with a unique personality by crafting a specific database of questions and answers. This approach ensured that each agent responded distinctly, aligning with its predefined character traits and narrative role within the visual novel.

The dialogue script was finalized by the storytelling department, creating a structured database of potential questions and corresponding responses for each character. This database included not just direct responses but also context-specific variations to reflect the character's personality and role in the narrative.

For each agent, there was a Chatterbot instance con-

[2]See https://pypi.org/project/ChatterBot/
[3]See https://crazytalk.reallusion.com/animator.html

**Figure 4:** Character design generated by AI

figured with custom logic adapters, one tasked with identifying and returning responses based on the user's input, which is determined based on which response matches the input best. The other logic adapter was introduced to handle inputs that didn't confidently match any response. This prompts the user to clarify their input and ensures that the agent's responses are relevant and appropriate regarding the context. This will later be brought up in a bit more detail.

### 3.3 Integration

In order to integrate the cognitive agents into the Monogatari visual novel engine a special Monogatary component was developed that allowed for the inclusion of the client cognitive agent code as characters in the visual novel.

To enable real-time interaction between the cognitive agents and the backend server, the team implemented a WebSocket communication protocol. A dedicated Python class on the Flask[4] based backend side managed WebSocket communication, enabling continuous messaging between the agents and the server. This setup included configuring the ChatterBot instance with custom logic adapters and a SQLite[5] database to store conversation history and manage response variability.

The Python class included methods to initialize a read-only bot instance and handle incoming and outgoing messages. The key features included continuous listening, which kept the bot in a permanent state of actively listening to the surrounding sounds, processing and responding to messages in real-time, as well as message buffering, with the buffer ensuring that responses remained unique by checking for duplicates and maintaining a history of recent interactions.

A Python script was developed to manage the integration of the cognitive agents with web-based functionalities. This script utilized Flask for serving web content and WebSocket capabilities to handle bidirectional communication. The script offered a flexible architecture, supporting both training mode for bot refinement and real-time interaction mode for seamless gameplay experiences.

The script operated in two modes. First, the training mode, meant for training the chatbot and using the provided data. Then the real-time interaction mode, which would initialize the server and Flask application to handle live user interactions.

Additionally, the WebSocket server ran on a separate thread, to ensure that web content was effectively managed via Flask on the main thread. This allowed for efficient multitasking and ensured a responsive user interface.

Integrating speech recognition was a necessary step, in which lied the key to the interactivity of our cognitive agents. Th team implemented this using the JavaScript speech recognition API[6], allowing the system to process user speech and generate appropriate responses.

A JavaScript object was initialized to capture and transcribe spoken inputs. This transcript was then sent to the WebSocket server for processing by the cognitive agents. The script also included handlers for various events as was previously mentioned, one which was triggered when speech was successfully recognized, prompting the bot to respond. The other being used for error handling; managed scenarios where speech recognition failed or returned low confidence, ensuring robust user interaction.

To reach our intended goal of the bot being visually and audibly supported by voice lines and character animations, a dedicated JavaScript function controlled audio and video playback based on user-triggered questions during conversations. This function synchronized audio segments with corresponding video segments, optimizing the timing and duration of multimedia elements based on contextual dialogue within the visual novel. This was achieved by using a single audio and video file containing all recorded answers. Each answer had a manually marked timestamp, allowing a playback of specific segments according to the user's input.

## 4 Results

By the end of the project, the team had successfully developed dynamic and contextually aware cognitive agents that significantly enhanced the interaction experience within the visual novel.[7] These agents were

[4]See https://flask.palletsprojects.com/en/3.0.x/

[5]See https://www.sqlite.org/index.html

[6]See https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API/Using_the_Web_Speech_API

[7]See the final GitHub repository here https://github.com/AILab-FOI/PRRI-CognitiveAgents2024 as well as the corre-

meticulously trained using the ChatterBot library, each equipped with a unique personality and a structured database of potential questions and answers, which ensured distinct and authentic responses. The agents' interactions were made more immersive through the integration of supporting audio and video materials, including voice lines and character animations, adding depth to the user experience. Fig.5 shows the opening screen of the visual novel game[8].

**Figure 5:** Opening screen of the visual novel with a cognitive agent shown

The cognitive agents were embedded within the visual novel through a WebSocket communication system that enabled real-time interaction between the agents and the game environment. This system, managed by a dedicated Python class, enabled continuous messaging and real-time responses, maintaining a fluid conversational flow. Additionally, the agents were enhanced with speech recognition capabilities, allowing them to respond to user inputs in real-time. This feature allowed recognized speech transcripts to be sent to the server, enabling the agents to process and respond dynamically.

Everything was further supported by an audio and video control system, which synchronized multimedia elements with the agents' dialogue based on user-triggered questions. This ensured an engaging user experience, blending visual and auditory components effectively to create a more believable and interactive storytelling environment.

# 5 Conclusion & Future Research

The integration of cognitive agents into visual novels, as demonstrated by the "Betrayed Trust" project, showcases a significant advancement in interactive digital storytelling. By employing cognitive agents capable of real-time dialogue and adaptive responses, the project has elevated the narrative experience beyond traditional static interactions. The use of the B.A.R.I.C.A. architecture and the Monogatari engine provided a robust framework for embedding these intelligent agents, resulting in a more immersive and engaging user experience. The successful implementation of speech recognition and real-time multimedia synchronization further enhanced the interactivity and realism of the game, marking a noteworthy achievement in student-led game development projects.

Despite the project's success, several challenges were encountered, particularly in integrating cognitive agents with the visual novel framework and ensuring smooth real-time interactions. These difficulties highlighted the need for more sophisticated tools and techniques for seamlessly blending AI-driven dialogue systems with traditional game engines. Future research should focus on refining these integration processes, exploring more advanced machine learning models for dialogue generation, and enhancing the scalability of cognitive agents in more complex game environments. Additionally, user feedback on the interactive experience might provide valuable insights for improving the design and functionality of cognitive agents in future projects.

Future research could also investigate the application of cognitive agents in other genres of interactive fiction and gaming. Expanding beyond visual novels to include role-playing games (RPGs), simulations, and other narrative-driven experiences can further demonstrate the versatility and potential of cognitive agents in enhancing user engagement. Moreover, integrating more advanced AI technologies, such as deep learning and natural language understanding, can significantly improve the responsiveness and realism of these agents. By continuing to innovate and explore new applications, the potential for cognitive agents to transform digital storytelling and interactive media remains vast and promising.

# Acknowledgements

This project was carried out under the scope of our faculty, specifically during the Game Development Platforms course, where we explored practical implementations of game development platforms. This was followed through by our mentor, prof. dr. sc. Markus Schatten.

We would like to extend our deepest gratitude to the entire team for their invaluable contributions. Project leaders Jakov Kadić and Sandra Sačarić coordinated tasks and motivated the team, handling documentation, backend support, and visual and content refinement.

Jana Jambrešić handled the UI and made prototypes for the interface, ensuring usability and responsiveness. Luka Pošta created backgrounds, voice lines for the agents, and edited the videos, and also single-handedly made the game's trailer.

Dorijan Kos initially trained the bot and later developed character animations and video materials, while Dora Garafolić designed characters, contributed to the story structure, and assisted with backend integration.

sponding itch.io site of the project here https://ailab-foi.itch.io/prri-cognitiveagents2024

[8]The Fantasy Ren'py GUI template asset pack by Sklolaztika has been used for the user interface https://skolaztika.itch.io/fantasy-renpy-gui-template

From the writing team, Sonja Kolarić shaped the narrative's vision and wrote the story, while Ivan Simić crafted dialogues and implemented narrative elements in JavaScript, ensuring all project requirements were met.

The backend and AI team, comprising Karlo Rosenthal, Frano Šimić, Karmelo Mrvica, and Mateo Zović, collaboratively trained the chatbot and developed server-side solutions, successfully integrating the bot's dynamic HTML display.

Each member's dedication and effort were pivotal in bringing this project to fruition, and their contributions are deeply appreciated.

ChatGPT has been used to to refine the phrasing of parts of this paper.

# References

Bhushan, R., Kulkarni, K., Pandey, V. K., Rawls, C., Mechtley, B., Jayasuriya, S., & Ziegler, C. (2020). Odo: Design of multimodal chatbot for an experiential media system. *Multimodal Technol. Interact.*, 68.

Cavallaro, D. (2009). *Anime and the visual novel: Narrative structure, design and play at the crossroads of animation and computer games*. McFarland.

Funge, J. D. (1999). *Ai for games and animation: A cognitive modeling approach*. A K Peters/CRC Press.

Khan, N., & Sabahat, N. (2024). The role of ai chatbots in revolutionizing gaming experiences-a survey. *VFAST Transactions on Software Engineering*, *12*(1), 93–104.

Ren, X., et al. (2020). Artificial intelligence and depression: How ai powered chatbots in virtual reality games may reduce anxiety and depression levels. *Journal of Artificial Intelligence Practice*, *3*(1), 48–58.

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

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

Schatten, M. [Markus], Protrka, R., et al. (2021). Conceptual architecture of a cognitive agent for telemedicine based on gamification. *Central European Conference on Information and Intelligent Systems*, 3–10.

Schatten, M. [Markus], Schatten, M. [Marinela], & Inkret-Martinčević, L. (2023). Development of gamified cognitive agents for primary schools-a case study. *Central European Conference on Information and Intelligent Systems*, 3–8.

Seering, J., Luria, M., Ye, C., Kaufman, G., & Hammer, J. (2020). It takes a village: Integrating an adaptive chatbot into an online gaming community. *Proceedings of the 2020 chi conference on human factors in computing systems*, 1–13.

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

Van Rosmalen, P., Eikelboom, J., Bloemers, E., Van Winzum, K., & Spronck, P. (2012). Towards a game-chatbot: Extending the interaction in serious games. *The 6th European Conference on Games Based Learning*.

# Public Funding Approaches to Video Game Industry Development in EU Member States

**Nikola Vlahović, Jasmina Pivar, Ivan Celić**
Faculty of Economics and Business
University of Zagreb
Trg J. F. Kennedyja 6, Zagreb
{nvlahovic, jpivar, icelic1}@net.efzg.hr

***Abstract**. The video game industry, with its proven resilience and flexibility, plays a crucial role in mitigating economic shocks in customer and resource markets, as well as overall economic recessions. This unique characteristic has led many countries to allocate public funding to this industry, recognising its potential to bolster national economies during times of crisis. The varying success of different public funding approaches in these countries underscores the industry's economic significance.*
*The goal of the paper is to analyse public funding approaches that have provided successful support for the development of gaming industry over several EU member states and draw the conclusions in what are the best approaches for particular instances.*

**Keywords.** Video game industry, public funding, European Union

## 1 Introduction

One of the most important endeavours of any economic policy is to maintain economic stability both in times of growth and especially in times of recession. Usually, as economic activity wanes, most industries follow due to their complex interrelationships and interdependencies. Several special cases may resist the trends, which can sometimes cushion the undesirable trends as they appear, such as in the case of an economic crisis. Over several previous economic crises and recessions, the gaming industry has drawn the attention of policymakers across various markets as an industry with exactly these strengths. Many countries have perceived these characteristics of the gaming industry as a useful strategy for bridging various economic strifes.

This paper will delve into the approaches of selected European Union member countries that have recognised the gaming industry's potential and implemented various funding mechanisms. These approaches aim to foster a stable and sustainable video game industry, which not only contributes to economic growth during periods of expansion but also serves as a buffer during economic crises, offering a promising outlook for the industry's future.

The goal of the paper is to detect mechanisms that are reliant directly or indirectly on public funds that policy makers have at their disposal and that have been most effective in particular socio-economic situations.

To detect and identify the relevant approaches, an overview of the German, French, Finnish, Croatian, and Polish video game industries, as well as the growth and development of the video game industry, will be considered.

The rest of the paper is structured as follows: Section 2 will briefly overview the video game industry and its particular properties that may explain its flexibility and resilience to disruptions in economic activities. Next, we will describe several distinct cases of countries that have recognised the potential of video game industry development and introduced support through public funds to encourage the establishment of a healthy and prosperous video game industry. Section 4 will discuss the data presented in previous Sections, underlining the most effective approaches. Finally, in Section 5, we will provide conclusions and outline the most important directions for further research of the presented topics.

## 2 Overview of the Video Game Industry

The video game industry is one of the most prominent industries in the knowledge-based economy and a paragon example of a quaternary sector of the contemporary economy (Zackariasson & Wilson, 2012). It is positioned as the meeting place of a number of creative and knowledge-intensive industries that have been enabled by digitalisation and digital revolution from the late 1990s to the early 2010s (Lyonnet & Rabineau, 2023).

The industry's scope couples various technological aspects with creative and cultural aspects, as well as business strategies, models, and solutions. In synergy,

these elements create practical software products and services that compete in the global market either as entertainment titles, educational tools, training aids (Schobel et al., 2021), or eSports as competitive outlets (Block & Haack, 2020).

From its early beginnings, the gaming industry has influenced the advancement of various technologies and information and communication technologies specifically. Hardware, video processing, human-computer interaction, computer networks, multimedia, streaming, artificial intelligence (Ishwarya et al., 2017), and software agents are just a few of these technologies. It has also enriched the opportunities for artistic talents in graphic design, storytelling and writing authors, various dramatic art talents, etc.

The early niche gaming industry started with the development of physical products in the form of arcades and the first game consoles (Palma-Ruiz et al., 2022). Following the increased availability of computers for personal use, the gaming industry was able to explore new business opportunities and create digital products in the form of game titles. This, in turn, initiated an exponential rise in the development of engaging mechanics and genres of games we have today. An additional boost to the development of the gaming industry was the invention and commercialisation of the Internet. It enabled one of the first genuinely fully digitalised life cycles of digital products (i.e. games), including both development, distribution and consumption, gaining increasing market share through this process.

As of 2018, there are 2 billion 341 million active video game players worldwide (Baltezarevic, 2018) with increasing trends additionally fuelled by the recent COVID-19 pandemic (Block & Haack, 2020). The global video game market is expected to reach US$282.30 billion in 2024, with average annual growth rate of 8,76% until 2027. The projected annual user penetration rate in the same period is between 15% to 18% (Alda et al., 2024).

Many countries and international professional organisations have acknowledged various disciplines through annual award distinctions, both by economic parameters and artistic value, to promote and encourage the development of games with additional value for players.

As we mentioned earlier, one of the most interesting properties is that the video game industry is a knowledge-based digital industry. This means that the main resources for the development of various software solutions are creative artists and proactive engineers and programmers proposing novel and ingenious solutions that are pushing the limits of currently available technologies (Politowski et al, 2021). Under appropriate socio-economic conditions, these experts can emerge in various countries, connect through globalised digital infrastructure, and access global markets without limitations of geographical locations. This is one of the main factors that enable the video gaming industry to be resilient to economic cycles and have an overall stabilising effect. This is even more pronounced in times of economic crises.

Any national economy can invest in creating the appropriate environment that can form a solid basis for such a knowledge-based industry on the same terms as any other country. The effect this may provide for the overall national economy may differ depending on the size, characteristics, structure and maturity of existing industries in each country. This is why, during the digital transition of national economies, a number of countries pioneered new approaches and strategies for establishing national gaming industry as the core of their entertainment industry. For this purpose, they employ various mechanisms, such as subsidies, grants, tax exemptions, etc. These measures achieve their goals in varying success rates depending on the circumstances of the national economy.

In the next Section, we will present the German, French, Finnish, Croatian and Polish video game industries and funding approaches to creating a positive entrepreneurial climate, with the goal of establishing a solid basis for the development of their respective gaming industries.

# 3 Overview of several selected EU countries

This chapter describes the German, French, Finnish, Croatian and Polish video game industries and related public funding to detect and identify the most relevant approaches to the video game industry.

## Germany

Germany's game industry market is the largest in Europe. According to Germany Trade & Invest (2021) around 750 companies publish and/or develop video games. Furthermore, around 60% of the German population plays video games, so its market provides opportunities for international investors in segments such as online games and in-app purchases. Since the number of smartphone devices grows constantly, there has also been a steady increase in mobile device gamers. In 2020, regarding games market revenue, in-app purchases were a leading business model. Gamers doubled the demand for online services for PC, console, and smartphone games. German gaming industry employs more than 10 thousand people. These people are trained at 109 public educational facilities offering game development programs (or 1,3 facilities per 1 M citizens.

Developer and publisher gaming in Germany is growing around major cities, with the Rhine-Ruhr region, Berlin, and Munich being the most important German gaming clusters. All these clusters have local success stories and the presence of prominent international gaming companies and gaming education and training institutions.

Start-ups also recognise the country as a preferred location for launching video game studios. This is due to moderate costs, supporting infrastructure, and access to talent. Germany is a focal point for the European and international gaming business communities.

Video game studios use the Creative Europe programme by the European Commission, particularly the Support for the Development of European Video Games subprogram. However, Germany supports the industry with the Federal German Games Funding Programme and 50 million euros annually for video game developers active in the country. The aim is to reinforce the country's position as a global hub for game development and enhance its international visibility. Funding is provided as a grant to prototype and development projects. Prototype development costs must be between 30 and 400 thousand euros, while production development costs must be at least 100 thousand euros. Around 190 projects are part of the initiative, with over 100 projects already completed (Federal Ministry for Economic Affairs and Climate Action, 2022). This complements two other supporting instruments for the computer and video games industry, the German Video Game Awards and the Federal German Games Funding Programme, launched in 2019.

Moreover, it is essential to state that federal states fund the video gaming industry. Funding approaches vary according to the individual funding organisation. Funding is usually available to developers and local, federal, and state project activities.

Germany offers cultural, R&D, and SME launching aids for the video game industry. It also provides start-up support and infrastructure for games development and hubs, with incubation and accelerator programmes. Germany offers various funding incentives, R&D grants, loans and tax credits, interest-reduced loans, loans and equity financing for launching a new game development studio, and access to regional funding and guarantees. The total amount of gaming funding exceeds 80 million euros annually.

## France

France is a leader among Member States. It has excellent educational institutions and a strong system of public funding that supports the video game industry. Since the 1980s, France has been the cradle of the European video game industry, with Ubisoft and Infogrames being the most important studios. It is part of the France-Germany-Poland axis, which employs most talent in video game development. Paris Region has the highest number of video game companies in France. Some clusters and incubators supporting the video game industry's growth, such as Capital Games, Cap Digital, C-19, and Ubisoft Entrepreneurs Lab, are also placed in the Paris Region. Furthermore, the Paris Region offers support incentives for the creation of video games.

According to Syndicat National du Jeu Video (2023) in 2022, the French video game industry comprised more than 1 thousand companies. The French regions with the largest number of video game companies are Ile-de-France and Auvergne-Rhone-Alpes. Game development studios mainly develop indie and AA types of video games. Regarding business models, the most popular are paid upon purchase with no additional content and paid upon purchase with additional content. Most developed games are distributed on PC, Steam, Epic, and alternative PC stores. Self-financing and video game tax credits are French video game studios' most popular financing mechanisms. Grants via the European Creative Media programme are the least popular. Regarding the popularity and attractiveness of France for the video game industry, 85% of companies consider France an attractive territory in 2022. The most important attractiveness factors are tax incentives and the quality of training provided by 130 educational institutions.

Regarding tax incentives, France has an efficient tax credit with almost 70 million euros in annual funding. Tax credits are available as part of cultural state aid for video game production and R&D aid supported by the French National Centre of Cinema. Its Fond d'Aida a la Creation de Jeu Video is in advance funding for video game projects. With a 30% tax credit for production expenditure, the French Video Games Tax Credit (CIJV) is one of the most competitive countries for the video game industry. It has supported around 400 projects since 2008. Applied projects must pass the cultural test to be eligible. Moreover, the Video Games Support Fund includes a Fund for Original Creation, Prototype Production, and Collective Actions Fund. The Centre also provides grants for video game pre-production and R&D. Many other tax provisions and public funding approaches are used to support the industry, such as an innovation tax credit of 20%, an R&D tax credit of 30%, grants, R&D loans, loans and equity funding for small and medium enterprises, equity funding for SMEs, and Young Innovative Company status.

## Finland

Rovio's and Supercell's success started the growth of the Finnish game industry in 2012. According to Hiltunen et al. (2023), by turnover, the Finnish video game industry is among Europe's top five largest national game industries. It surpassed 3,000 million euros of turnover in 2021. By the end of 2022, there were more than 230 active game development studios and employed 4100 people. It is a European leader in mobile game development, which is the primary driver of the industry's growth. Finland also leads in the number of facilities offering game development programmes by having 6,7 educational facilities per 1M citizens. With almost half of active Finnish game studios, Finland's most significant game industry hub

is the capital region, which consists of Helsinki, Espoo, and Vantaa. Interestingly, mobile is no longer as important platform as it used to be since PC is now nearly as popular as mobile.

The industry's success in Finland is built on a supportive ecosystem. Finnish public funding for the game industry is R&D-oriented, offering grants and loans to studios. However, nowadays, the range of support extends to the industry, such as technology development and marketing. Cultural funding approaches that are common in many other European countries are scarce. Grants are generally available but are dedicated to more than just game development studios. An exception to this is DigiDemo funding. Regarding public R&D funding, Business Finland is the leading funding source for Finnish studios. It has supported the Finnish video game industry through various approaches. Business Finland funding is dedicated to companies registered in Finland. It comes in the form of a grant or a loan. There is also Talent funding that can be used to create new jobs and practices that support the ability to recruit international talent. Furthermore, the financing company Finnvera supports the industry with guarantees and loans. Moreover, DigiDemo by AVEK and the Ministry of Education and Culture supports the development of new digital content, including video games. Therefore, the most important funding sources for the Finnish video game industry include Business Finland, DigiDemo and EU funding.

## Croatia

The Croatian video game industry is growing, with 65 game developer studios and 326 employees as of the end of 2022. Seven educational institutions offer game development programs.

An example of a government incentive for the development of the Croatian video game industry is the project Game Industry Center in Novska led by the Regional Coordinator of the Sisak-Moslavina County, with partners Sisak-Moslavina County, Development Agency SIMORA and the City of Novska (Central Finance and Contracting Agency, 2024). The agreement is related to releasing 26.3 m euros from the EU's Just Transition Fund for a Game Industry Center. The Game Industry Center will include an incubator with a hall for testing video games, a gaming campus, a five-year higher education programme and a high school programme. The City of Novska has been profiled as a centre for the Croatian gaming industry.

In 2023, funding for the video game industry was significantly increased through the National Plan for Recovery and Resilience, which had a budget of 33.2 m EUR for cultural and creative industries, including the video game industry. Furthermore, game developer studios in Croatia can receive grants from the Croatian Audiovisual Centres as part of cultural state aid for video game production. Croatian public gaming industry funding is also R&D oriented, offering studios grants, loans, and equity funding. Moreover, the Ministry of Culture - Entrepreneurship in Culture provides aid in the form of grants for launching a new studio. Equity funding is also available for a new studio.

## Poland

In 2023, the Polish video Game industry had around 500 game producers and publishers and a workforce of more than 15,000 people, more than the German video game industry. It is important to state that due to the Russo-Ukrainian War, the video game industry in Poland is experiencing an influx of talent from Ukraine. According to Marszalkowsk et al. (2023) the Polish video game industry's success is due to studios' significant efforts in exporting their titles and translating them into more than two languages. Exports account for 96% of Polish gaming revenues. Poland offers 65 game-related degree courses at 52 universities. The Witcher franchise is the most famous Polish video game franchise, released in 2007. Inspired by Central Europe's legends and myths, the Witcher is an excellent example of how the video game industry can help promote the European cultural identity, history, and heritage of different countries.

Some supporting institutions for the Polish game industry are The Polish Games Association and The Indie Games Poland Foundation. The government supports the industry via The Polish Investment and Trade Agency. It supports establishing business relations between entrepreneurs and enables companies to reach their full potential in exporting their products and services. It also supports potential investors in Poland by providing information on legal aspects, tax and available financial incentives. The Polish Investment Zone operated by the Kraków Technology Park should also be mentioned in relation to public funding approaches in Poland. It is a form of support for businesses planning new investments. Companies investing in PIZ are eligible for 12 and 15-year tax relief packages for investments.

Moreover, Polish Ministerstwo Kultury i Dziedzictwa Narodowego offers grants, loans, and tax relief incentives as cultural aid for the video game industry. Game developer studios can use funding for R&D activities through grants, loans, tax credits, and equity funding. Grants, loans, tax credits, and equity funding are also available to aid in launching new game development studios.

# 4 Discussion

Tables 1 and 2 clearly and structuredly present the video game industry in Germany, France, Finland, Croatia, and Poland. The first table summarises the industry's economic data, while the second outlines the available approaches to public funding of the video game industry in these countries. Table 1 contains data

on 2021 and 2020 since, for the later years, there are not yet available data on all of the selected countries.

**Table 1.** Video Gaming Industry Economic Data for Germany, France, Finland, Croatia and Poland in 2021 and 2020

| Video game industry | DE | FR | FI | HR | PL |
|---|---|---|---|---|---|
| Turnover (mil. eur) 2021 | 3740 | 3880 | 3200 | 64,4 | 1191 |
| Turnover (mil. eur) 2020 | 3600 | 2860 | 2400 | 79,3 | 1335 |
| No. of video game dev studios 2021 | 717 | 1147 | 215 | 65 | 439 |
| No. of video game dev studios 2020 | 591 | 780 | 200 | 45 | 424 |
| Size of workforce 2021 (in k) | 10,9 | 18 | 3,55 | n/a | 12,11 |
| Size of workforce 2020 (in k) | 10,07 | 15 | 3,35 | 0,33 | 9,71 |
| No. of edu facilities | 107 | 130 | 42 | 7 | 52 |

**Table 2.** Available Public Funding for the Video Game Industry in Germany, France, Finland, Croatia and Poland

| Type of Public Funding for Game Dev. Industry | | DE | FR | FI | HR | PL |
|---|---|---|---|---|---|---|
| Cultural | Grants | + | + | + | + | + |
| | Loans | | + | | | + |
| | Tax credits | | + | | | + |
| | Equity funding | | + | | | |
| R&D | Grants | + | + | + | + | + |
| | Loans | + | + | + | + | + |
| | Tax credits | + | + | | | + |
| | Equity funding | | + | | + | + |
| SME Launching | Grants | + | + | + | + | + |
| | Loans | + | + | | | + |
| | Tax credits | | + | | | + |
| | Equity funding | + | + | | + | |

Table 2 aligns with the conclusions of the European Game Developers Federation's report (2023) about the video game industry's public funding approaches.

Public authorities need to be made aware of the video game industry, which is why there are few parliamentary activities within European bodies. For the development and visibility of the gaming industry in the European Union's member states, proactive public funding approaches to support companies, promoting collaboration between EU countries, and creating a framework that will prevent talent from leaving the European Union are essential.

Public support schemes for video game development at the European Union level exist. Examples include the Horizon Europe research programme and the Creative Europe Programme MEDIA chapter. However, making these incentives visible is still necessary because resources are not used at a sufficient level. Different bodies are working with European and national institutions on behalf of video game industry stakeholders. An example is the European Games Developer Federation, which represents more than 2500 game development studios. Video game unions promote industry and act as representatives in Germany, France, and Finland, such as GAME in Germany, SNJV in France, and Neogames in Finland. Furthermore, regional associations such as French Capital Games and Push Start and Game Only help to structure the local video game industry.

Significant disparities exist between countries regarding public funding approaches to the video game industry. However, it can be noticed that the video game industry is susceptible to the effects of tax credits. Therefore, Member States should consider that when designing their policies regarding developing the video game industry. Also, reliable and easy-to-anticipate schemes need to be developed which are not complex and do not require considerable administrative effort. Additionally, countries should introduce priority grants to finance the work of various stakeholders and partners to open up the sector to partnerships.

Therefore, we recommend the following. Governments of Member States must know the video game industry and its value. Frameworks conducive to the industry's development should be developed. The video game industry should also be used to construct a solid European and state identity. Specialised networks such as French SNJV can disseminate good practices and present efficient public funding approaches. National funds should be available, especially public R&D and cultural financing.

# 5 Conclusions

In this paper, we presented the results of a comparative analysis of different funding approaches for supporting the development of the gaming industry by

selected member states of the European Union. As previous research and statistics testify, the contemporary gaming industry is highly resilient to economic shocks, which makes it valuable asses in the structure of the national economy as a buffer in the face of economic instabilities. To support the development of this industry, various countries have opted for different approaches to create and maintain favourable business environments for gaming entrepreneurs, developers, publishers and providers of related services and dependant industries.

The paper's goal was to investigate and analyse funding approaches used to support the development of video gaming industries in several European Union countries, such as Germany, France, Finland, Croatia, and Poland.

We have shown that the selection of the publicly funded measure is highly dependent on the current state, strengths, and maturity of the national economy, the strategic positioning of the gaming industry, and the awareness of its full potential.

Key determinants are to raise awareness of the flexibility and strengths of nurturing the national gaming industry sector, not only for the exceptional economic benefits but also for the potential to activate and maintain the cultural heritage that can be used to promote national brands and international recognisability.

The research's limitation is that it focused on the video game industry and public funding approaches of only five countries in the European Union. As a recommendation for future research, it is suggested that the study be expanded to include the video game industry of all EU countries. Additionally, the research could be extended to include private funding approaches for the video game industry.

# References


Alda, M., Mega, Y., Mizuno, K., Biagi, L., Dulin, C. (2024, July 3). *Video Games – Worldwide* STATISTA. https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide

Baltezarevic, R., Baltezarevic, B., Baltezarevic, V. (2018) The Video Gaming Industry (from play to revenue). International Review Faculty of Business Economics and Entrepreneurship, No. 3-4, doi:10.5937/IntRev1804071B.

Block, S. & Haack, F. (2020) eSports: a new industry. SHS Web of Conf., Volume 92, 2021, The 20th International Scientific Conference Globalization and its Socio-Economic Consequences, doi:10.1051/shsconf/20219204002

Central Finance and Contracting Agency. (2024, June 5). *Novska is becoming the center of Croatia's gaming industry*. Central Finance and Contracting Agency. https://www.safu.hr/en/novska-is-becoming-the-center-of-croatias-gaming-industry/

European Game Developers Federation. (2023). *2022 European Video Games Industry Insight Report*. European Game Developers Federation – EGDF. https://www.egdf.eu/wp-content/uploads/2024/06/2022-European-video-games-industry-insight-report.pdf

Federal Ministry for Economic Affairs and Climate Action. (2022). *Video Games.* Federal Ministry for Economic Affairs and Climate Action. https://www.bmwk.de/Redaktion/EN/Dossier/games.html

Germany Trade & Invest. (2021). *The Gaming Industry in Germany* [Fact sheet]. https://www.gtai.de/en/invest/service/publications/the-gaming-industry-in-germany-64164

Hiltunen, K., Latva, S., Kaleva, J-P., Tyynela, E. Sauri, L. (2023). *The Game Industry of Finland – Report 2022.* Team Neogames. https://neogames.fi/finnish-game-industry-report-2022/

Ishwarya, T.A.S.K., Naidu, R. C. A., Meghana, K., Reddy, G. P. (2017) A modern approach to design and integrate conceptual methods in video games with artificial intelligence, *Materials Today: Proceedings*, Volume 4, Issue 8, pp 9100-9106, doi: 10.1016/j.matpr.2017.07.265.

Lyonnet, L. & Rabineau, D. (2023) The video games industry in Europe: current situation, issues and prospects., *Shuman paper N724*, 31st October, 2023, Fondation Robert Shuman.

Marszalkowski, J., Biedermann, S., Rutkowski, E. (2023). *The Game Industry of Polan Report*. Polish Agency for Enterprise Development. https://en.parp.gov.pl/publications/publication/the-game-industry-of-poland-report-2023

Palma-Ruiz, J. M., Torres-Toukoumidis, A., Gonzalez-Moreno, S. E. & Valles-Baca H. G. (2022). An overview of the gaming industry across nations: using analytics with power BI to forecast and identify key influencers. *Heliyon*, 8, CellPress. doi: 10.1016/j.heliyon.2022.e08959

Politowski, C., Petrillo, F., Ullmann, G. C. & Guéhéneuc, Y.-G. (2021) Game industry problems: An extensive analysis of the gray literature, *Information and Software Technology*, Volume 134, 2021, doi: 10.1016/j.infsof.2021.106538.

Syndicat National du Jeu Video. (2023). *Annual Survey of Video Games in France – 2022 Edition*. Syndicat National du Jeu Video - SNJV. https://frenchgamesmap.fr/en/barometre

Schöbel, S., Saqr, M. & Janson, A. (2021) Two decades of game concepts in digital learning

environments – A bibliometric study and research agenda. *Computers & Education*, Volume 173, Elsevier, doi: 10.1016/j.compedu.2021.104296

Zackariasson, P. & Wilson, Z. (eds) (2012) The Video Game Industry: Formation, Present State, and Future, 1st Ed., Routledge.

# Exploring Hypersexualization in Video Games: A Literature Review of Female Characters

**Antonela Devčić**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
acizmesi@foi.unizg.hr

**Mihaela Sabol**
University North
Business Economics
Juraj Križanić Street 31, 42000 Varaždin, Croatia
mihsabol@unin.hr

***Abstract***. *The main aim of this paper is to improve the understanding of the representation of female characters in video games, focusing on the issue of their hypersexual representation. To achieve this, a literature meta-analysis was conducted, which included searches in relevant academic research databases: WoS, Scopus, IEEE, and ACM. A total of 19 papers were identified in the literature meta-analysis. The results, aligned with the two research questions outlined in this paper, show that female characters in video games have been predominantly hypersexualized over the last decade, are targeted at a male audience, and are often portrayed as sexual objects, victims, or in passive roles. Such portrayals have a negative impact on perceptions and attitudes towards women in the gaming community and reinforce gender stereotypes. They are also linked to self-objectification among female gamers. However, recent trends show a positive development towards more realistic representations and strong female protagonists who actively contribute to the storyline, as well as an increasing number of main female characters in popular games*

**Keywords.** female characters, hypersexualization video games

## 1 Introduction

In today's culture, video games are more than just a form of entertainment and leisure activity that was once thought to be reserved mainly for the younger population. The video game industry is rapidly following technological changes and trends in the software industry and artificial intelligence, with the aim of improving the personalization of the gaming experience for players (Paraschos & Koulouriotis, 2023). According to *Technopedia*, the global video game industry reached a total value of US$184.0 billion in 2023 (Mascellino, 2024), and the video game market is forecast to reach US$282.30 billion in 2024 (Statista, 2024a).

Video games have a number of positive influences in different contexts. For example, they have a strong impact on education and change teaching and learning activities (Toh & Kirschner, 2020). Video games such as *Minecraft* can also serve as a backbone for the adoption of 21st-century 4C skills in virtual environments: critical thinking, creativity, collaboration, and communication (Hewett, Zeng, & Pletcher, 2020). In marketing, video games offer in-game advertising (Abbasi, Rehman, Hussain, Ting, & Islam, 2021) that influences consumer behaviour and purchasing decisions.

The number of active video players (users) in 2023 was 2.42 billion and is expected to rise to 3.02 billion by the end of 2029 (Statista, 2024b). As influential media, video games affect popular culture (Desoniq Studio, 2023) and reach a large number of users worldwide. Therefore, the creation of effective storylines, main characters and the promotion of positive values in video games have the potential to break down stereotypes in society and overcome intercultural barriers in real life (Shliakhovchuk & García, 2020).

Statistics reveal that men made up 53% of gamers in the US in 2023, compared to 46% of female video gamers (Howarth, 2024), highlighting the strong interest of women in gamingAccording to this year's *Entertainment Software Association* (ESA) report, which covers the American population, 190.6 million respondents play video games every week. Thanks to the widespread use of mobile devices and the availability and affordability of the internet, 78% of gamers are mobile device users, while 90% participate in some form of online gaming (Entertainment Software Association, 2024). The ESA's study is of great value to the gaming industry as it provides a long-term insight into the changes in the number of female and male gamers over the last two decades. In 2004, female gamers made up less than 40% of the total number of players. This trend continued until 2012, when the proportion of female players rose to 47%, while the proportion of men was 53%. After 2014, according to ESA data, the number of women in the

gaming community drops to around 40%, but by 2022, the difference between women and men is back down to less than 10%, similar to 2014 (Entertainment Software Association, 2024).

It is important to emphasize that the data presented pertains to the United States; however, similar trends in narrowing the gap between the number of male and female players have also been confirmed by other studies and reports (see for example, UK statistics (Statista, 2024c) or growing number of female players in Asia (Niko Partners, 2023)). Despite growing interest, around 50% of female gamers are not interested in multiplayer games (Deloitte, 2024) and almost seven in ten female gamers play video games on their smartphones (YouGov PLC., 2023).

Although there is growing interest among women in video games, research on gender differences in the context of video (computer) games provides a valid reason why the role of women as players, game developers, or characters in stories continues to be suppressed. Female gamers frequently face gender-based, racial, and other forms of harassment within gaming environments. Gaming communities often propagate stereotypes that undermine women's competencies in shooter and action video games, subjecting them to insults during live streaming on platforms like Twitch (Ruberg, Cullen, & Brewster, 2019).

A comprehensive global study shows that in 2021, 30% of those working in professions related to video game development were women (Statista, 2024d). The results of the same study indicate that this percentage was 21% in 2014. Some of the issues women face in the game development industry include feeling like they don't belong in the traditionally male culture, feeling undervalued, being insulted, and other challenges (Drummond, Salgado, & Viterbo, 2022). The underrepresentation of women in the game industry opens up space for male domination in game development, which determines what types of games are produced and how the characters in the storyline are designed.

Popular games and gaming culture often promote hyper-masculinized protagonists (Sarda, Zerhouni, Gentile, Bry, & Bègue, 2022), making female characters less frequently depicted as main characters and more often in secondary roles (Lynch, Tompkins, van Driel, & Fritz, 2016; Miller & Summers, 2007). Female characters in games are often dependent on male characters. Another significant concern is the hypersexualization (Anupama & Chithra, 2022; Lynch et al., 2016) and objectification (Sarda et al., 2022) of female characters in video games. As cited from Cote (2020, p. 5): "…women continue to struggle with hypersexualized female characters, games that lack playable female characters…These behaviours and themes work to relegate women to the margins of gaming rather than allowing them to enter the cultural core."

The main aim of this paper is to critically examine the portrayal of female characters in video games, with a focus on gender stereotypes and the hypersexualization of female protagonists. The negative effects of such representation are also examined and discussed in relation to the gaming community's attitudes towards female video game players and the self-perception of women in games. The paper is organized as follows: After an introductory section, female characters are analysed within the context of relevant theories. The development of the most prominent female protagonists in video games is presented, highlighting positive examples in the face of the negative effects of hypersexualization in video games. After explaining the motivation and research questions, the process of a literature meta-analysis is presented. The core of the paper focuses on the results of the two research questions. The conclusion summarizes the main findings and outlines future guidelines for research, practice, and the gaming community in relation to female characters in video games.

# 2 Portrayal of female characters in video games

In this section of the paper, relevant theories related to understanding the problem are discussed. Additionally, significant female characters are described and analyzed in the context of the evolution of their representation in video games.

## 2.1 Theories explaining portrayal of female characters in video games

In the existing literature, the problem of stereotyped and hypersexualised female characters in video games is explained with different theories from media studies, gender studies and psychology.

Cultivation theory, proposed by George Gerbner and Larry Gross, suggests that long-term exposure to media content, in this case, video games, can influence users' perceptions of reality (Martins, Williams, Harrison, & Ratan, 2009). For example, if a gamer is exposed to sexualized and objectified portrayals of female characters in supporting roles from a young age, this exposure can potentially shape the gamer's views on gender roles and reinforce sexist attitudes.

According to the assumptions of Laura Mulvey's male gaze theory, main characters in video games are often created as visual spectacles primarily from a heterosexual male perspective. Female characters in these games are typically portrayed as objects of male desire and are depicted in passive roles (Lavrador, 2023). Furthermore, female characters are often portrayed with hypersexualized characteristics and possess weaker abilities or powers compared to male characters.

Based on the assumptions of cultivation theory and the social-cognitive theory of mass communication, Dill and Thill (2007) emphasized that gender representations in media significantly influence people's perceptions in real life. For example, the hypersexualized portrayal of female characters in video games reinforces stereotypes about male and female roles and relationships, portraying women as weak, submissive, and often as objects (Dill & Thill, 2007).

According to Fox et al. (2015), objectification theory, proposed by Fredrickson and Roberts in 1997, explains the problem of hypersexualized portrayals of women in video games. This theory suggests that women's exposure to sexualized depictions in media leads to self-objectification.

## 2.2 Evolution of female characters in video games

Even in the early days of video games, characters without a defined gender, such as *Pac-Man*, were explicitly portrayed as male. Later, in 1982, the character of Ms. Pac-Man was created for a popular arcade game to potentially attract more female players and increase the game's appeal (Kirkpatrick, 2017, p. 461).

Although the entire *Legend of Zelda* series is named after her, Zelda is never the main character or heroine of the various games. In different game stories, the female characters are passive, while the male characters are heroes who save them based on their skills and abilities. The story is told from a predominantly male perspective, and the female characters, as princesses, reflect the so-called "damsel in distress" stereotype (Pugh, 2018).

One of the first examples of a female character in video games is Lara Croft, the capable protagonist in the 1996 game *Tomb Raider*. With her distinctive characteristics and personality, Lara Croft breaks stereotypes about female roles and portrayals in video games. The appearance of a strong female character in an action video game is known in culture and literature as the "Lara phenomenon" (Engelbrecht, 2020; Jansz & Martis, 2007). This term refers to the portrayal of the female protagonist as strong and competent. In the early stages of the game series, Lara Croft was portrayed as an action figure with distinctly feminine features. As the series evolved and new editions were released, Lara took on a more realistic body size, was dressed more appropriately, and was portrayed as a non-sexualized woman (Cross et al., 2024).

Aloy, a female character from the video game *Horizon Zero Dawn*, embodies a young woman growing up alone in a post-apocalyptic world (Woolbright, 2018). The game follows her journey, and the story is based on feminist perspectives. The female characters in this game are not overly sexualized, helpless, or passive (Brusuelas, 2021). Aloy is portrayed realistically in terms of her appearance and is a distinctive character characterized by combat training, strong leadership skills, emotional intelligence, and negotiation abilities.

Jill Valentine represents a female protagonist in *Resident Evil*. She is brave, smart, loyal, and skilled with weapons due to her military experience (Lewis, 2024). Regarding her appearance and clothing, she typically wears cargo trousers or combat-ready outfits suitable for her missions. However, in the third installment of *Resident Evil*, she wore a short skirt and knee-high boots, hinting at a certain level of sexualization.

The creators of the video game *Assassin's Creed Odyssey* designed Kassandra as a protagonist whom players can choose to play as. She is portrayed as strong and muscular, yet at the same time gentle and emotionally intelligent (Lance, 2018). However, players are not limited to choosing her, as her brother Alexios is also available as a playable character. It is important to note that both characters are not portrayed in a sexualized manner and are appropriately dressed according to the game's context, even though some parts of their bodies are exposed (The Odyssey of Woman Representation, 2021). Although the creators of the game, Ubisoft Quebec, developed Kassandra and Alexios as equally competent characters and designed them independently of gender, statistics show that the male protagonist was the more common choice among players; only a third of the players chose a female character as the protagonist (Duckworth, 2018). This was despite the fact that the Kassandra character provided a better gameplay experience, as reported by players.

More recently, the video game *The Last of Us* features a protagonist named Ellie, a young girl who is brave, smart, and capable, guiding an older male character, Joel, through challenges in a post-apocalyptic world (McCullough, Wong, & Stevenson, 2020).

# 3 Research problem

A review of the literature found that the hypersexualized portrayal of female characters in video games discourages women from participating in gaming (Lynch et al., 2016). For this reason, women tend to choose games that do not feature hypersexualized characters or avatars, but rather have strong female protagonists. In the context of video games, avatars of characters meant to be human-like are often portrayed as objects or somehow dehumanized due to the digital nature of their image representation (Fox et al., 2015, p. 351). Cross et al. (2024) have highlighted two main issues regarding women in video games: the relationship between female protagonists and the way they are portrayed.

Overly sexualized portrayals of female characters, restricted or repressed behaviors, and characterizations of female characters in video games negatively impact

the treatment of issues related to gender discrimination. The identified imbalance between female and male protagonists exacerbates this problem (Cross et al., 2024). The enjoyment and immersion of a game are related to the player's ability to identify with the main character (Heron, Belford, & Goker, 2014). Therefore, the lack of female protagonists in video games reduces the applicability of this experience for women. Self-objectification could explain the association between video games, body image disorders, and gender stereotyping, especially in adolescents. However, sexualized female avatars affect all players, regardless of gender and gender identity (Vandenbosch et al., 2017). Based on the introductory section and the issues highlighted in the literature regarding the portrayal of female characters in video games, this research addresses two key research questions (RQs):

***RQ1:*** What are the typical representations and roles assigned to female characters in video games?

***RQ2:*** How do sexualised and stereotypical female characters influence the general attitudes and perceptions of women towards games, as well as the attitudes and perceptions of other players towards females?

# 4 Methodology

To answer the two research questions for meta-analysis, a literature search was conducted according to the guidelines of Kitchenham and Charters (2007). In order to find relevant papers, four recognized scientific research databases were included in the search: Scopus, ACM, Web of Science (WoS), and IEEE.

The keywords used for the database search were chosen to target relevant areas related to the research problem: (1) the representation and issues related to women, (2) a focus on sexual themes, including sexualization and hypersexualization, and (3) the specific context of video games and online gaming. For each academic research database, the query was customized with a Boolean operator and a combination of keywords to refine the scope of the search. An example of a query with keywords and operators used for searching in IEEE was: *("Woman" OR "Female") AND ("sex*" OR "sexual*" OR "hypersexual*" OR "sexualization" OR "hypersexualisation") AND ("video gam*" OR "gam*" OR "internet gam*").* The search was limited to journal articles and conference papers published in English between 2014 and 2024 and was conducted in May and June 2024. The initial searches returned the following results:

- Scopus: 404 papers
- ACM: 329 papers
- WoS: 283 papers
- IEEE: 47 papers

The next step in the literature search was to identify relevant papers by reading the title, abstract, and keywords. If a paper appeared to be relevant to the topic of the literature search, its data (title, authors, year of publication) were added to a common list of identified research papers for all scientific databases. After a detailed screening process, this step resulted in a list of 71 relevant research papers. After duplicates were removed, 63 research papers were downloaded and read in full. This step was crucial to determine whether the topic of each paper was relevant to the research problem and research questions. Papers dealing exclusively with females in gaming phenomena, characters in virtual realities, or issues related to gender differences in gaming style were excluded from the final analysis. Finally, the sample for the exploratory literature analysis comprised 19 relevant papers on the topic of the hypersexualization of female characters in video games. A list of the papers included in the meta-analysis is shown in Table 1.

**Table 1.** Papers included in the meta-analysis on female representation in video game

| | **Authors and year** |
|---|---|
| 1 | Ansari, R., Jaffar, B. A., Riaz, S., Kaur, M. J., & Mushtaq, A. (2019) |
| 2 | Anupama, K., & Chithra, G. K. (2022) |
| 3 | Blomquist, G. (2021) |
| 4 | Cross, L., Kaye, L. K., Savostijanovs, J., McLatchie, N., Johnston, M., Whiteman, L., … Atherton, G. (2024) |
| 5 | Davies, B., Headleand, C. J., & Hicks, K. (2020) |
| 6 | Gao, G., Min, A., & Shih, P. C. (2017) |
| 7 | Heron, M., Belford, P., & Goker, A. (2014) |
| 8 | Hollett, R. C., Morgan, H., Chen, N. T. M., & Gignac, G. E. (2020) |
| 9 | Lynch, T., Tompkins, J. E., Gilbert, M., & Burridge, S. (2024) |
| 10 | Lynch, T., Tompkins, J. E., van Driel, I. I., & Fritz, N. (2016) |
| 11 | McCullough, K. M., Wong, Y. J., & Stevenson, N. J. (2020) |
| 12 | Perreault, M. F., Perreault, G. P., Jenkins, J., & Morrison, A. (2018) |
| 13 | Perry, K. (2022) |
| 14 | Ruberg, B., Cullen, A. L. L., & Brewster, K. (2019) |
| 15 | Sarda, E., Zerhouni, O., Gentile, D. A., Bry, C., & Bègue, L. (2022) |
| 16 | Schwind, V., & Henze, N. (2018) |
| 17 | Silva, J. P. N., Valadares, G. C., Pedrosa, G., Rezende, D. C., Cappelle, M. C. A., & Assis, F. A. A. (2023) |
| 18 | Tompkins, J. E., Lynch, T., Van Driel, I. I., & Fritz, N. (2020) |
| 19 | Vandenbosch, L., Driesmans, K., Trekels, J., & Eggermont, S. (2017) |

# 4 Results of the meta literature review

In this section, exploratory literature review findings will be presented and summarized in the context of two research questions.

## 4.1 Findings related to the RQ1

In relation to the first research question, a) representations and b) roles of female characters in selected academic papers were analyzed. Particular attention was paid to the hypersexualized representation of female protagonists in video games over the last ten years.

The most common stereotypes associated with female characters in video games are sexism, scantily clad women, and the notion of beauty (Ansari et al., 2019). One of the most studied elements of video game content is the sexual objectification of female characters, particularly in relation to their physical appearance. Women and female characters in video games are often portrayed through sexualized elements, such as unrealistic body shapes with emphasis on large breasts, small waists, and skimpy clothing that highlights their sexual characteristics (Lynch, Tompkins, Gilbert, & Burridge, 2024; Sarda et al., 2022). Female characters in video games are often portrayed as hypersexualized and objectified, and are primarily aimed at a male audience (Anupama & Chithra, 2022) or in a seductive way (Cross et al., 2024). Characters such as Sarah Kerrigan from StarCraft, a monstrously feminine character, are depicted in skimpy clothing that emphasizes sexual attributes and with unreal body proportions, regardless of their abilities and importance to the story's plot. Such depictions tend to draw the player's attention to the physical appearance rather than the character of the female protagonists (Blomquist, 2021). In video games, female characters' bodies are depicted as objects or as weapons, and the theme of female characters' (in)dependence is recognized (Blomquist, 2021).

Advertising materials for video games show women in highly sexualised poses. Women are portrayed as victims in video games, with little clothing or objectified to fulfil men's desires (Heron et al., 2014). For example, Joker's ally, Harley Quinn, is created as a marketing object to attract the younger male target audience (Anupama & Chithra, 2022).

In AAA video games (high-end and high-budget games such as *Grand Theft Auto, Call of Duty, and Red Dead Redemption*), female characters are rarely the main protagonists and are often portrayed as sexual objects or victims (Perry, 2022). Ruberg (2019)) emphasizes that these characters are not only objectified but are frequently depicted as sex workers or in passive roles, such as prostitutes or private dancers, and are mistreated by male characters (Sarda et al., 2022)

The marginalisation of women in video games is closely linked to gender stereotypes that often portray them as helpless victims (e.g. damsels in distress who need to be rescued by a man) and secondary characters or often placing female characters in a lower status than male protagonists (McCullough et al., 2020). Sexualised female characters in video games promote stereotypes and women are portrayed as objects and are less important than men (Perreault, et al., 2018).

A study of 398 video games developed in the USA and Japan between 1983 and 2017 shows that only a third of female characters are main characters. Therefore, female characters are mostly minor characters and conform to traditional gender values (Lynch et al., 2024). The roles that female characters take in video games are usually passive and portrayed as a visual attraction for the male players (Anupama & Chithra, 2022) and have emphasised sexualised characters (Ansari et al., 2019). In *World of Worcraft*, female characters usually do not have a leading role, they are in the minority and are introduced as girls in trouble or as a reward for achieving a goal in the game (Silva et al., 2023). In the game *Legaue of Legends*, typical female characters usually do not have attack skills, but focus on defence or healing male characters such as "tanks" (Gao, Min, & Shih, 2017) or supporting allies, which improves team performance (Davies, Headleand, & Hicks, 2020). On the other hand, female characters sometimes sacrifice their abilities as a character trait in order to develop or advance in the game (e.g. in the video game *Heroes of the Strom*). Female characters have more abilities to help other heroes in the game, which naturally puts them in a subordinate/inferior role (Davies et al., 2020).

It is worth mentioning that the (hyper)sexualized representation of female characters varies across different cultures. In Japanese culture, for example, in anime games, female characters are strong and have supernatural abilities, but are still represented sexually. On the other hand, some female characters are portrayed as innocent, pale, and sympathetic—traits that are desirable in this culture—which aligns with the empowerment of such characters (Tompkins et al., 2020). Nevertheless, games developed in the USA and Japan are quite similar in terms of the sexualized portrayal of female characters. In Japanese games, women tend to be portrayed in traditional female roles (e.g., princess, cheerleader) and do not dominate the game storyline (Lynch et al., 2024).

However, this literature review also found some positive examples and figures of female characters in video games. Female protagonists in video games such as *The Last of Us*, *Bioshock Infinite*, *Tomb Raider*, and *Beyond: Two Souls* feature female protagonists with developed skills and roles that are not merely passive but actively contribute to the game's plot (Perreault et al., 2018, p. 843). The results show that 80% of female

characters in games, regardless of the excessive sexualization of their physical representation, are capable within the plot (Lynch et al., 2024). Recently, however, there has been a positive trend toward a more realistic portrayal of female characters and those that can be chosen for the game. The primary female characters are less sexualized than the secondary female characters, and their meaning and character go beyond the sexualized portrayal. The secondary female characters are depicted in a sexualized way to increase their attractiveness in the game (Lynch et al., 2024). In *League of Legends*, the authors noted a positive trend toward equal competition between male and female characters in the game and an increase in the number of female champions (Davies et al., 2020).

## 4.2 Findings related to the RQ2

This section presents the main findings on the influence of sexualized, stereotypical female characters in video games on: (a) the perception of women within the gaming community and (b) the attitudes and self-perception of women when playing video games.

Sexualized depictions of female characters in video games negatively impact the perception and attitudes of gaming communities towards women (Blomquist, 2021), and contribute to the negative self-perception and self-objectification of women (Ansari et al., 2019). Players of hardcore games tend to be men, while women are more likely to play casual or single-player games. Male players often develop sexist attitudes towards women, perceiving them as sex objects (Ansari et al., 2019). This is supported by the results of a study involving the gaming experiences of 319 American women, which indicate that frequent video game play is associated with the formation of negative beliefs about women—such as the notions that they are less capable than men and lack leadership qualities—leading to numerous prejudices against female gamers (McCullough et al., 2020).

In forum discussions, chats or online communities, female casual gamers rarely participate in discussions or express their opinions (Heron et al., 2014). In the virtual world of gaming, female gamers are often exposed to sexist jokes, insults and rape threats, which can force them into such actions that can later be linked to a lack of interest or abandonment of STEM careers due to exposure to negative gender values (Davies et al., 2020). While some women suffer the consequences of internalised misogyny (a form of sexist behavior and attitudes that women hold towards themselves or other women and girls), others resist harmful ideas about women portrayed in games (McCullough et al., 2020).

Researchers highlight the concept of 'gender camouflage', which posits that female players choose male characters to avoid possible negative comments, attitudes, and behaviors from male players, even when female characters are an option (Silva et al., 2023). Interestingly, older female players are more likely to create male avatars and often reduce the femininity of their avatars, but invest more effort in customising their in-game appearance (Schwind & Henze, 2018). The negative self-perception of female gamers is associated with the use of sexualized avatars in video games and other virtual environments. Furthermore, these depictions contribute to a negative perception of women in general, which affects attitudes both in the virtual environment and in the real world (Fox et al., 2015).

Younger populations, including both girls and boys, develop increased self-objectification when playing with sexualized avatars compared to non-sexualized ones (Vandenbosch et al., 2017). Virtual portrayals of female characters in adult-only games trigger a sexually objectifying gaze in both male and female participants when exposed to them (Hollett, Morgan, Chen, & Gignac, 2020).

In the popular game *League of Legends*, gender stereotypes influence both player character selection and character design in line with players' real-life genders (Gao et al., 2017). In games like *Grand Theft Auto* (GTA), the roles of female characters as sex workers contribute to negative perceptions of women and devalue their work (Ruberg, 2019). The impact of identifying with video game characters varies; for example, it predicts a stronger association between self-concept and masculinity but does not correlate with the objectification of women (Schwind & Henze, 2018). Conversely, female players often seek a connection to video game characters that extends beyond physical appearance, encompassing both the character's personality and the game's storyline (Tompkins et al., 2020).

Distorted representations of the female body perpetuate stereotypes, and female gamers are less respected, even by members of the same gender (Blomquist, 2021). Sexualized representations of female characters in games or their portrayal in subordinate roles lead to so-called benevolent sexism (Davies et al., 2020) and are linked to the negative self-perception of women and self-objectification (Ansari et al., 2019). Additionally, the passive roles of female characters affect the common understanding that their role is secondary in games (Anupama & Chithra, 2022), and also in real-life situations.

In recent work, Lynch et al. (2024) point to positive trends in the balanced number of male and female gamers and a growing interest among women in playing games and participating in content creation. These trends have been further reinforced by feminist debates that have highlighted the problem of the stereotypical representation of female characters in video games, hypersexualized representation, and the underrepresentation of women in the gaming industry. Nevertheless, research findings suggest that short-term exposure of gamers to video games in which female characters are portrayed in a passive, hypersexualized

manner did not have a significant impact on the adoption of negative attitudes towards gender stereotypes or attitudes towards women (Cross et al., 2024). Similarly, recent analyses suggest that female characters in video games are portrayed in a less sexualized way (Lynch et al., 2024; Lynch, 2016), indicating a positive change in game development.

# 5 Conclusion

Although the review of the literature has shown that the problem of oversexualization and objectification still exists in relation to female characters in video games, recent work has shown positive developments where women are portrayed more realistically and play important roles in the story (e.g., Ellie in the game *The Last of Us*). Perreault et al., (2018) note that the development and inclusion of a greater number of female protagonists in video games, along with the creation of more complex female characters, can have a positive effect by reducing negative attitudes and stereotypes towards women.

Changes are taking place, and there is a positive trend towards more gender inclusivity and feminism, also driven by educating video game developers about gender stereotypes and the negative effects of hypersexualization (Tompkins et al., 2020). In the future, artificial intelligence and data mining could be used to build solutions to recognize elements of the objectification of women in video games (Ansari et al., 2019). One approach to gaining insights from the gaming community regarding female gaming and related issues is to use netnography to identify the current state (Silva et al., 2023).

In the future, more female developers should create video games to ensure a female perspective that could increase the interest of women in gaming and improve character portrayal (Anupama & Chithra, 2022). According to *Forbes* (2023), women take the role of representatives for gaming companies (so-called ambassadors) to attract more female players. Also, cosmetics and makeup brands are forming collaborations with popular female gamers in the industry.

The games industry also recognizes the need to integrate the principles of diversity, equality, and inclusion (DEI) into job roles, leading character creation, storylines, etc., to continue the positive progress for women and their real and virtual images in the industry and in society as a whole. These challenges are acknowledged in the media, professional and scientific literature, and by various organizations. Therefore, ongoing efforts are needed to sustain the progress made in breaking stereotypes.

The limitation of this paper is that a limited number of articles were found when searching the scientific databases because of using specific keywords. The topic of the hypersexualization of female characters, as well as female characters in general in video games, is a relevant topic in academic, professional, and media circles. Expanding the sources to include perspectives from female gamers, interviews with game developers, and insights from the gaming community would improve the understanding of perceptions of female characters in gaming world. Nonetheless, this paper makes an academic contribution by providing an overview of academic research from the last ten years and highlighting the positive progress made towards gender equality in video game characters.

# Acknowledgments

This paper was written based on the topic of the bachelor thesis: *The portrayal of female characters in video games* by student Mihaela Sabol, whose supervisor in 2023 was Antonela Devčić.

# References

Abbasi, A. Z., Rehman, U., Hussain, A., Ting, D. H., & Islam, J. U. (2021). The impact of advertising value of in-game pop-up ads in online gaming on gamers' inspiration: An empirical investigation. *Telematics and Informatics, 62*, 101630.

Ansari, R., Jaffar, B. A., Riaz, S., Kaur, M. J., & Mushtaq, A. (2019). Datamining to alert the formation of women objectification stereotypes in video games. *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, 521–526. https://doi.org/10.1109/AICAI.2019.8701256

Anupama, K., & Chithra, G. K. (2022). Of castration anxiety and hypersexualized female bodies: A critical assessment of the objectifying gaze in *Batman: Arkham Video Game Series*. *Theory and Practice in Language Studies, 12*(9), 1921–1929. https://doi.org/10.17507/tpls.1209.26

Blomquist, G. (2021). Postmodern Medusa: The monstrous-feminine in *StarCraft II: Heart of the Swarm*. *Games and Culture, 16*(7), 885–906. https://doi.org/10.1177/15554120211005232

Brusuelas, C. (2021, February 18). *Horizon Zero Dawn*: PS4 feminist reviews. *Medium*. https://cbrucewillis.medium.com/horizon-zero-dawn-ps4-feminist-reviews-bf786898d380

Cote, A. C. (2020). *Gaming sexism: Gender and identity in the era of casual video games*. New York University Press.

Cross, L., Kaye, L. K., Savostijanovs, J., McLatchie, N., Johnston, M., Whiteman, L., … Atherton, G. (2024). Gendered violence and sexualized representations in video games: (Lack of) effect on gender-related attitudes. *New Media and Society, 26*(3), 1648–1669.

https://doi.org/10.1177/14614448221075736

Davies, B., Headleand, C. J., & Hicks, K. (2020). Exploring the impact of gender on character mechanics in multiplayer online battle arena (MOBA) games. *Proceedings - 2020 International Conference on Cyberworlds, CW 2020*, 71–78. https://doi.org/10.1109/CW49994.2020.00018

Deloitte. (2024). For women playing video games, it's (still) a man's world. Retrieved 29 June 2024, from https://www2.deloitte.com/xe/en/insights/industry/technology/digital-media-trends-consumption-habits-survey/2024/how-can-gaming-industry-get-more-female-gamers.html

Desoniq Studio. (2023). Game on: How gaming revolutionized pop culture - From movies to music and runways. Retrieved 24 June 2024, from https://www.linkedin.com/pulse/game-how-gaming-revolutionized-pop-culture-from-movies/

Dill, K. E., & Thill, K. P. (2007). Video game characters and the socialization of gender roles: Young people's perceptions mirror sexist media depictions. *Sex Roles, 57*(11–12), 851–864. https://doi.org/10.1007/s11199-007-9278-1

Drummond, B. M., Salgado, L. C. de C., & Viterbo, J. (2022). What are the challenges faced by women in the games industry? *Proceedings - 2022 Workshop on Interactive Technologies and Games*, 167–178. https://doi.org/10.5753/wit.2022.223130

Duckworth, J. (2018, December 16). *Assassin's Creed Odyssey* director reveals how many players picked Alexios. *Game Rant*. Retrieved September 3, 2024, from https://gamerant.com/assassins-creed-odyssey-alexios/

Engelbrecht, J. (2020). The new Lara phenomenon: A postfeminist analysis of *Rise of the Tomb Raider*. *Game Studies, 20*(3), 9.

Entertainment Software Association. (2024). Essential facts about the U.S. video game industry: 2024 data. Retrieved September 2, 2024, from https://www.theesa.com/resources/essential-facts-about-the-us-video-game-industry/2024-data/

Forbes. (2023). Leveling up the gaming gender gap. Retrieved 27 June 2024, from https://www.forbes.com/sites/forbesbusinesscouncil/2023/08/24/leveling-up-the-gaming-gender-gap/

Fox, J., Ralston, R. A., Cooper, C. K., & Jones, K. A. (2015). Sexualized avatars lead to women's self-objectification and acceptance of rape myths. *Psychology of Women Quarterly, 39*(3), 349–362. https://doi.org/10.1177/0361684314553578

Gao, G., Min, A., & Shih, P. C. (2017). Gendered design bias: Gender differences of in-game character choice and playing style in *League of Legends*. *ACM International Conference Proceeding Series*, 307–317. https://doi.org/10.1145/3152771.3152804

Heron, M., Belford, P., & Goker, A. (2014). Sexism in the circuitry: Female participation in male-dominated popular computer culture. *ACM SIGCAS Computers and Society, 44*(4), 18–29.

Hewett, K. J. E., Zeng, G., & Pletcher, B. C. (2020). The acquisition of 21st-century skills through video games: *Minecraft* design process models and their web of class roles. *Simulation and Gaming, 51*(3), 336–364. https://doi.org/10.1177/1046878120904976

Hollett, R. C., Morgan, H., Chen, N. T. M., & Gignac, G. E. (2020). Female characters from adult-only video games elicit a sexually objectifying gaze in both men and women. *Sex Roles, 83*(1–2), 29–42. https://doi.org/10.1007/s11199-019-01096-y

Howarth, J. (2024). How many gamers are there? (New 2024 statistics). Retrieved 29 June 2024, from https://explodingtopics.com/blog/number-of-gamers

Jansz, J., & Martis, R. G. (2007). The Lara phenomenon: Powerful female characters in video games. *Sex Roles, 56*(3–4), 141–148. https://doi.org/10.1007/s11199-006-9158-0

Kirkpatrick, G. (2017). How gaming became sexist: A study of UK gaming magazines 1981–1995. *Media, Culture and Society, 39*(4), 453–468. https://doi.org/10.1177/0163443716646177

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in SE. *Guidelines for Performing Systematic Literature Reviews in SE*, 1–44.

Lance, S. (2018, December 3). Ubisoft should have trusted *Assassin's Creed: Odyssey* with a female protagonist. *Merry-Go-Round Magazine*. Retrieved September 3, 2024, from https://merrygoroundmagazine.com/assassins-creed-odyssey-trusting-the-female-protagonist/

Lavrador, E. (2023). Video games against the male gaze. *Journal of Science and Technology of the Arts, 15*(2), 135–153. https://doi.org/10.34632/jsta.2023.15706

Lewis, E. (2024). *Resident Evil*: Every iteration of Jill Valentine, ranked. *Gamerant*. Retrieved 25 June 2024, from https://gamerant.com/resident-evil-best-jill-valentine-versions/

Lynch, T., Tompkins, J. E., Gilbert, M., & Burridge, S. (2024). Evidence of ambivalent sexism in female video game character designs. *Mass Communication and Society, 00*(00), 1–26. https://doi.org/10.1080/15205436.2024.2311229

Lynch, T., Tompkins, J. E., van Driel, I. I., & Fritz, N. (2016). Sexy, strong, and secondary: A content analysis of female characters in video games across

31 years. *Journal of Communication, 66*(4), 564–584. https://doi.org/10.1111/jcom.12237

Martins, N., Williams, D. C., Harrison, K., & Ratan, R. A. (2009). A content analysis of female body imagery in video games. *Sex Roles, 61*, 824–836.

Mascellino, A. (2024). 70+ video game statistics you need to know in 2024: Market growth, emerging trends, and more. Retrieved 24 June 2024, from Technopedia website: https://www.techopedia.com/video-game-statistics

McCullough, K. M., Wong, Y. J., & Stevenson, N. J. (2020). Female video game players and the protective effect of feminist identity against internalized misogyny. *Sex Roles, 82*(5–6), 266–276. https://doi.org/10.1007/s11199-019-01055-7

Miller, M. K., & Summers, A. (2007). Gender differences in video game characters' roles, appearances, and attire as portrayed in video game magazines. *Sex Roles, 57*(9–10), 733–742. https://doi.org/10.1007/s11199-007-9307-0

Niko Partners. (2023). Female gamers in Asia. Retrieved September 2, 2024, from https://nikopartners.com/female-gamers-in-asia/

Paraschos, P. D., & Koulouriotis, D. E. (2023). Game difficulty adaptation and experience personalization: A literature review. *International Journal of Human-Computer Interaction, 39*(1), 1–22. https://doi.org/10.1080/10447318.2021.2020008

Perreault, M. F., Perreault, G. P., Jenkins, J., & Morrison, A. (2018). Depictions of female protagonists in digital games: A narrative analysis of 2013 DICE award-winning digital games. *Games and Culture, 13*(8), 843–860. https://doi.org/10.1177/1555412016679584

Perry, K. (2022). Damsels and darlings: Decoding gender equality in video game communities. *Feminist Media Studies, 22*(5), 1102–1119. https://doi.org/10.1080/14680777.2021.1883085

Pugh, T. (2018). The queer narrativity of the hero's journey in Nintendo's *The Legend of Zelda* video games. *Journal of Narrative Theory, 48*(2), 225–251.

Ruberg, B. (2019). Representing sex workers in video games: Feminisms, fantasies of exceptionalism, and the value of erotic labor. *Feminist Media Studies, 19*(3), 313–330. https://doi.org/10.1080/14680777.2018.1477815

Ruberg, B., Cullen, A. L. L., & Brewster, K. (2019). Nothing but a "titty streamer": Legitimacy, labor, and the debate over women's breasts in video game live streaming. *Critical Studies in Media Communication, 36*(5), 466–481. https://doi.org/10.1080/15295036.2019.1658886

Sarda, E., Zerhouni, O., Gentile, D. A., Bry, C., & Bègue, L. (2022). Some effects of sexist video games on self-masculinity associations. *Information Communication and Society, 25*(12), 1683–1698. https://doi.org/10.1080/1369118X.2021.1877770

Schwind, V., & Henze, N. (2018). Gender- and age-related differences in designing the characteristics of stereotypical virtual faces. *CHI PLAY 2018 - Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 462–475. https://doi.org/10.1145/3242671.3242692

Shliakhovchuk, E., & García, A. M. (2020). Intercultural perspective on impact of video games on players: Insights from a systematic review of recent literature. *Educational Sciences: Theory and Practice, 20*(1), 40–58. https://doi.org/10.12738/jestp.2020.1.004

Silva, J. P. N., Valadares, G. C., Pedrosa, G., Rezende, D. C., Cappelle, M. C. A., & Assis, F. A. A. (2023). Gender imbalance in MMORPG: The case of World of Warcraft in Brazil. *Feminist Media Studies, 23*(1), 289–305. https://doi.org/10.1080/14680777.2021.1973060

Statista. (2024a). Distribution of video game developers worldwide from 2014 to 2021, by gender. Retrieved 23 June 2024, from https://www.statista.com/statistics/453634/game-developer-gender-distribution-worldwide/

Statista. (2024b). Number of video game users worldwide from 2019 to 2029. Retrieved 24 June 2024, from https://www.statista.com/statistics/748044/number-video-gamers-world/

Statista. (2024c). Digital games users by age and gender in the UK. *Statista Digital Market Outlook*. Retrieved September 2, 2024, from https://www.statista.com/forecasts/461273/digital-games-users-age-gender-digital-market-outlook-uk

Statista. (2024d). *Video games - Worldwide*. Retrieved 24 June 2024, from https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide?currency=USD

The odyssey of woman representation. (2021, April 16). *Just a Game*. Retrieved September 3, 2024, from https://justagame.uk/2021/04/16/the-odyssey-of-woman-representation/

Toh, W., & Kirschner, D. (2020). Self-directed learning in video games, affordances and pedagogical implications for teaching and learning. *Computers & Education, 154*, 103912.

Tompkins, J. E., Lynch, T., Van Driel, I. I., & Fritz, N. (2020). Kawaii killers and femme fatales: A textual analysis of female characters signifying benevolent

and hostile sexism in video games. *Journal of Broadcasting and Electronic Media, 64*(2), 236–254. https://doi.org/10.1080/08838151.2020.1718960

Vandenbosch, L., Driesmans, K., Trekels, J., & Eggermont, S. (2017). Sexualized video game avatars and self-objectification in adolescents: The role of gender congruency and activation frequency. *Media Psychology, 20*(2), 221–239. https://doi.org/10.1080/15213269.2016.1142380

Woolbright, L. (2018). Ecofeminism and gaia theory in *Horizon Zero Dawn*. *Trace Journal*, 2.

YouGov PLC. (2023). *Gaming & Esports Report 2023*.

# Cyber-Physical and Embedded Systems

Ivan Sládek, Gabriel Gašpar, Juraj Repka, Roman Budjač
**Stereo Vision Based 3D Positioning for Real-Time Control Applications**

Juraj Ďuďák, Jakub Perička, Martin Skovajsa
**A Lightweight communication protocol for Body Area Networks**

Roman Budjač, Gabriel Gašpar, Maros Valasek, Martin Barton
**Proposal for a weather monitoring system utilizing IoT technologies**

Roman Ruzarovsky, Richard Skýpala, Nikolas Blahušiak, Ján Šido, Martin Csekei, Tibor Horák, Peter Střelec
**Virtual Commissioning of Industrial Robot through Software-in-the-Loop Strategies for Cyber-Physical Systems**

Marija Kuštelega, Ahmed Shareef, Renata Mekovec
**The Role of Digital Twins in Future Skill Development**

Snježana Križanić, Neven Vrček, Matija Habuš
**From Classroom to Real-World: University Students' Experience in Learning Solidity and Blockchain**

Matija Habuš, Neven Vrček, Snježana Križanić
**Optimization of urban road network and digital twins: a literature review**

Goran Filinić, Ivan Magdalenić
**Agile Software Development Life Cycle Model for Internet of Things**

# Stereo vision based 3D positioning for real-time control applications

**Ivan Sládek**
University of Žilina
Research centre
Univerzitná 8215/1, 010 26 Žilina, Slovak republic
ivan.sladek@uniza.sk

**Gabriel Gašpar**
University of Žilina / Faculty of Materials Science and Technology in Trnava
Research centre / Slovak University of Technology in Bratislava
Univerzitná 8215/1, 010 26 Žilina, Slovak republic / Jána Bottu 25, 917 24 Trnava, Slovak republic
gabriel.gaspar@uniza.sk / gabriel.gaspar@stuba.sk

**Juraj Repka**
Faculty of Materials Science and Technology in Trnava
Slovak University of Technology in Bratislava
Jána Bottu 25, 917 24 Trnava, Slovak republic
xrepkaj@stuba.sk

**Roman Budjač**
University of Žilina
Research centre
Univerzitná 8215/1, 010 26 Žilina, Slovak republic
roman.budjac@uniza.sk

**Abstract.** *Humans observe the world predominantly through vision, naturally perceiving depth and distance. This ability is thanks to binocular vision, where our two eyes collaborate to form a single, three-dimensional image. This principle is also applied in computers through a technology known as stereo vision. Stereo vision is a pivotal technology with applications across various fields, including medicine, entertainment, robotics, and industry. This paper uses this innovative technology to enhance gaming experiences and human-machine interactions. By integrating stereo vision with deep learning techniques for object detection, we estimate the 3D positions of detected objects and utilize this information to control specific aspects of computer games.*
**Keywords.** Stereo Vision, Object Detection, 3D Position Estimation, Game Controller

## 1 Introduction

Humans observe the world primarily through their eyes, which possess the natural ability to perceive the depth of objects. This depth perception allows us to determine whether something is near or far. This ability, known as binocular vision, relies on our two eyes to create a three-dimensional image. Inspired by this biological phenomenon, computer vision employs a similar approach called stereovision. Stereovision enables machines to interpret 3D space with applications across various fields [1, 2]. In medicine, it aids in performing precise surgical procedures [3]. In entertainment, it allows for the creation of immersive 3D movies; in robotics, it facilitates navigation and manipulation; and in industry, it enhances automated quality control. As technology advances, stereovision continues integrating into more segments of our lives.

Gaming has also become a significant part of daily life for many individuals, providing a means to relax and socialize. While gaming can benefit mental health, it often negatively impacts physical health due to the sedentary nature of most modern games. This paper addresses this issue by exploring using the human body as a game controller. By integrating stereovision with deep learning techniques, we propose a system for real-time human tracking. This system detects objects in images, estimates their 3D positions, and uses this information to control various aspects of computer games, leading to more immersive and interactive ex-

periences. Additionally, we propose a cost-effective stereovision solution, addressing the high expense of traditional systems.

Our motivation for this project began last year when we aimed to create an engaging project to attract the broad public into scientific projects. It led to the developing of a game controller project based on augmented reality. The controller captured images with an ESP32-CAM camera, which were then processed by a server running a program that detected characters in the image and recorded their positions on the X and Y axes. Such coordinates were used in an application environment, such as controlling the keyboard or character movement in a game. We tested this system during our faculty's open days, and it was well-received by students. However, the system had several issues, including problems with motion recognition in images with multiple people, low FPS causing lag, and inaccurate position tracking using only two axes.

The most prevalent issue was inaccurate tracking. This problem arose due to the camera's perception of objects at different depths. For instance, consider two actors of the same height, each on a plane with uniform characteristics in Fig. 1.

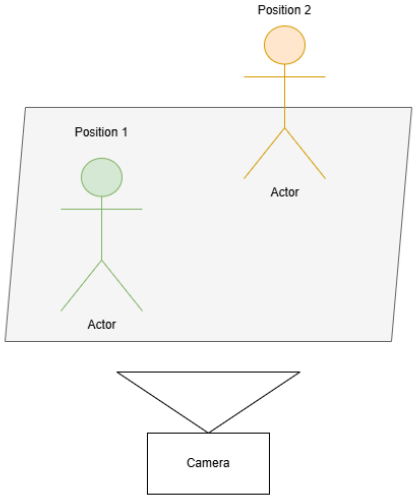



**Figure 1:** Real world Actors Position

For this example, we use the actors pelvises as reference points in Fig. 2. In the real world, their pelvises are at the same height (Y coordinate).

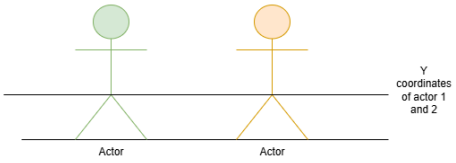



**Figure 2:** Height of actors pelvises

However, when viewed through the camera, the closer actor is projected at a larger scale compared to the one further away in Fig. 3. This means that in the image, the actor who is closer, but at the same height in the real world, appears higher on the image plane.

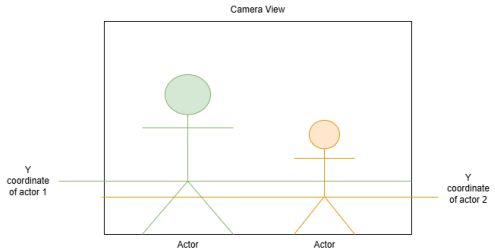



**Figure 3:** Real height of actors pelvises

This issue is significant because we are trying to detect when an actor jumps. This is usually done using Human Action Recognition techniques [4, 5, 6, 7]. However, these techniques are computationally demanding for our system. Therefore, in the older version of our system, we used a threshold to detect jumps. We detected persons in the image and took the middle of the bounding box of each person. We then checked if this midpoint was above a certain threshold. If it was, it meant that the person had jumped. This threshold was adaptive because not every person and environment are the same. While this approach worked, it introduced many false jump detections into our system, which was useful but not as enjoyable in the game play.

Due to these problems, we decided to develop a new version, focusing primarily on addressing the issue of incorrect tracking with only two axes.

Additionally, one of the motivations for this article was the high price of stereo cameras on the market. Stereo cameras are usually costly for regular use. However, our system is built using inexpensive webcams, improving accessibility to various applications. While these cheaper cameras are not as high-quality as the commercially available stereo systems, this limitation does not affect our use case and is suitable for character tracking.

## 2 Hardware and Software

In our initial attempt to build a stereo vision system, we used the ESP32-CAM module primarily for its affordability. The ESP32-CAM is a microcontroller with an integrated camera, offering both wireless and wired connectivity, which provides flexibility in our implementation. However, during testing, we encountered significant limitations. The image quality was poor, adversely affecting the performance of our deep learning model, which struggled to detect persons accurately. The frame rate was too low for real-time applications, rendering it unsuitable for our needs.

We replaced the ESP32-CAM with the Powerton Universal Webcams in Fig. 4 to address these shortcomings. This webcam offers a good balance between cost and performance, providing higher-resolution images and a better frame rate, which significantly improves the quality of our stereo vision system. The primary drawback of this webcam is that it can only be used with a wired connection.

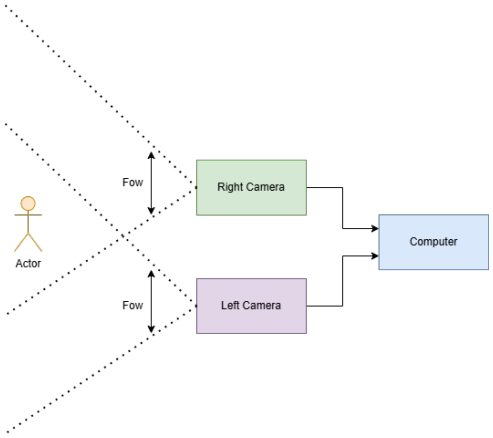



**Figure 4:** Stereo vision application hardware design

For the computational tasks required for image processing and deep learning algorithms, we used a notebook computer with the following specifications:

- **Processor**: Intel Core i5 - 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz
- **Graphics Card**: NVIDIA GeForce MX450
- **RAM**: 16.0 GB

The processor handles data input from the cameras, while the graphics card accelerates the computations of deep learning algorithms necessary for real-time object detection.

We utilized the OpenCV library for image processing, which is widely used for computer vision applications. Next, we employed MediaPipe, a Python library designed to implement deep learning algorithms in projects for object detection within images.

## 3 Implementation

The implementation section details the practical steps and methodologies employed to develop the stereovision system for real-time human tracking and interaction.

### 3.1 Camera Mounting

We secured both Powerton Webcams to the fixed base using a strong adhesive to ensure a stable and reliable stereovision system prototype in Fig. 5. Fixing the cameras in place is essential for maintaining consistent calibration, as any movement could disrupt the alignment and accuracy of the stereovision setup. The cameras were positioned at a predetermined distance from each other and oriented parallel to ensure appropriate fields of view overlap.

**Figure 5:** Stereo vision application

### 3.2 Image Collection for Calibration

For accurate calibration, capturing images from both cameras simultaneously is crucial. This synchronization ensures that corresponding points in each image pair are captured under identical conditions. We collected a series of images from each camera, capturing various objects and scenes at different distances and angles. In these images, we used a chessboard pattern that provides well-defined corners essential for calibration. These images served as the dataset for calibrating the stereovision system.

### 3.3 Camera Calibration

We used a series of collected images in Fig. 6, containing a chessboard pattern to perform camera calibration. The goal was to detect the chessboard corners in these images using the OpenCV library. Identifying these corners in both the left and right camera images is essential to ensure accurate calibration. The detected corner points from both images were then used as reference points for calculating the camera's intrinsic pa-

rameters. These parameters are crucial for accurately aligning the stereo-vision system and ensuring precise depth perception.

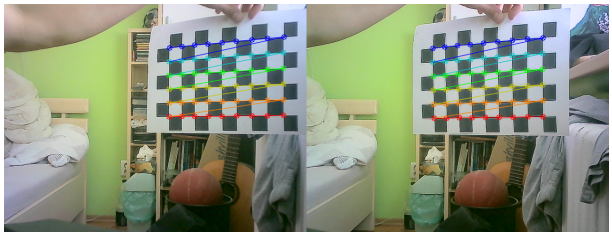


**Figure 6:** Left and Right image corner detection

Intrinsic calibration of a camera determines internal parameters that affect how images are captured. These parameters include the focal length, optical center, and lens distortion coefficients. We calculated these parameters using OpenCV functions and optimized them to reduce distortion and enhance image quality. This optimization process helps minimize the effect of lens distortion and improve the accuracy of the camera parameters.

Stereo calibration ensures that the spatial relationships between cameras are accurately determined and their images are aligned to enable depth perception [8, 9, 10]. This process involves determining the relative positions and orientations of two cameras, computing the rotation and translation vectors, and the essential and fundamental matrices, as shown in Fig. 7. These matrices describe the geometric relationships between corresponding points in the two images, which are necessary for accurate 3D reconstruction.

Using the parameters obtained from stereo calibration, we perform stereo rectification. Stereo rectification aligns the image planes of the two cameras, making subsequent depth estimation more straightforward and more accurate. This process involves computing rectification maps for both cameras, which are then stored for future use.

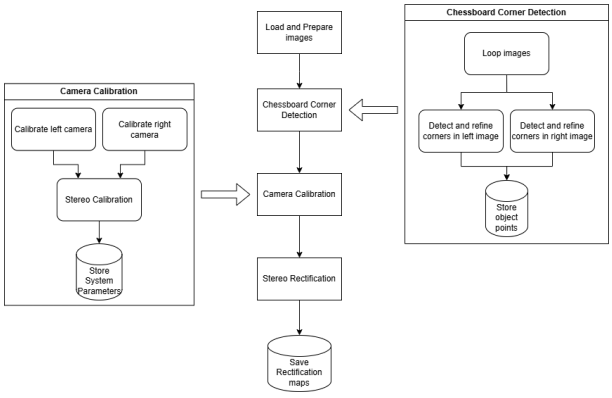



**Figure 7:** Camera Calibration Diagram

## 3.4 3D Tracking

For 3D tracking, we utilize a live stream of images from our stereo vision system, simultaneously capturing left and right images, as displayed in Fig. 8. We cannot use the raw images directly for depth estimation due to inherent lens distortions and misalignments. These distortions can significantly affect the accuracy of the depth calculation. Therefore, we apply stereo rectification and undistortion techniques using precomputed stereo maps. These stereo maps, calculated as described in Section 3.3, are essential for correcting lens distortions and aligning the images from both cameras. The rectification process involves remapping the pixels of the images to a common plane, ensuring that corresponding points in the left and right images are horizontally aligned. This alignment is crucial for accurate depth measurement. By applying the stereo maps to every pair of images in the live stream, we ensure that each frame is corrected for any distortions, providing a reliable basis for subsequent depth calculations. We use the deep learning library MediaPipe to detect and track person landmarks. MediaPipe is a versatile framework designed for building machine learning pipelines. It provides a robust solution for detecting and tracking person landmarks in real time. Once the hand landmarks are detected in both images, we can calculate the 3D position of these landmarks using triangulation. Triangulation involves measuring the disparity between the positions of corresponding landmarks in the left and right images [11]. The disparity is the horizontal shift between the same point in the two images, which is directly related to the depth of the point. In addition to depth, we can also compute the X and Y coordinates of the landmark in the 3D space relative to the camera's coordinate system.

## 3.5 Deep Learning implementation

For person detection, we utilize the MediaPipe deep learning framework, which includes multiple pipelines such as person detection, face detection, hand tracking, and pose estimation. MediaPipe enables rapid implementation of these pipelines, but other algorithms, such as Convolutional Neural Networks (CNNs), can also be used for object detection.

MediaPipe's person detection pipeline leverages advanced deep learning techniques to accurately detect and localize humans in images or video frames. The pipeline begins by preprocessing the data to ensure it meets the model's requirements. This typically involves resizing the image to a standard dimension, normalizing pixel values, and converting the image format if necessary.

The framework then employs Convolutional Neural Networks (CNNs) [12] for feature extraction. The lower layers of the CNN capture simple features such as edges and textures, while the deeper layers identify more complex patterns like shapes and object parts.

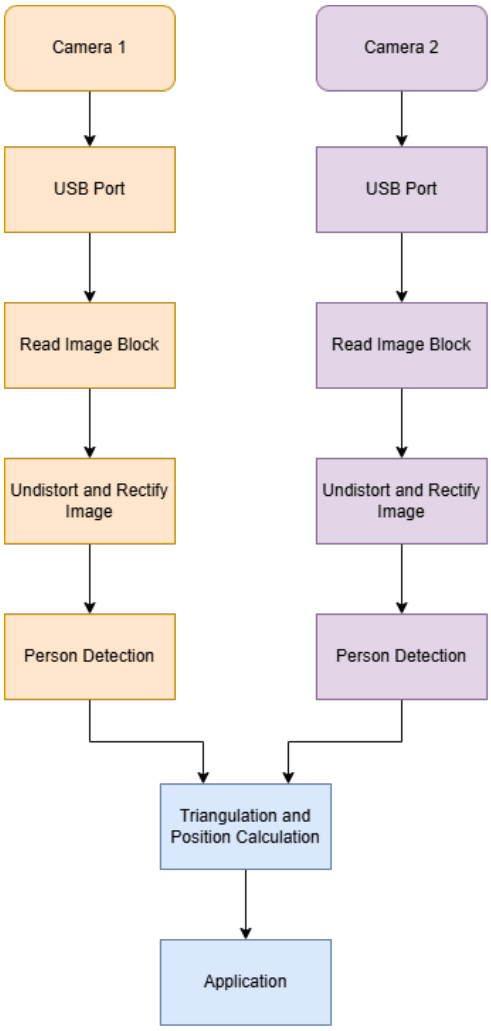



**Figure 8:** Tracking Diagram

These extracted features are used for the detection and localization of persons within the image.

MediaPipe utilizes networks such as Single-Shot MultiBox Detector (SSD) [13] or Region-Based Convolutional Neural Network (R-CNN) [14] for this process. These networks detect objects and draw bounding boxes around them, allowing for precise localization within the image. The person detection pipeline also includes post-processing steps, which refine the raw detection output to enhance accuracy and readability.

## 3.6 System Aplication

To demonstrate the application of our system, we selected the Chrome Dinosaur game. This choice stems from a previous project in which we developed a computer vision gaming controller for this game. However, the earlier controller had a limitation: it could only capture movement in the 2D plane, detecting changes along the X and Y axes. It posed a problem because our system could not accurately interpret its positional change when a detected point was closer to the camera.

Our new system overcomes this limitation by detecting objects in 3D space. This advancement ensures that the distance from the camera does not affect the object's perceived position on the X and Y axes. To simplify the process, we adjusted the system to detect the index finger in images and calculate its 3D position shown in Fig. 9. With this capability, we can accurately track the finger's position.

We implemented a threshold on the Y (height) axis to facilitate interaction as displayed in Fig. 10. The system can detect when the finger crosses the defined Y axis threshold by live tracking the finger's movement. When the finger crosses this threshold, the system simulates a press of the spacebar on the keyboard. After pressing the spacebar, the system waits for the finger to move back across the threshold before triggering another press. This mechanism enables the user to control the jumping action in the Chrome Dinosaur game. The player can effectively control the dinosaur's jumps by moving the finger up and down.

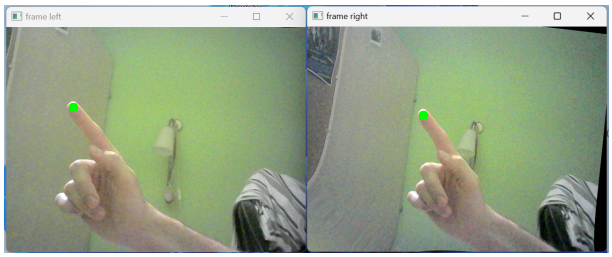



**Figure 9:** Detected finger Left and right image

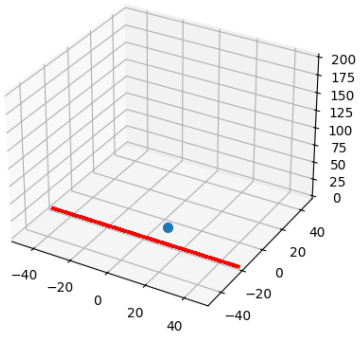



**Figure 10:** 3D Space with Y threshold

Compared to the first system, the new system does not require dynamic thresholding, enhancing its precision and reducing false jump click occurrence.

Additionally, our new camera system employs higher-resolution cameras, enabling the deep learning algorithm to recognize objects in images better.

The system operates at an average of 20 FPS when using a GPU. When using a CPU, the FPS drops to about 10. While both setups are playable, utilizing the

CPU results in noticeable latency, causing a slight delay between the finger crossing the threshold and the jump action. Although this latency does not render the game unplayable, it negatively impacts the gaming experience.

### 3.7 Performance Results

To evaluate the precision of our system, we conducted an experiment. We set detection to person detection and performed 100 jumps for each system and recorded instances where the system falsely registered a jump when no jump occurred. These 100 jumps were performed consecutively, and data was recorded in intervals of 10 jumps. The counter did not reset after each interval, but continued counting from the previous total.

We used an external button to manually count the jumps. While the system automatically counted the jumps, we manually pressed the button after each actual jump, incrementing a separate counter for real jumps. After every 10 real jumps, we calculated the difference between the system's jump count and the real jump count to determine the number of false jumps.

| Jumps | False Sys 1 | False Sys 2 |
|---|---|---|
| 10 | 4 | 0 |
| 20 | 7 | 0 |
| 30 | 8 | 2 |
| 40 | 13 | 3 |
| 50 | 14 | 3 |
| 60 | 17 | 5 |
| 70 | 22 | 5 |
| 80 | 23 | 5 |
| 90 | 27 | 5 |
| 100 | 29 | 6 |
| **Percentage:** | **29%** | **6%** |

**Table 1:** Comparison of False Jumps in Old System (Sys 1) and New System (Sys 2)

Table 1 shows that after 100 jumps, System 1 performed poorly, registering false jumps 29% of the time. In contrast, System 2 recorded only 6% false jumps, demonstrating a substantial improvement in performance. It's important to note that these metrics were obtained in a controlled environment and may vary in real-world situations where movements are more unpredictable.

Table 2 show a comparison of both systems. The old system is limited to 15 fps due to its camera setup, which affects its GPU performance. In contrast, the new system does not have this limitation, allowing it to achieve higher performance on a GPU.

However, the new system's performance is significantly impacted by the need to process two images simultaneously, rather than one. This impact is evident when running on a CPU, where the old system outperforms the new one. This means that if we used the old system with the new camera hardware, it would outperform the new system in terms of speed.

| Hardware | Old System FPS | New System FPS |
|---|---|---|
| CPU | 15 | 10 |
| GPU | 15 | 20 |

**Table 2:** FPS Comparison of Systems

## 4 Conclusion

This paper explored the application of stereo vision and deep learning techniques to enhance the gaming experience. By leveraging stereo vision combined with deep learning object detection, we developed a cost-effective system capable of tracking human movement in a 3D coordinate system.

Our approach integrates stereo vision technology with a deep learning model to detect and track person landmarks in real time. We addressed initial hardware limitations by transitioning to more suitable components, such as the Powerton Universal Webcam, which provides the image quality and frame rates necessary for accurate and responsive game control.

A comprehensive calibration process involving both intrinsic and extrinsic stereo calibrations ensures the precision of our 3D positioning system. Intrinsic calibration determines the internal parameters of each camera, while extrinsic calibration determines the relative position and orientation of the cameras, which is essential for accurate position triangulation in 3D space.

For object recognition, specifically identifying people and hands in the image, we utilized the Python library MediaPipe, which includes several deep-learning models for recognizing different objects. We applied our system to the Google Dino game, controlling a jumping dinosaur by moving a finger up and down. Our system demonstrated a significant improvement over the previous version, with fewer false jumps.

For performance comparison, we configured both systems for person detection and recorded the number of false jumps over 100 jumps. The results demonstrated that the new system is superior, achieving a 6% false jump rate compared to the old system's 29% false jump rate.

However, we observed a performance issue when using the system with a CPU instead of a GPU, resulting in slow jump responses. This problem is mitigated when running the system on a GPU.

## Acknowledgments

This work was supported by VEGA through the Research on inertial data analysis methods for ap-

plications in rehabilitation adjuvants under Grant 1/0095/24.

# References

E. Dandil and K. K. Çevik, "Computer Vision Based Distance Measurement System using Stereo Camera View," *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Ankara, Turkey, 2019, pp. 1-4, doi: 10.1109/ISMSIT.2019.8932817.

C. Loop and Zhengyou Zhang, "Computing rectifying homographies for stereo vision," *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, Fort Collins, CO, USA, 1999, pp. 125-131 Vol. 1, doi: 10.1109/CVPR.1999.786928.

Suenaga, H., Tran, H.H., Liao, H. et al. Vision-based markerless registration using stereo vision and an augmented reality surgical navigation system: a pilot study. BMC Med Imaging 15, 51 (2015). https://doi.org/10.1186/s12880-015-0089-5

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., Velastin, S. A. (2022). Video-based Human Action Recognition using Deep Learning: A Review. arXiv preprint arXiv:2208.03775. Retrieved from https://arxiv.org/abs/2208.03775

Kulbacki, M., Segen, J., Chaczko, Z., Rozenblit, J. W., Kulbacki, M., Klempous, R., Wojciechowski, K. (2023). Intelligent Video Analytics for Human Action Recognition: The State of Knowledge. Sensors, 23(9), 4258. https://doi.org/10.3390/s23094258

Shaikh, M. B., Chai, D., Islam, S. M. S., Akhtar, N. (2024). From CNNs to Transformers in Multimodal Human Action Recognition: A Survey. ACM Transactions on Multimedia Computing, Communications, and Applications. Association for Computing Machinery (ACM). http://dx.doi.org/10.1145/3664815

Xie, Y. (2024). Deep Learning Approaches for Human Action Recognition in Video Data. arXiv preprint arXiv:2403.06810. Retrieved from https://arxiv.org/abs/2403.06810

Nico Nielsen, "Stereo Vision Depth Estimation," *GitHub repository*, 2023. Available: `https://github.com/niconielsen32/ComputerVision/tree/master/StereoVisionDepthEstimation`.

Satya Mallick, "Depth Perception Using Stereo Camera – Python C++," *LearnOpenCV*, 2023. Available: `https://learnopencv.com/depth-perception-using-stereo-camera-python-c/`.

Satya Mallick, "Stereo Vision and Depth Estimation Using OpenCV AI Kit," *LearnOpenCV*, 2023. Available: `https://learnopencv.com/stereo-vision-and-depth-estimation-using-opencv-ai-kit/`.

Hartley, R. and Zisserman, A. Multiple View Geometry in Computer Vision. 2nd ed. Cambridge University Press; (2004).

O'Shea, K. and Nash, R. An Introduction to Convolutional Neural Networks. (2015). arXiv:1511.08458. https://arxiv.org/abs/1511.08458

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A.C. SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science*. Springer International Publishing. (2016), 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Girshick, R., Donahue, J., Darrell, T., and Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. (2014). arXiv:1311.2524. https://arxiv.org/abs/1311.2524

# A Lightweight communication protocol for Body Area Networks

**Juraj Ďud'ák**
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava
Jana Bottu 2781/25, Trnava, Slovakia and
University of Zilina, Research center
Univerzitná 8215/1, Žilina, Slovakia
juraj.dudak@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

**Martin Skovajsa**
University of Zilina
Research center
Univerzitná 8215/1, Žilina, Slovakia
martin.skovajsa@uniza.sk

**Abstract.** *This article introduces nanoBus (nBus), an innovative RS485-based communication protocol designed specifically for Body Area Networks (BANs). To enhance real-time monitoring and feedback of health parameters, nBus addresses the growing need for reliable and secure patient data transmission, especially in clinical and home rehabilitation contexts. Utilizing the RS-485 serial interface, nBus ensures high data transfer rates and robust resistance to interference, a notable improvement over wireless technologies in terms of security and reliability. The protocol operates on a master/slave request/response model, facilitating node and sensor addressing within a BAN. The architecture of nBus is thoroughly detailed, highlighting its three primary layers: Physical, Link, and Application. This comprehensive design allows for scalable integration of up to 132 nodes with 32 sensors each, promising significant advancements in BAN communication. The paper concludes with prospects, emphasizing the protocol's potential implementation in intelligent clothing and the challenges therein, particularly in cable management.*

**Keywords.** nBus Protocol, RS-485 Communication, Body Area Network, Health Monitoring, Wearable Sensors

## 1 Introduction

As the World Health Organization notes, the number of aging people and patients with chronic diseases, obesity, and other disabilities is steadily increasing and becoming an unavoidable problem. It will likely contribute to a significant deterioration in the quality of care in hospitals and health centers as more patients are readmitted daily (Rahat, 2018). The need for optimal and rapid healthcare solutions to tackle the looming health crisis is the focus of many researchers (Ullahm, 2012).

In outpatient practice, many vital functions (heart rate, blood pressure, body temperature, and similar) are monitored. However, they are rarely monitored simultaneously, often resulting in an incomplete picture of the patient's health status. To solve this problem, non-intrusive and real-time ambulatory health monitoring offers an efficient and cost-effective way to treat patients, as vital signs can be recorded and medical data updated in real-time (Ullahm, 2012).

Small sensor devices that can monitor various patient data have become increasingly popular in the healthcare industry in recent decades. It is mainly due to advances in nanotechnology, which have enabled the development of nano biosensors with high sensitivity and a range of functions for monitoring and diagnosis even at the cellular level (Javaid, 2022). This technological progress is particularly noticeable in the age of the Internet of Things (IoT), big data, and cloud computing. In terms of human health monitoring, the Body Area Network (BAN), also known as the Body Sensor Network (BSN), seems to be a prime candidate among the most promising communication technologies (Ullahm, 2012).

These networks consist of small, low-power, and intelligent wearable sensors that are placed in, on, or around the human body. These sensors constantly collect data on the health status and activities of the person and store it for further processing (Rahat, 2018; Ullahm, 2012). As part of these networks, multiple actuators (displays, speakers, buzzers,

and others) can also be implemented to provide feedback to the user in an interactive way (Zhang, 2009). The communication hierarchy in BANs is generally divided into three parts, as shown in Fig. 1:

- *Intra-BAN*, which includes communication between sensors, actuators, and a control unit,
- *Inter-BAN* consisting of personal devices and other BAN instances,
- *Beyond-BAN* relates to data transmission over the WAN, where it can be stored, analyzed, and action taken in an emergency.

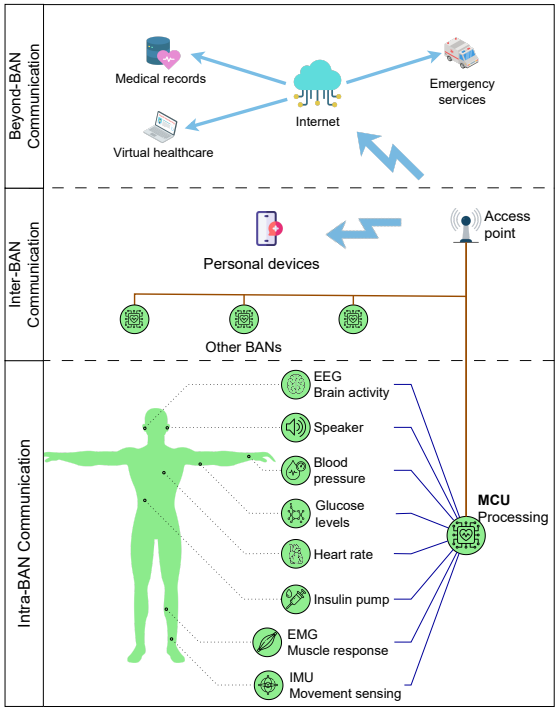



**Figure 1:** BAN Communication hierarchy

### 1.1 Related Work

Recently, the main focus has been put on the development of new communication protocols for intra-BAN communication, especially for wireless applications. Wireless technologies (IEEE 802.15.6, Bluetooth, ZigBee, and others) have many advantages. The most important one is mobility, as the central concept of a BAN is that it cannot restrict the user in any way. There are also low-power technologies that allow battery-powered devices to last a long time. However, the protocols based on wireless sensor networks (WSN), which consist of stationary and redundant sensor nodes, are not often reliable in a BAN environment with much movement and suffer from interference, packet loss, and latency issues.

Using air as a transmission medium also raises the issue of security, especially as patient data is susceptible. These challenges put researchers and manufacturers alike under considerable pressure (Shadi, 2021). In this detailed study, the author deals with the placement of sensors and the human body, the possibilities of their mutual communication.

Patel in his paper (Patel, 2010) lists appropriate technologies, minimum transmission rates, minimum latency, and required uptime as parameters to use for tracking the desired biosignal. He also considers QoS and communication security as a key factor in selecting the appropriate technology. All the technologies described are exclusively wireless.

In the paper (Li, 2008), the authors describe the standardization of BANs according to the IEEE 802.15.BAN standard. Low power consumption, guaranteed response to external stimulus were defined as key parameters in the implementation. As transmission medium they define: air, vicinity of human body, inside human body. The transmission rate is defined from 100 bps (heartbeat, blood pressure, body temperature) up to 2500 bps (Electro cardiogram).

The proposed nBus protocol aims to meet the requirements of high transmission rate (up to 2 Mbps), wireless link, higher number of connected sensors (up to 32 sensors), transmission reliability and energy efficiency.

## 2 NanoBus (nBus) Protocol

The architecture of nBus is thoroughly detailed, highlighting its three primary layers: Physical, Link, and Application. The Physical Layer emphasizes using RS-485 for its industrial-grade reliability and high data rates, which are suitable for short BAN bus lengths. The Link Layer focuses on efficient node addressing and error checking, including an 8-bit cyclic redundancy check (CRC-8). The Application Layer encompasses the translation of requests and responses, defining up to 128 function codes for various operations.

The nBus protocol is an application layer protocol for connecting sensor nodes to a higher-level control unit in the BAN. The main development goal is to provide a reliable and secure way for real-time monitoring and feedback of health parameters during clinical or home rehabilitation and physiotherapy. The communication model of nBus is shown in Fig. 2.

### 2.1 Physical Layer

The communication is a master/slave request/response model, allowing node addressing and specific sensor addressing, as a communication node might consist of sensors and actuators. The packet structure consists of a link frame and an application packet as shown in Fig. 3.

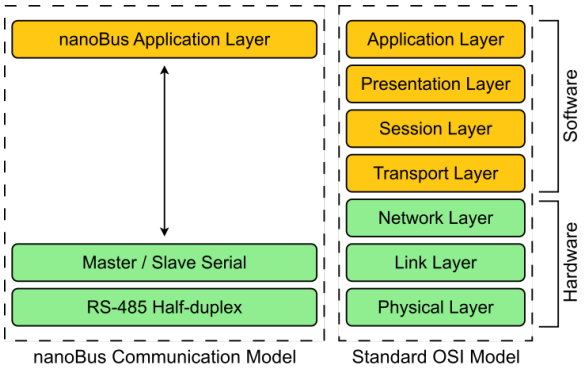



**Figure 2:** Communication Layer Model

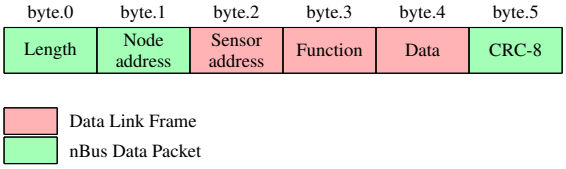



**Figure 3:** nBus packet structure

To achieve reliable and secure data transmission between the nodes, we have opted for an RS-485 serial interface, an industry standard with high reliability and resistance to interference. Since the transmission is via cable, it is more tap-proof than wireless technologies. RS-485 is a half-duplex mode that uses only two twisted differential wires for data transmission, often combined with a common ground wire in a single cable. It was developed for connections in harsh industrial environments, enabling high data rates over long distances.

The standard states that a rate of 100 kbit can be achieved with a bus length of approx. 1200 m (Texas Instruments, 2008) with a maximum of 32 nodes on one bus due to current loads. However, as the bus of a BAN is very short ($< 5\ m$), even 100 Mbit rates are achievable if high-end transceivers are used. These have a 1/5 bus load, meaning up to 160 nodes can be connected to a single bus without repeaters. In addition, modern integrated circuits can be put into a low-power standby mode when no transmission occurs, extending battery life for portable devices (Renesas, 2017).

A crucial thing to consider is cable management. The data rates for high-speed transmission use a CAT5 cable as a reference. The ANSI/TIA/EIA-568-B standard dictates a minimum wire gauge of 26 American Wire Gauges (AWG) or 0.14 mm$^2$ (ANSI/TIA/EIA, 2001). By using thin 3-wire cables, we could reduce the impact on patient's mobility and increase the potential for smart clothing use. However, it also poses new challenges, with maintenance being the prime candidate as thin wires are prone to breaking with repeated movement.

## 2.2 Link Layer

The link layer is responsible for node addressing and error checking. Each node in a BAN has its unique node address. Each frame contains an 8-bit cyclic redundancy check (CRC-8) to implement error detection during transmission. This check is the last byte transmitted, and to account for the dynamic frame length, the frame length is included as the first byte. The maximum frame length is currently set to 132 bytes.

## 2.3 Application Layer

The application layer is responsible for the translation of requests and responses. It works either in slave or master mode. In a single nBus network, only one master is permitted, which sends data requests, slave commands, and application error handling. The information about the target service is contained in a single function byte and supports up to 128 function codes. The MSB of the function byte is reserved for an error flag.

In the following terminology, the following terms are used: master node — a device connected to the BAN that communicates with slave node devices, slave node — a device in the BAN to which sensors sensing signals of interest are connected, sensor — a console sensing element, a single slave node may contain multiple sensors. Each sensor can be addressed individually or as a group.

These functions range from version information, data request, node, and sensor parameter configuration to node echo, calibration, sleep, and wake requests. nBus supports broadcast for the entire network (e.g., clock synchronization across nodes or calibration of all inertial sensors), broadcast for individual node sensors (e.g., setting the same gain for each force sensor in a single node), and also node sensor unicast (i.e., sending data for display on an integrated display).

| bit.7 | bit.6 | bit.5 | bit.4 | bit.3 | bit.2 | bit.1 | bit.0 | |
|---|---|---|---|---|---|---|---|---|
| Type | | | Sensor address | | | | | group |
| $I/\bar{O}$ | Res | Res | $a_4$ | $a_3$ | $a_2$ | $a_1$ | $a_0$ | bit |

**Figure 4:** Sensor Address byte structure

The target sensor is specified in the node address byte as shown in Fig. 4. The first three bits are used to indicate the sensor type, with the first bit's designation as an input/output selector and two reserved for later versions. The remaining bits are used for sensor addressing. This way, we can integrate up to 31 sensors and actuators into a single node.

The slave node listens to the master's requests and generates corresponding responses. The distinction between transmission types is made by node and sensor address bytes, as shown in Tab. 1 where $X \in (1,\ 127)$ represents a node address and Y being an sensor address ($Y \in (1,\ 31)$ for outputs, $Y \in (129,\ 159)$ for

inputs). Both slave and master nodes implement transmission error detection and correction. In the event of a transmission error, the node that detects the error sends a data packet with a set error flag and an error code as data. This code can be anything from a bad CRC result to an invalid parameter request.

**Table 1:** nBus address bytes based on transmission type

| Node Address | Sensor Address | Description |
|---|---|---|
| 0 | 0 | Broadcast to all nodes and sensors. |
| 0 | Y | Broadcast to specific sensors. |
| X | 0 | Unicast to slave node (to all node sensors). |
| X | Y | Unicast to node sensor. |

# 3 nBus Application Protocol

The nBus protocol was designed with an emphasis on minimizing communication overhead by allowing the use of high data rates and advanced addressing options (unicast/broadcast/group broadcast). This protocol is designed as a library independent of the hardware platform used. In the current version, this library is implemented for the STM32L031 (Slave module) and Nordic NRF52820 (Master module) hardware platforms. The nBus library has been programmed in C, which allows it to be used in both C and C++ projects. The control firmware was programmed in the STM32CubeIDE development environment (the STM32L01x microcontroller was used) and Visual Studio Code/Platformio (for the NRF52820 microcontroller).

For easy implementation of the nBus protocol on the selected hardware platform, the *nBusPlatformInterface_t* interface is available to provide a connection to the platform on which the library is implemented. This interface consists of two layers containing elementary functions. These functions need to be implemented on a specific microcontroller:

- **Communication layer**. Receiving and sending data packets:
  - `uart_receive()` — receiving a data packet in non-blocking mode,
  - `uart_transmit()` — sending a response from the application layer of the nBus protocol.
- **System Layer**. Notification of the status of connected sensors:
  - `loop_callback()` — nBus application layer notification of external events,
  - `delay_ms()` — blocking delay.

The second part of the slave module functionality consists of the connected sensors. For a uniform implementation of the connected sensors in the slave module, the *nBusAppInterface_t* interface has been prepared to provide uniform access to the data and settings of the left-handed sensor connected to the module. This interface provides the following functionality:

- initialization and reset of sensor settings,
- obtaining information about the number and type of connected sensors within the hardware module,
- setting and reading the configuration parameters of the connected sensors,
- start and stop the automatic sensor reading process,
- read the value from the sensor(s) of the module.

These 2 interfaces ensure the independence of the hardware platform used. When you need to connect a module with a new type of sensors to the BAN, you only need to implement the *nBusAppInterface_t* interface.

Each module must store the configuration of the connected sensors in non-volatile memory. The internal design of the memory map was designed to be able to store the configuration of 16 sensors, each sensor can have up to 8 parameters. The parameters were chosen to suit the widest possible use (time base, resolution, gain, offset, sample rate, range 1, range 2, filter). The use of these settings is optional. The *nBusAppInterface_t* interface, which implements the functionality for each existing sensor, must indicate whether a given parameter is implemented for a particular sensor.

## 3.1 Sensoric module nBus slave

In the actual version, we have prototyped an nBus module with a 6-axis IMU sensor. The module is built on the STM32L0130 MCU — Fig. 5.

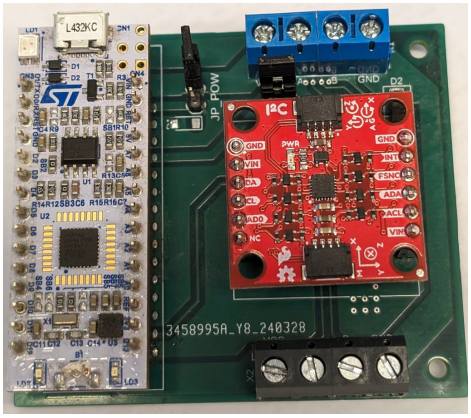



**Figure 5:** nBus module with IMU sensor

Properties of nBus slave module on Fig. 5:

1. communication interface RS485, with baud rate up to 2 Mbps,
2. Connected sensors: 1 * accelerometer, 1 * gyroscope (in three axes),
3. maximum frequency of IMU sensor readings: 260 Hz,
4. it contains internal non-volatile memory for storing the configuration of the module itself and all connected sensors,
5. device does not contain a battery, the power supply of the module is solved via RS-485 bus.

An important aspect of the overall design is the ability to ensure data transmission at the required quality. In the context of a BAN, we consider bus transmission capacity as one of the main quality indicators. According to the above proposal, a BAN can have up to 127 nodes, with each node having up to 31 sensors. Thus, in total, the network can contain up to 3937 sensors.

Our design includes 2 network interfaces: an RS-485 interface for the internal network and an RF interface for communication with the external system. The communication in the internal network must be conflict-free and with maximum speed. This is ensured by the master/client architecture.

$$C = (N * Data_{size} + C_{overhead}) * f \qquad (1)$$

To calculate the required baud rate of the internal bus, we use equation 1.

Where C is the required communication capacity of the bus for a single slave module. N is the number of active sensors, $Data_{size}$ is the word width (number of bytes per data), $C_{overhead}$ is the communication overhead and $f$ is the communication frequency. We consider with N=31, $Data_{size} = 2$, $C_{overhead} = 10B$ (request length and overhead in the slave module response) and $f = 100Hz$.

$$C = (31 * 2B + 10B) * 100Hz = 7200B/s \qquad (2)$$

For a theoretical number of modules n=127, we get the required capacity of the communication link $C = 914400B/s$.

### 3.2 Future work

The current aim of the development is to implement the nBus protocol in embedded applications, with the target processing unit being low-power 32-bit microprocessors. A hardware platform consisting of a high-speed RS-485 interface, analog-to-digital converters, a USB debugging interface and an STM32L031 microprocessor from STMicroelectronics is being developed for testing purposes. Power management is also an important part of the overall design. Setting the minimum power consumption of the nBus slave is determined by using the appropriate operating modes of the microcontroller, but also by taking advantage of the sleep modes of the sensors used. Another criterion is the actual mechanical design of the whole module. As this module will be attached to a human limb, the overall weight needs to be minimized so that the weight of the device itself does not affect the measurement results.

We are currently working on the nBus master module, which will additionally contain a battery to power the entire nBus network and a hardware interface for wireless transmission of measured data from the nBus slave modules.

## 4 Conclusion

In this article, a new communication protocol for Body Area Networks has been proposed. The main goal of the design is to develop a reliable and straightforward protocol for real-time monitoring of health parameters with high scalability and the ability to implement feedback elements in the system. nBus is an application layer protocol and uses request/response communication between master and slave nodes. An RS-485 industrial bus is the physical layer to ensure high data rates over 2-wire connections. It lightens data security and encryption requirements and, therefore, processing times and energy consumption, as the wired transmission of sensitive patient data is less susceptible to interception. It also opens up new possibilities for implementation in sensor-enabled clothing and poses new challenges in cable management and maintenance.

The nBus data link layer addresses the nodes and checks for transmission errors using CRC-8. Up to 127 nodes are currently supported by the protocol, with each node containing up to 62 sensors (31 inputs and 31 outputs). The central data unit of the protocol is the nBus data packet, which contains the sensor address, function, and data bytes. Further protocol testing still needs to be carried out to obtain objective indicators of the protocol's performance. However, current development focuses on implementing a customized hardware platform based on an STM32L031 32-bit microprocessor (32 kB FLASH, 8 kB RAM, $f_{max} = 32MHz$) and NRF52840 (512 kB FLASH, 128 kB RAM, $f_{max} = 64MHz$)

To summarise, nBus is a novel communication protocol that has the potential to simplify data transmission in BANs and enable reliable communication in real-time applications.

nBus is a pivotal development in the field, offering an efficient and scalable solution for BAN communications, crucial for modern healthcare monitoring systems

## Acknowledgments

This work was supported by VEGA through the Research on inertial data analysis methods for applications in rehabilitation adjuvants under Grant 1/0095/24.

## References

Min Chen, Sergio Gonzalez, Athanasios Vasilakos, Huasong Cao, and Victor C. M. Leung. (2011) Body area networks: A survey. *Mobile Networks and Applications*, 16(2):171–193. DOI: 10.1007/s11036-010-0260-8.

Sana Ullah, Henry Higgins, Bart Braem, Benoit Latre, Chris Blondia, Ingrid Moerman, Shahnaz Saleem, Ziaur Rahman, and Kyung Sup Kwak. (2012) A comprehensive survey of wireless body area networks. *Journal of Medical Systems*, 36(3):1065–1094. DOI: 10.1007/s10916-010-9571-3.

Shumaila Javaid, Sherali Zeadally, Hamza Fahim, and Bin He. (2022) Medical sensors and their integration in wireless body area networks for pervasive healthcare delivery: A review. *IEEE Sensors Journal*, 22(5):3860–3877.

Khalid Hasan, Kamanashis Biswas, Khandakar Ahmed, Nazmus S. Nafi, and Md Saiful Islam. (2019) A comprehensive review of wireless body area network. *Journal of Network and Computer Applications*, 143:178–198. DOI: 10.1016/j.jnca.2019.06.016.

Rahat Ali Khan and Al-Sakib Khan Pathan. (2018) The state-of-the-art wireless body area sensor networks: A survey. *International Journal of Distributed Sensor Networks*, 14(4):1550147718768994. DOI: 10.1177/1550147718768994.

Mi Zhang and Alexander A. Sawchuk. (2009) A customizable framework of body area sensor network for rehabilitation. In 2009 *2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 1–6. DOI: 10.1109/ISABEL.2009.5373703.

Rim Negra, Imen Jemili, and Abdelfettah Belghith. (2016) Wireless body area networks: Applications and technologies. *Procedia Computer Science*, 83:1274–1281. DOI: 10.1016/j.procs.2016.04.266.

Rui Li, Daniel T H Lai, and Weesit Lee. (2017) A survey on biofeedback and actuation in wireless body area networks (WBANs). *IEEE Rev Biomed Eng*, 10:162–173, DOI: 10.1109/RBME.2017.2738009.

Muhammad Shadi Hajar, M. Omar Al-Kadri, and Harsha Kumara Kalutarage (2021). A survey on wireless body area networks: architecture, security challenges and research opportunities. *Computers & Security*, 104:102211, DOI: 10.1016/j.cose.2021.102211.

Texas Instruments (2008). The RS-485 Design Guide,. Rev. D.

Renesas. ISL3259E RS-485/RS-232 *Transciever Datasheet*, 8 2017. Rev. 2.

ANSI/TIA/EIA (2001). TIA/EIA-568-B.2: *Commercial Building Telecommunications Cabling Standard*. Rev. B.2

M. Patel and J. Wang, (2010) "Applications, challenges, and prospective in emerging body area networking technologies," in IEEE Wireless Communications, vol. 17, no. 1, pp. 80-88, February 2010, doi: 10.1109/MWC.2010.5416354

Li, HB., Kohno, R. (2008). Body Area Network and Its Standardization at IEEE 802.15.BAN. In: István, F., Bitó, J., Bakki, P. (eds) Advances in Mobile and Wireless Communications. Lecture Notes Electrical Engineering, vol 16. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-79041-9_12

# Proposal for a weather monitoring system utilizing IoT technologies

**Roman Budjač, Gabriel Gašpar**
University of Žilina
Research Centre
Univerzitna 8215/1, 010 26 Žilina, Slovakia
{roman.budjac, gabriel.gaspar}@uniza.sk

**Maroš Valášek**
University of Žilina, Faculty of Electrical Engineering and Information Technology
Department of control and information systems
Univerzitna 8215/1, 010 26 Žilina, Slovakia
valasek.maros@gmail.com

**Martin Bartoň**
Slovak University of Technology, Faculty of Materials Science and Technology in Trnava
Institute of Applied Informatics, Automation and Mechatronics
Jana Bottu 25, 917 24 Trnava
martin.barton@stuba.sk

**Abstract.** *The rapid integration of modern technologies in everyday life, embodied by the Internet of Things (IoT), has significantly enhanced the efficiency and convenience of various sectors, including industrial systems. This article explores the design and implementation of a meteorological data collection, processing, and visualization system, focusing on the development and implementation of an IoT device capable of providing and transmitting these data. Additionally, we have developed a comprehensive dashboard to store, visualize, and analyze all incoming data in real-time. This system provides the capability for real-time data analysis and environmental monitoring.*

**Keywords.** data acquisition, IoT, microcontrollers, esp32, MQTT, Smart City

## 1 Introduction

In the current digital age, the integration of modern technologies into the daily functioning of society has become inevitable. This phenomenon extends to a wider range of devices that are categorized as the Internet of Things (IoT). These devices, found in homes, workplaces, or public spaces, for example, significantly increase the efficiency and convenience of our daily activities. IoT devices find applications across a wide range of sectors, from industrial production systems to the provision of data within cities and transport. The spread of IoT devices creates vast new opportunities for data collection, ultimately leading to improved efficiency, simplified tasks, and better decision-making.

Currently, the concept of Smart Cities could be implemented more effectively, and cities and regions worldwide are facing similar challenges and problems. Rapid urbanization and rapid population growth increasing demands on city infrastructure. This causes increasing demands on housing, rising energy prices, and also the need to reduce the impact of various crises, whether climatic, economic, or health, increasing the burden on the budget of local governments and the state. As we mentioned, rapid urbanization is demanding and putting pressure on all aspects of living in the cities(Ministry of Investments, 2023).

The Smart City concept not only offers solutions for so-called Smart Cities, but it is a concept that can be applied to many sectors. These can include, for example, smart housing and infrastructure, smart public transport, smart industry, smart energy grids, smart nature and more. IoT devices are used in the concept. The Smart City framework is part of the Industry 4.0 concept (Bauer et al., 2021).

In the Industry 4.0 concept, IoT can be used to develop so-called smart products. Things like the Internet of Things, the Internet of Services, the Internet of Energy, and others can be considered elements that create a link between the concepts of Smart City and Industry 4.0. Industry 4.0 can be considered as part of the Smart City concept (Lom et al., 2016). The development of

Industry 4.0 concepts and Smart Cities can solve problems related to electrical efficiency, urban production, demographic changes in large cities, and more.

This paper focuses on designing and implementing a system for collecting, processing, and visualizing meteorological data, as well as developing and implementing an IoT device that will provide and send this data. The solution assumes that using server containers will allow for efficient scalability and adaptability of the system when performance increases are needed. The server solution will also include functionality for monitoring and managing the status of the weather station and its sensors, including an automatic notification system for administrators when a malfunction of the device or sensors is detected. This project is a first attempt at creating a system to collect meteorological data. In the future, we plan to include a neural network for predictive capabilities.

The design phase will primarily focus on the weather station connectivity and ability to process and send data to server as well as communication architecture of the whole system will be created. The selected software and hardware solutions will reflect current technology standards and free open source technologies in the field.

# 2 Materials and Methods

The use of meteorological data in the Smart City concept is indispensable. Mainly because the environmental conditions in and around the city directly impact the comfort and quality of life not only of the inhabitants but also of the large number of working people who commute to the cities every day for work. Environmental conditions are also attractive for working people who commute to cities for work, perhaps even more interesting. According to the European Commission (EC), 15 percent of workers commute to cities for work (European Commission, 2019). This shows the importance of building information systems and making measured data available on the Internet so that it is also accessible to people who do not live in the city but work there. Measuring, collecting, and evaluating meteorological data in different parts of the city is even more important as cities and their inhabitants will have to face the challenges of extreme weather events, climate change, and their impact on urban infrastructure and inhabitants. High temperatures and poor air quality can significantly affect the health of urban residents. Smart City concepts use advanced technological solutions to monitor and predict air quality, enabling cities to implement effective adaptation measures (Camarillo-Escobedo et al., 2022).

Measuring and monitoring air quality and other environmental data in Smart Cities is an essential part of mapping the environment in which residents live, work, and spend their leisure time. This data then needs to be stored and analyzed appropriately. For example, high temperatures and poor air quality can significantly affect the health of urban residents. The Smart City concepts use advanced technological solutions such as deep learning to monitor and predict air quality based on historical data, enabling cities to implement effective adaptation measures (Tahirkheli et al., 2021). Intelligent air quality monitoring systems in Smart Cities often use a combination of low-cost sensors and advanced algorithms to create accurate pollution maps. These systems allow cities not only to monitor current pollution levels but also to predict future pollution levels.(Bibri and Krogstie, 2020).

This allows cities to introduce targeted measures to reduce residents' exposure to air pollution, for example, at times of high concentrations of dust particles and high temperatures. With the detailed mapping of parts of the city, such alerts and warnings can be specified for individual mapped areas, and thus, specific measures can be developed for individual areas and zones within the city (Neo et al., 2023). Using meteorological data in urban traffic management is essential, especially during severe weather conditions. For example, research shows how deep learning-based models, Bekkar et al. use meteorological data to predict air pollution, which has a direct impact on traffic management and public health. These data allow cities to take precautions to minimize the impacts of pollution (Neo et al., 2023).

Meteorological stations are essential in the Smart Cities concept because they not only provide the necessary meteorological data but also support decision-making processes in managing environmental objectives and planning infrastructure development. They collect important weather information critical for traffic management, energy management and crisis response to extreme weather conditions. A fundamental need for Smart City development is to collect data on what is happening in the city (Bibri and Krogstie, 2020). In Smart Cities, IoT technologies integrate this data into city information systems, enabling more accurate forecasts and adequate responses to weather. Advanced models based on deep learning, studies show, effectively predict air pollutants by leveraging data from these stations, allowing cities to take preventive action to protect citizens' health. Thus, integrating these stations into Smart Cities is crucial for developing Smart City services that respond to citizens' needs and improve safety and the environment in urbanized areas (Cheon and Mun, 2023).

## 2.1 Proposal of weather station based on IoT concept

The proposed system will be designed to collect sensor data from IoT devices. The submitted manuscript consists of several parts. The design of the weather station hardware, the design of the sensors, and the physical connection to the server. The next part deals with

the server's design and the system's implementation to collect data from the microcontrollers and transmit it to the server using MQTT protocol (Message Queuing Telemetry Transport) and Node-RED. The last part of the paper discusses in detail the implementation of the InfluxDB database system for storing real-time sensor data in the database and visualizing it using Grafana. The paper aims to design a weather station data acquisition system for collecting, storing, and visualizing the measured data. Also, the system provides capabilities for subsequent data analysis of historical data that has been measured. Above, the server will have alarm values set when some environmental data values are exceeded and will alert the users by email. All software is stored in a containerized docker system for easy future modifications and to ensure consistency across multiple development, release cycles, and production environments.

For these purposes, we have made our own weather station prototype. This is the first prototype of the weather station and also of the proposed data collection system on the server. This is the first step in our effort to create a network of measurement points and to create a detailed map between weather stations based on environmental data in order to obtain detailed environmental data from the campus environment in order to implement our devices concerning the Smart City concept.

Once the data is received into the system, processing, and computation will take place, such as calculating other data, sending alerts based on the value of the given data, or modifying the data into the correct data structure before storing it. After the data has been processed, all the retrieved and calculated data will be stored in the database system. The system will also be used to monitor the logs from the system to ensure that they can be accessible if needed. The proposed data storage system will monitor whether there is communication between the device and the server. The last part of the system will be data visualization through dashboards. In this proposal, the microcontroller will be used in the role of an IoT device will be used to collect data for the server.

The device will be constructed as a small weather station that will provide all the necessary data about the current weather. The device will also monitor the voltage of the batteries from which it will be powered. A solar panel will also be connected to the device. In addition to the data from the weather station, the device will send the functional status of the individual sensors.

In the first part, we will discuss the theoretical knowledge about the hardware part of our system for both the server part and the microcontroller part. We will discuss the sensors used, the microcontroller and the device on which we will run the server part. We have chosen sensors that will allow us to measure air temperature, relative humidity, air pressure, wind speed and direction, and precipitation. In this section, we will discuss the different sensors that we use in our implementation.

### 2.1.1 Hardware components

*ESP32 microcontroller selection*
The ESP32 refers to a chip called ESP32. However, the term ESP32 is also used to refer to development boards that are built on ESP32 chips. Using pure ESP32 chips is neither easy nor practical, especially when learning, testing, or prototyping. In most cases, you will want to use an ESP32 development board. These boards already include all the necessary components to power and program the chip, connect the chip to the computer, connect pins to peripherals, the necessary antennas for Wi-Fi and Bluetooth, and other valuable conveniences. Also, some of these boards may already contain other useful devices such as sensors or modules, screens, or cameras (RTD, 2022).
*Server based on Raspberry Pi 4*
The Raspberry Pi is a single-board computer developed by the Raspberry Pi Foundation. It has a small footprint but enough power for classic uses such as web browsing or creating small home servers. Raspbian is recommended for classical device use up to operating systems designed for servers. There are also various other operating systems defined e.g. as emulators, systems designed for multimedia, but also for 3D printing and home automation. There are also several versions of the Raspberry Pi, ranging from small computers such as the Raspberry Pi 4B to even smaller microcontrollers called Raspberry Pi Pico, which are suitable for IoT devices whose size is negligible. The Raspberry Pi is an inexpensive computer running Linux, but it also provides a certain amount of pins that allow them to control electronic components for physical computing and IoT exploration (RaspberryPi Ltd., 2024). The Raspberry Pi is a low-cost computer running Linux, but it also provides a certain amount of pins that allow them to control electronic components for physical computation and IoT exploration. We chose Linux as the operating system, as most servers run on Linux, and it also presented an opportunity to learn something new. Among the Linux distributions, we chose Ubuntu Server, which can be considered as a distribution suitable for beginners as well as for professional use.
*Temperature and humidity sensor DHT22*
The DHT22 temperature and humidity sensor, known as the AM2302, has high accuracy in both temperature and humidity measurement. It consists of a capacitive part, which is used to measure relative humidity, and an NTC thermistor, which is used to measure temperature. Each sensor is calibrated at the time of manufacture. The small size of the sensor, the very low power consumption, and the possibility of data transmission via cable up to a distance of about 20 meters supports a large range of applications of this sensor, such as IoT devices (Sparkfun, 2024).
*Temperature, humidity and barometric pressure sensor*

*BME280*
The BME280 is a sensor that measures the environment's temperature, barometric pressure, and humidity. The sensor is suitable for a large number of applications, such as the measurement of environmental variables in weather measurement. This sensor can be connected to your device using I2C protocol or SPI. We can use the I2C protocol for simple wiring, but care must be taken to ensure that the sensors do not collide between I2C addresses. The module has its own pull-up resistor, connected between the power supply and the SDA/SCL pins by default. When using multiple sensors, we can change the I2C address by soldering a jumper on the back of the PCB. The module has a total of 4 pins: two for communication and 2 for power. The power supply can range from 1.8 to 3.6VDC, but typically it is 3.3VDC. (Adafruit.com, 2024)
*Wind speed sensor WH-SP-WS01*
The sensor named WH-SP-WS01 is designed for wind speed measurement, which works on the mechanical principle. The sensor is designed in a non-contact way using a magnetic reed contact. The sensor has a cable that is terminated with an RJ11 connector. The sensor can be used alone, but it can also be used in conjunction with a sensor for measuring wind direction, which has a separate connector for connecting this sensor. Subsequently, these sensors are routed through one telephone quad line. The wind speed is calculated based on the switching frequency of the sensor, where 1Hz equals 2.4km/h.
*Precipitation measuring sensor MS-WH-SP-RG*
The rainfall sensor works on the principle of a self-emptying rainwater collector. Therefore, there are no other sensors or devices below the sensor into which water could flow, thus harming them. Also, the sensor must not be covered by drain holes and should have sufficient space to absorb or drain water. For the most accurate sensing, the sensor should be positioned as horizontally as possible, which is what the small spirit level on the sensor is designed to do. The signal from the sensor is routed through an RJ11 connector cable.
*Sensor for measuring electrical quantities INA219*
The INA219 sensor is an integrated circuit developed by Texas Instruments to measure current and voltage in electrical circuits. This sensor is often used to monitor and control the power consumption of electrical equipment. Its main purpose is to provide an accurate and reliable way to measure input voltage and current consumption for a variety of applications, such as powering embedded systems, solar panels, batteries, motors, LED lighting, and other devices. The main purpose of this device is to provide an accurate and reliable way to measure input voltage and current consumption for a variety of applications, such as powering embedded systems, solar panels, batteries, motors, LED lighting, and other devices (Texas Instrument, 2015).

The sensor supports measurement of shunt voltages ranging from 0 to 26VDC. To power the device, we connect to the 3.3VDC input pin, and the device uses a maximum current of 1mA. The device can operate at temperatures ranging from -40°C to 125°C.

The sensor contains a total of 4 pins, 2 for power and two for communication using the I2C protocol. The sensor also includes a contact interconnect option on the back, allowing the I2C address of the device to be changed. The sensor contains a total of 4 usable addresses. The sensor also has a pull-up resistor on the back, which is connected to the basic setup.

### 2.1.2 Software components

*MQTT protocol*
The MQTT is a lightweight publish/subscribe message transfer protocol that runs on top of the TCP/IP protocol. MQTT is designed for orderly, lossless, and bidirectional connections for embedded devices, microcontrollers, and networks. The publish-subscribe model of the MQTT protocol is based on the Client-Server model, but the server on which MQTT runs is more like a broker or gateway. An MQTT client can send a message to a specific "Topic", or subscribe to messages to a specific "Topic". The MQTT Broker will then be in charge of processing the messages received and sending them to the correct device. The Broker will keep a list of Topics and process publish/subscribe requests from clients (Cui, 2017). For the implementation, we will also use a program called Mosquitto, which will serve as an MQTT broker that will allow us to receive data from the device and send it to the server. In this section, we will explain the program and the MQTT protocol. The MQTT protocol provides a lightweight method for transmitting messages that use the Publish/Subscribe model. This makes MQTT suitable for use in IoT devices such as sensors with low power consumption, mobile devices, and others such as cell phones or microcontrollers (Eclipse Foundation , 2018). First, we need to define the communication between the device and the MQTT Broker, which is done using Publish/Subscribe. The MQTT Broker then communicates with Node-RED in the same way. After processing the data in Node-RED, we will send an array to InfluxDB containing two JSON objects. In the first object, we will have all the data we want to store in each measurement in the database. In the second object, there will be auxiliary variables, called tags, that express auxiliary values, such as sensor names, the device name that provided us with the data, and others. The communication between InfluxDB and Grafana will be bidirectional so that when the display period changes, the data provided by the database will be updated.
***Data collection using Node-RED***
Node-RED is an open-source web programming tool that uses a flow-based approach to application programming. This is based on nodes, each of which performs a specific task, according to which it processes incoming data and sends it to the next node in the sequence. The user then creates the independent be-

haviour of the application by simply placing and linking nodes in a so-called data flow. In addition to being user-friendly, this approach also allows for clearer programming and greatly facilitates troubleshooting of potential problems. The Node-RED user environment is accessible via a web browser and provides a wide variety of nodes that perform basic automation tasks. Users can also create their own nodes or use community nodes. Storage of measured data

To store the measured data, we chose a time series database named InfluxDB.

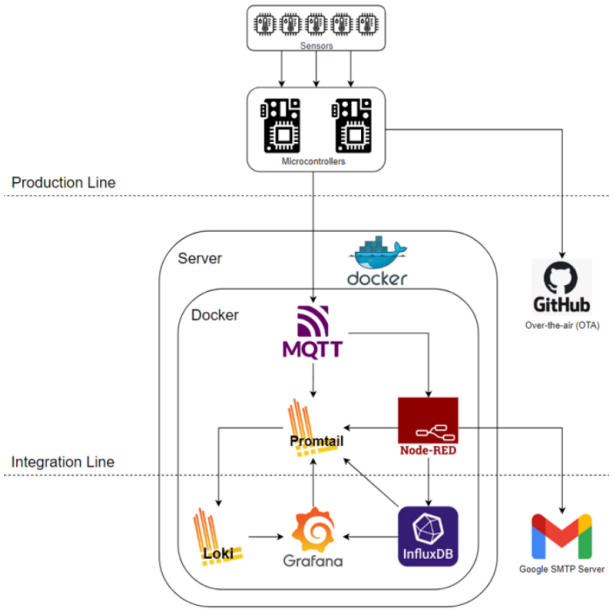



**Figure 1:** Proposed server software components

***Data storage***

InfluxDB was designed to store sensor data, calculated data, and logging logs from the application. InfluxDB is an open-source time series database written in the Go programming language and being developed by InfluxData. This database is optimized for faster, high-availability data retrieval and storage of time-dependent data in areas such as operations monitoring, IoT sensor data storage, and real-time analytics. SDBs are optimized to measure change over time. The characteristics that make time-dependent data different from other data are, e.g., data life-cycle and especially the large amount of different data (InfluxData, 2022).

To visualize the data for the user, we used the Grafana tool. Grafana is an open-source interactive data visualization platform created by Grafana Labs that allows users to visualize data (Redhat, 2024).

***Storing and managing logs***

The system logs are stored and managed by Grafana Loki and Grafana Promtail. Grafana Loki is a horizontally scalable log aggregation system built for log collection efficiency. It focuses on log streaming, storage, and indexing. Promtail is the agent responsible for collecting logs from various sources, such as system logs and Docker logs, and then sending them to Grafana Loki.

Promtail and Loki are usually deployed on each device running applications that need monitoring. Promtail and Loki are only responsible for collecting and storing logs. To use them, e.g., for visualization, we need to link Loki and Grafana together. Fig. 1 shows all software components running on the server together.

***Hardware design and wiring***

Fig. 4 shows the device installed. In total, there are five sensors that collect seven different quantities. The housing in the picture is then fitted over the BME280 and FHT22 sensors to protect them from the weather. The solar panel is mounted on a heavy concrete block, which protects it from blowing away and avoids the need to mount the panel directly on the roof. Up to this stage, the visualisation in Grafana has only been simple serial graphs. Once the latest version of the data collection device was installed, the creation of the main dashboard began, as described in the server-side implementation.

In the final version, the device's behaviour was changed when an error condition occurred or when sensors were not connected. The device's connection to Wi-Fi was redesigned, resulting in longer battery life, which in turn allowed more frequent data collection from every 15 minutes to 5 minutes. This was followed by incorporating rainfall amount collection from an approximate value to continuous rainfall amount data collection, even while the device was asleep.

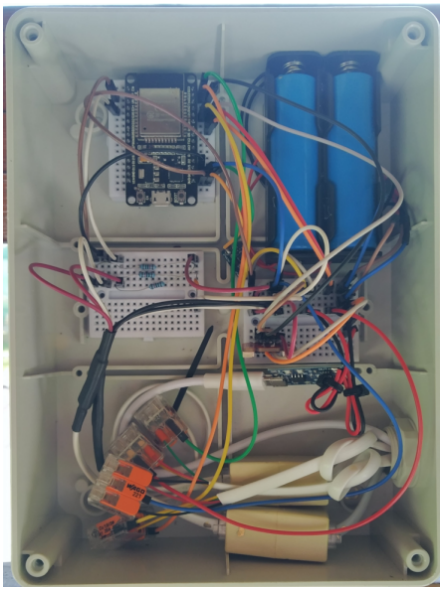


**Figure 2:** Hardware components housed in a box on a meteorological station.

Also, device error conditions, such as overflow of variables or division by zero, etc., were treated in the latest version. After fixing the flaws from the last version, the device was moved to a real environment where it collects weather data. The device is mounted on the roof along with the solar panel, and the box with

the microcontroller was placed under the roof.

# 3 Results

The weather data collection device has gone through several development versions to reach the final version that is currently implemented in the real environment. This final version includes several improvements and modifications based on experience and testing of previous versions.

The final device is equipped with several types of sensors to measure different meteorological variables. These sensors include the BME280 to measure temperature, humidity, and air pressure, the DHT22 to measure temperature and humidity, an anemometer and wind rosette to measure wind speed and direction, and a rain gauge to measure the amount of precipitation. These sensors are strategically placed to accurately and reliably collect data from the surrounding area.

Power for the device is provided by two 2500mAh 18650 batteries, which are wired in parallel to increase the overall capacity. The device is also equipped with a solar panel that provides battery charging, ensuring continuous operation of the device even in the absence of other power sources. The batteries are connected to a battery charger that regulates the charge from the solar panel to prevent overcharging or damage.

The device is housed in a SCABOX (I68) shown in Fig. 2 , which provides protection from external influences such as rain, dust, and extreme temperatures. Prior to installation, a hole was drilled into the box, and a grommet for the cabling was subsequently inserted into the box to ensure that the cables were protected from damage whilst allowing easy access for maintenance.

**Figure 3:** Final dashboard.

# 4 Conclusion

The entire equipment programme consists of four parts. The first part is the auth.py file, which contains the data for connecting to Wi-Fi and the MQTT Broker. The second part is the version.json file where the current firmware version of the device is stored. The third part contains all the libraries needed for the device to function. The fourth part contains the boot.py and main.py files, where the main code of the device is located. The

**Figure 4:** Meteo station

boot file connects to Wi-Fi and checks and updates the firmware. The main file initializes pins and variables, collects sensor data, processes it, and sends it to the MQTT Broker. Once connected to the MQTT Broker, the device sends the collected data, disconnects, and sleeps for 5 minutes. If an error occurs, such as an unconnected sensor or a failed connection to the MQTT Broker, the device also falls asleep for 5 minutes and then repeats the entire process.

Once all the improvements were implemented, the final version of the device was placed on the roof along with the solar panel, with the microcontroller box placed under the roof for protection from the weather. This placement allows optimal meteorological data collection, which is then processed, stored, and visualized on the server. The collected data include variables such as temperature, humidity, air pressure, wind speed, and direction, amount of precipitation, battery status, and the status of equipment and sensors. This ensures continuous and reliable collection and processing of meteorological data in real-time.

The latest version of the weather station is conceived as an IoT device that is powered by a solar panel and batteries that send data to a server every 5 minutes using the MQTT protocol. When collecting rainfall data, a time window of 5 minutes would represent a large inaccuracy, which is why this part was subsequently redesigned to allow for continuous rain data collection, even with the device in a sleep state and not measuring ambient variables. Furthermore, an OTA updates approach was used for the weather station, which allows a new firmware version to be uploaded to the weather station via Wi-Fi without having to access it directly.

In this study, we designed and built a specialized IoT-enabled weather station to advance monitoring environmental conditions within urban areas. We carefully selected and proposed hardware and software components that align with the IoT and Smart City

concepts. The weather station is equipped to gather, process, and visualize meteorological data effectively. It operates with a high degree of reliability, transmitting data every five minutes using MQTT protocol and maintaining functionality even during low-power conditions to ensure continuous rainfall data collection. Additionally, we implemented over-the-air firmware updates, enhancing the station's adaptability without needing physical adjustments.

Our proposal shows the integration of cutting-edge IoT technologies, utilizing MQTT, Node-RED, InfluxDB, and Grafana to enhance the functionality and efficiency of the weather station, thereby facilitating sophisticated data analysis and real-time environmental monitoring in urban settings.

The final deployment of the IoT meteorological station underscores a significant advancement in Smart City environmental monitoring. The system's ability to consistently transmit accurate data every five minutes via the MQTT protocol and its novel capability to collect continuous rainfall data even in low-power states confirm its operational efficacy and reliability. With over-the-air firmware updates facilitating effortless upgrades, the station remains at the forefront of technological innovation. This project proves the feasibility and critical importance of IoT systems in enhancing urban resilience and sustainability.

## Acknowledgments

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00562.

## References

Adafruit.com (2024). Adafruit BME280 Humidity + Barometric Pressure + Temperature Sensor Breakout. Available online: https://learn.adafruit.com/adafruit-bme280-humidity-barometric-pressure-temperature-sensor-breakout/overview.

Bauer, M., Sanchez, L., and Song, J. (2021). IoT-Enabled Smart Cities: Evolution and Outlook. *Sensors*, 21(13):4511. Number: 13 Publisher: Multidisciplinary Digital Publishing Institute.

Bibri, S. E. and Krogstie, J. (2020). The emerging data–driven Smart City and its innovative applied solutions for sustainability: the cases of London and Barcelona. *Energy Informatics*, 3(1):5.

Camarillo-Escobedo, R., Flores, J. L., Marin-Montoya, P., García-Torales, G., and Camarillo-Escobedo, J. M. (2022). Smart multi-sensor system for remote air quality monitoring using unmanned aerial vehicle and lorawan. *Sensors*, 22(5).

Cheon, M. and Mun, C. (2023). The Climate of Innovation: AI's Growing Influence in Weather Prediction Patents and Its Future Prospects. *Sustainability*, 15(24):16681. Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

Cui, P. (2017). Comparison Of IoT Application Layer Protocols. Master's thesis, Auburn University, United States – Alabama. ISBN: 9798368493824.

Eclipse Foundation (2018). Eclipse Mosquitto. Available online: https://mosquitto.org/.

European Commission, u. (2019). Majority commuted less than 30 minutes in 2019. Available online: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20201021-2.

InfluxData (2022). Time Series Database (TSDB) Guide | InfluxDB. Available online: https://www.influxdata.com/time-series-database/.

Lom, M., Pribyl, O., and Svitek, M. (2016). Industry 4.0 as a part of smart cities. In *2016 Smart Cities Symposium Prague (SCSP)*, pages 1–6.

Ministry of Investments (2023). Slovakia's digital transformation strategy 2030 and digital action plan. Available online: https://www.smartcity.gov.sk/wp-content/uploads/2023/05/Ak%C4%8Dn%C3%BD-pl%C3%A1n-inteligentn%C3%BDch-miest-a-regi%C3%B3nov-na-roky-2023-2026.pdf.

Neo, E. X., Hasikin, K., Lai, K. W., Mokhtar, M. I., Azizan, M. M., Hizaddin, H. F., Razak, S. A., and Yanto (2023). Artificial intelligence-assisted air quality monitoring for smart city management. *PeerJ Computer Science*, 9:e1306. Publisher: PeerJ Inc.

RaspberryPi Ltd. (2024). Raspberry Pi 4 Model B specifications. Available online: https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/.

Redhat (2024). What is Grafana? Available online: https://www.redhat.com/en/topics/data-services/what-is-grafana.

RTD (2022). Getting Started with the ESP32 Development Board | Random Nerd Tutorials. Available online: https://randomnerdtutorials.com/getting-started-with-esp32/.

Sparkfun (2024). RHT03 (DHT22) Humidity and Temperature Sensor Hookup Guide - SparkFun Learn. Available online: https://learn.sparkfun.com/tutorials/rht03-dht22-humidity-and-temperature-sensor-hookup-guide/all.

Tahirkheli, A. I., Shiraz, M., Hayat, B., Idrees, M. S., Sajid, A., Ullah, R., Ayub, N., and Kim, K.-I.

(2021). A survey on modern cloud computing security over smart city networks: Threats, vulnerabilities, consequences, countermeasures, and challenges. *Electronics*.

Texas Instrument (2015). Ina219 zero drift, bidirectional current power monitor with. Available online: https://www.ti.com/lit/ds/symlink/ina219.pdf.

# Virtual Commissioning of Industrial Robot through Software-in-the-Loop Strategies for Cyber-Physical Systems

**Roman Ružarovský, Richard Skýpala, Nikolas Blahušiak, Ján Šido, Martin Csekei**
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava/ Institute of Production Technologies
Jana Bottu 25, 917 24 Trnava, Slovakia
{roman.ruzarovsky, Richard.skypala,xblahusiak, jan.sido, martin.csekei}@stuba.sk

**Tibor Horák, Peter Střelec**
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava/ Institute of Applied Informatics, Automation and Mechatronics
Jana Bottu 25, 917 24 Trnava, Slovakia
{tibor.horak, peter.strelec}@stuba.sk

***Abstract****. This paper considers the virtual commissioning of an industrial robot. An industrial robot with peripherals is formed as a Cyber-Physical System. Virtual commissioning uses the Software-in-the-Loop method, where a co-simulation interfaces the emulation of a virtual controller from ABB RobotStudio with the digital twin of the industrial robot application in Tecnomatix Process Simulate through the VRC server. The objective was to provide a functional interconnection between the simulation model of a robot workstation and the control software for the industrial robot with peripherals. The result of this interfacing is to replace the RCS system and test the functionality of the proposed VRC system on a case study.*

**Keywords.** Virtual Commissioning, Industrial Robot, Software-in-the-Loop, Co-Simulation

## 1 Introduction

The digitalisation and virtualisation of industrial processes have made significant advances in recent years, transforming traditional production methods and processes. (Gašpar et al., 2023; Sharma et al., 2023) Digitisation enables the conversion of physical processes into digital form. Virtual models and simulations provide a detailed view of the functionality and behaviour of systems before they are deployed, allowing us to identify and correct potential problems in the early stages of development.

Virtual Commissioning (Pullnig, 2023) is the key tool in this process. It uses advanced software tools to create faithful digital twins of industrial equipment and systems that can be tested and optimised in detail in a simulated environment. This approach reduces the risks associated with physical installation and commissioning and significantly reduces the time and cost required to implement new technologies.

Industrial enterprises must integrate cyber-physical systems (CPS) that combine physical and digital components into a single whole as part of their digital transformation. (Jazdi, 2014) Such systems enable comprehensive real-time analysis and optimisation of manufacturing processes, thereby improving their performance and reliability. Virtual commissioning using Software-in-the-Loop (SiL) strategies is essential in this context. It enables realistic testing and debugging of systems, leading to smoother integration and higher quality of the resulting solutions.

Software in the Loop (SiL) is one of several methods used for virtual commissioning. SiL is the most effective method for testing and validating control software in a simulated environment where the software interacts with virtual models of equipment. (Ayed et al., 2017)It's particularly advantageous when physical components are not available or when their use is costly or impractical. SiL enables engineers to detect and fix software bugs before deployment in real operations, significantly reducing implementation costs and risks.

In addition to SIL, there are other methods such as Hardware-in-the-Loop (HiL) and Model-in-the-Loop (MiL). (Shylla et al., 2023) HiL integrates physical components into a simulated environment, allowing real devices to be tested under the control of simulated software. MIL, in turn, allows software testing using models of mathematical and physical systems. Each method has its own specific advantages and is suitable for different stages of development and testing.

SiL is particularly suitable for projects where detailed and repeated testing is required without the

risk of damaging expensive physical equipment. It provides flexibility and the ability to iterate quickly, which is crucial for developing complex systems in a dynamically changing industrial environment.

There are two main ways of programming industrial robots. (Kelemen et al., 2018; Vagas & Romancik, 2023) One is online programming, which is ideal for small projects in SMEs where it is easy to manipulate products without a large number of peripherals, or in the case of robotic devices for arc welding, where workflows and products are often changed flexibly. In this case, the programmer connects to the real controller of the industrial robot, configures the communication with the peripherals and directly creates the program. (Kohrt et al., 2013; Marcinko et al., 2024)

The second way is the offline programming method (OLP, OLRP). OLP is the general term for offline programming, which is used for industrial robots. (Bilancia et al., 2023; Vocetka et al., 2020)OLRP is a method for programming industrial robots through software that includes a virtual controller. Offline Robot Programming (OLRP) is often used by integrators and manufacturers of whole robot solutions. This solution is an introduction to Virtual Commissioning. It allows the simulator to program, deploy and reprogram its solution several times without the use of a physical device. The program can be debugged on a virtual model. By programming robot functions in a virtual, offline environment (on a computer), OLRP eliminates the need to take robots out of production. This significantly increases productivity and the bottom line. The industrial market has a number of software solutions. You have two options: an OEM solution from the robot manufacturer (ABB RobotStudio, Fanuc RoboGuide, KUKAsim, Motosim, etc.) or an agnostic solution (RoboDK, Tecnomatix Process Simulate, Delmia, Robomaster, Visual Components, etc.).

# 2 Related Work and Background

The issue of virtual commissioning has been addressed by a number of authors. These contributions focus on several areas, on the creation of simulation models (J. Wang et al., 2023), behavioural models (Strahilov & Damrath, 2015; Süß et al., 2016), or for case studies (Ugarte et al., 2022). Most of the literature focuses on the PLC systems themselves, which is understandable given that these are the main focus of research in the field. However, if we are talking about IRC systems and virtual commissioning, the situation is different. The authors (Ben Ayed et al., 2017; Shylla et al., 2023) dealt with this issue, but used Open Source solutions such as ROS and Gazebo. Authors (Wallner et al., 2023) were interested in the concept of a robotic cell, but primarily with signals for the machine and not for the robot.

Before we deploy virtual commissioning for industrial robots, we must define and characterise the relevant terminologies and technologies.

- RC/ IRC: Robot Controller/ Industrial Robot Controller. Robotic controllers are an essential part of industrial robots. The robot's controller is a computer system that connects to the robot to control its movements and movements of the robot's arm. The controller also controls the end-effector and prevents interference from occurring within the robots work area. All industrial robot must be paired with a controller in order to operate. (Khatib & Siciliano, 2008)
- PLC: PLC stands for Programmable Logic Controller. They are the gold standard for controlling automated systems in industries. They are the pinnacle of control systems, and they are now replacing hard-wired logic relays at a large scale. (Bolton, 2006)
- Co-simulation: Co-simulation is an emerging enabling technique, where global simulation of a coupled system can be achieved by composing the simulations of its parts. (Gomes et al., 2018; Li et al., 2014)
- Emulation: Emulation is the use of an application, program, or device to imitate the behaviour of a different program or device. We use a PLCSim for the emluation of the real PLC or Virtual Controller ABB for industrial robot e.g. (McGregor, 2002; Mykoniatis & Harris, 2021)
- Behavioral model: Behavioral Model is specially designed to make us understand behavior and factors that influence behavior of a System. In our case, it is the components of a manufacturing system and a mathematical description of the behaviour of these components. (Stegmaier et al., 2023)
- RRS/ RCS: Realistic Robot/ Controller Simulation. This interface enables a standardized integration of motion software for robot controllers with simulators. Hence, the motion software of any robot manufacturer can be integrated into any simulation system. Today the RCS-Interface is the world-wide de-facto standard for precise simulation of robot motion behavior. (ROZITEK, 2023)
- VRC: Virtual Robot Controller. The VRC-Interface Specification aims at integrating robot controller software almost completely into simulators. Furthermore it supports the over-all data flow during the development and maintenance of installations with robots and automates several technical integration aspects. (Fraunhofer Institute for Production Systems and Design Technology, 2024)

When performing virtual commissioning with ABB robots, a connection between Process Simulate and RobotStudio can be created and used to simulate a

complete virtual robot control (VRC) system. Process Simulate is designed to read and write information from and to RobotStudio. This includes joint values, signal values and other relevant data. This allows the simulation to visualise and interact with the VRC.

### 2.1 Defining the problem

The integration with the robot will be very tight thanks to the OEM solution. The advantage of this solution is that we use directly the virtual controller from the specific robot we want to use, including it's functionalities. However, these software are limited to direct programming of the robot and do not have other functionalities like large agnostic solutions. When integrating robots into production workplaces, it is necessary to maintain a precise cycle time. In this case, we need to have built-in RCS (Realistic Controller Simulation) modules to provide accurate cycle times. 100% accurate cycle times are important for certain industries such as automotive. Another concern is safety, collisions, dealing with singularities, etc.

The implementation process (calibrating the virtual world with the real world) is an extremely important part of the solution. OLRP is of little value if it does not guide the robot to exactly the right coordinates in the real world. The OLRP is a program generated for the robot via a postprocessor that contains the basic trajectories and motions of the robot, and often times we know from experience that the online roboticist will fine-tune the program on the jobsite or via an OEM robotics solution so that the robot interacts with the peripherals.

### 2.2 The aim of the work

The objective is to establish a seamless link between the robotic workstation simulator and the industrial robot controller, integrating inputs and outputs. This will leverage the benefits of both solutions. The outcome of this integration will be the replacement of the RCS system and the testing of the proposed VRC system in a real-world case study.

## 3 Materials and Methods

We will use two methods to link and verify functionality. The first method is the co-simulation method (Hasmukhbhai et al., 2021), which will create a joint simulation of the robotic workstation on two simulators. The first one will provide the simulation of the robotic workstation in the Tecnomatix Process Simulate environment. It is necessary to create a full model of the workplace with the industrial robot and its peripherals and to prepare the model for Virtual Commissioning. This means that it is necessary to create the necessary signals for the control of the industrial robot and the input-output commutation with the pefiferia such as the robotic tool. On the other hand, it is necessary to create a simulation model for the ABB RobotStudio industrial robot with the necessary specification and the created virtual controller. The time synchronization in the co-simulation between the simulators will be taken care of by the co-simulation algorithm itself. (Sadjina et al., 2019; L. Wang et al., 2023) The interactions between these subsimulators are synchronized only at discrete communication points. The sub-simulators can run in parallel, which can speed up simulations of the whole system. Models can be efficiently developed in parallel and can be easily reused. Ability to include hardware, software, and human in the loop relatively easily and consistently. Due to its features and benefits, co-simulation is best used for complex and heterogeneous cyber-physical systems.

## 4 Development of the Virtual Commissioning Model

The model workplace will be developed in four basic steps. The first step is to create a model robotic workstation in the Tecnomatix Process Simulate environment. The second step is to create behavioural models for the individual peripherals and the industrial robot. We then distribute the individual control and command signals created for the individual connections to the virtual controllers, using the Software-In-the-Loop (SiL) method in Virtual Commissioning. The last step is to connect the virtual controller from the ABB RobotStudio environment via the VRC server and test the co-simulation.

### 4.1 Simulation model of a robotic workplace

To operate an industrial robot and its workplace and create a control programme for the robot and its controller using the OLRP method, you have to model the workplace. This is done using CAD, where you create fixture models, end-effector models and other components if they are not already in the database.

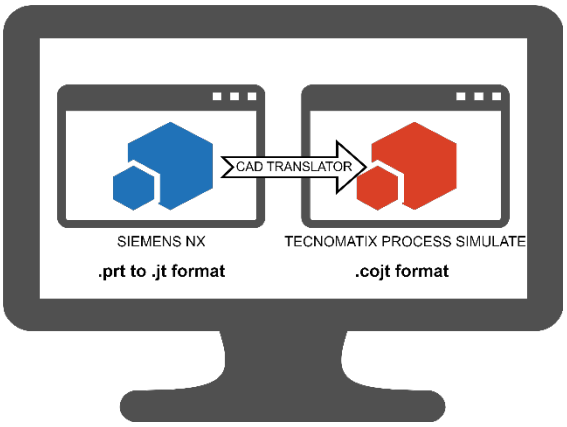



**Figure 1.** General procedure for the creation of models for simulation

The manufacturer of the single-label markers provides models of industrial robots. These models are ready for simulation in both types of software, whether OEM or agnostic large solution. If we have the models available, we need to translate them via CAD translator into the necessary format for the simulation. In our case it is the *.jt format, Fig. 1.

The entire simulation model of the workplace is designed to verify the collision-free nature of the workplace, the reaches of industrial robots, and to generate robot trajectories and motions for a specific manufacturing application. The simulation model works with a virtual controller and MOP (Motion Planner) that can control robot motion including individual axis, speed and zone configurations. The entire robotic program can be translated through the postprocessor, where it is necessary to have a virtual robot controller installed for ABB robots. This can be characterized as offline programming of industrial robots, Fig. 2.

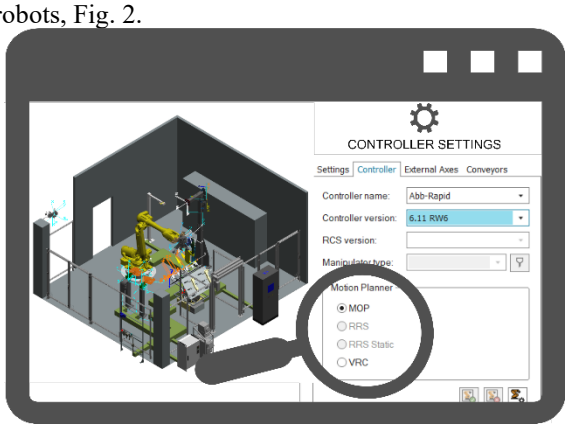



**Figure 2.** MOP controller setup for the robot

## 4.2 Behavioural model for VC

The entire simulation model can be upgraded to a higher level and prepared for virtual commissioning of the system. Virtual commissioning is the process of using virtual simulation technology to commission - design, install or test control software with a virtual model of a robotic workstation before connecting it to the real system. The created workstation can be used to simulate real robot control code and program in combination with emulated control hardware (SiL) or in combination with real hardware (HiL) in a realistic virtual commissioning environment to verify complete system function before production starts.

The kinematic models created for the implementation of the robotic workplace simulation and offline programming of the industrial robot are the basis for the implementation of the behavioral models for Virtual Commissioning. Virtual commissioning of a robotic system brings together three important parts: the digital model (sometimes referred to as the digital twin), the controller code that controls motion and responds to sensor feedback, and the development environment that enables them to run together. The digital twin needs a behavioural model, according to our research. Models that represent the temporal and logical behavior of a production device in relation to the corresponding control system are referred to as behavioral models. (Skýpala & Ružarovský, 2022; Striffler & Voigt, 2023) Behavioural models are modelled in a separate simulation and modelling tool, which is linked to the control at the input and output level. From a management perspective, behavioural models represent a 'black box'. The behaviour models also serve as a link between the control system and the extended geometry/kinematics mode. (Kiefer et al., 2009) In our case, we created a behavioral model in the Tecnomatix Process Simulate LB environment, where we defined Entries, Exits, Parameters and Constants for the individual kinematic models, Fig. 3.

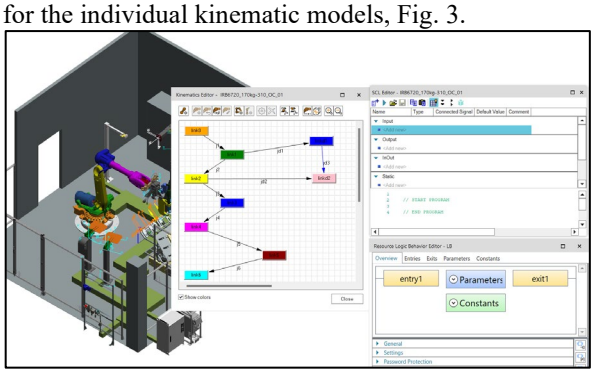



**Figure 3.** Kinematic model of robot and creation of behavioural model

Based on the nature of the motion and behavior of the model, we generated inputs and outputs for model control, individual variables and data types. The same had to be done for the robot control except for its peripherals. As can be seen, we thus generated 4 inputs and 5 outputs, Fig. 4.

| InterfaceName | RobotInternalName | I_Q | TYPE | Comment |
|---|---|---|---|---|
| HX165_hub40_startProgram | startProgram | Q | BOOL | (Starting Program) |
| HX165_hub40_programNumber | programNumber | Q | BYTE | (Program Number) |
| HX165_hub40_emergencyStop | emergencyStop | Q | BOOL | (Program Emergency Stop) |
| HX165_hub40_programPause | programPause | Q | BOOL | (Program Pause) |
| HX165_hub40_reduceSpeed | reduceSpeed | Q | BYTE | (Reducing Speed) |
| HX165_hub40_programEnded | programEnded | I | BOOL | (Ending Program) |
| HX165_hub40_mirrorProgramNumber | mirrorProgramNumber | I | BYTE | (Mirror Program Number) |
| HX165_hub40_errorProgramNumber | errorProgramNumber | I | BOOL | (Error Program Number) |
| HX165_hub40_robotReady | robotReady | I | BOOL | (Robot Ready) |

**Figure 4.** Created signals for robot control

The behaviour model is thus ready for Virtual Commissioning. In our case, we have currently prepared the model only for communication between the virtual controller emulator and the robotic workstation simulation model, in order to verify the functionality of co-simulation and to use the proposed VRC system instead of the MOP or RCS system. The latter needs to be installed and run. If the robot name in Process Simulate and the controller name in RobotStudio do not match, a new string parameter called OLP_VRC_NAME must be created in Process Simulate on the corresponding robot instance and set to the controller name in RobotStudio. When Process Simulate attempts to find a matching VRC instance for the robot, it will attempt to do so using the value of the parameter in addition to the robot name. In this case, either the <RobotName> in Process Simulate or the <value OLP_VRC_NAME> must match the name of the VRC instance.

### 4.3 Virtual controller emulation

As mentioned, the simulation model must be prepared to communicate with the VRC instance. A new folder RobotParameterEditor was created in the Abb-Rapid folder at \eMPower\Robotics\OLP\.

```
<RobotController Name="Abb-Rapid">
   <RoboticParams>
      <Param Name="OLP_VRC_NAME" ValueType="String" />
   </RoboticParams>
</RobotController>
```

**Figure 5.** Created a folder for communication

At the same time set other parameters for communication. Communication is done on localhost via communication port 8523 and the service runs on Tecnomatix VRC Server ABB Real Time. Each robot instance in Process Simulate (and therefore each VRC instance) requires a unique port. An emulation of the virtual controller is then created in ABB RobotStudio. In RobotStudio, the program or module must exist as a global procedure nested under the robot's TASK1. The operation can be completely empty (no OLP commands or locations) or it can have nested information underneath, but the robot only executes the code stored in RobotStudio. We can use the Process Simulate Upload Programs and Download to Robot commands to exchange programs between Process Simulate and RobotStudio.

### 4.4 Testing the functionality of co-simulation

Tecnomatix Virtual Robot Controller Server ABB Real Time provides communication between Process Simulate and RobotStudio VRC. The server allows you to execute robot programs using the ABB VRC, which is part of RobotStudio, while connecting to Process Simulate for simulation. For proper functioning, several parameters need to be set for the server to run on localhost, the parameters need to be set on the simulation side of the VRC and the same on the ABB virtual emulator, and an important point is to have the same name for both controllers on both sides.

## 5 Results

The communication between the two controllers was successful.

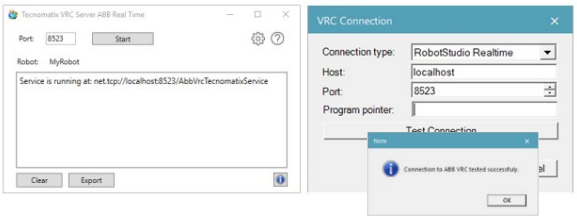



**Figure 6.** Successful communication via the configured VRC server

We sent several data between the two controllers in two software, the robot moved according to the instructions we entered through the RAPID control program. (ABB Robotics, 2004) We monitored the response, but it was not measured, we could monitor it currently through the ABB Real Time application. The delay was felt just when the 3D simulation model was complex and the problem was on the hardware performance side.

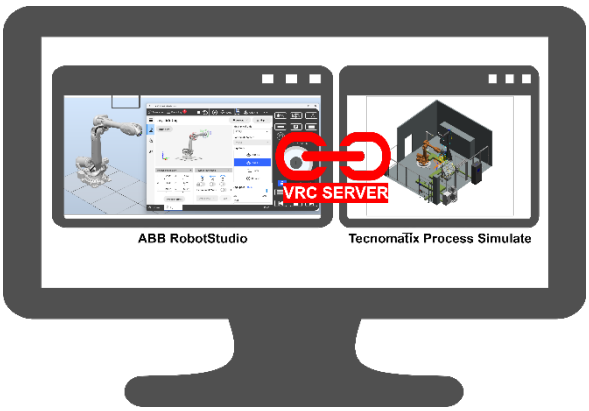



**Figure 7.** Communication between the emulated virtual controller and the simulated model via VRC

Overall, the testing was successful, but future testing needs to be done. There is no guarantee that all the different parts of the joint simulation will work well together. So far there is no general and easy to use co-simulation solution (software or algorithm). This is mainly due to technical problems and the fact that it is very difficult to define a universal method that works for all kinds of situations and requirements.

Co-simulation in most cases requires sufficient knowledge at the system level and potentially also knowledge of the individual models. It is not easy to determine where and how to partition the system into sub-models, leading to a trade-off between modularity, accuracy and performance. This has far-reaching implications for sub-model development, collaboration and ease of use.

It is also important to consider how to define links between sub-models. In addition, it is necessary to define how computational causality between sub-models is defined.There is a wide range of co-simulation algorithms to choose from, ranging from very simple (which may be slow and inaccurate) to quite complex (which may be difficult to tune and may have other requirements). The simplest co-simulation algorithms support only constant communication intervals, so the user must choose the correct interval for a given simulation.

## 5 Conclusion and Future Work

Our project involved co-simulation between a robot simulation model and a virtual controller emulator. Testing was carried out using the Virtual

Commissioning and Software-in-the-Loop methods. Virtual Commissioning is the best way to create a digital twin of the system, thus enabling testing and verification of functionality prior to its physical deployment. This involves the creation of detailed models of the behaviour of individual components and their interactions, thus reducing the risk of errors and increasing reliability.

We created behavioural models for this project to simulate the robot's motion in detail, including its movements, speed, acceleration and interaction with the environment. These models included mathematical representations of the control algorithms that govern the robot's movements. For example, the simulations included controlling the robot's arm motion, its interaction with objects, and its responses to various sensory inputs.

The virtual controller emulator has been designed to faithfully mimic the logic and responses of a real controller. This included the processing of control commands, synchronization with the robot simulation model, and consideration of time delays that may arise in real control.

Testing using the Software-in-the-Loop method allowed us to simulate the control algorithms in a software environment before implementing them in a physical controller. This approach provided the ability to thoroughly verify and optimize the control algorithms in a controlled virtual environment, minimizing the risk of errors when deployed in a real system. Preliminary testing results indicate that the implementation is functional, however, more detailed testing will need to be performed in the future for full confirmation. It is also necessary to validate the communication at the signal level through a PLC emulation such as PLCSim Advanced. This would confirm the relevance of deploying the Virtual Commissioning method for industrial robots and their peripherals as cyber-physical systems.

## Acknowledgments

This research was funded by the Scientific Grant Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic and the Slovak Academy of Sciences, grant number VEGA 1/0391/24 Research and development of sensor calibration methodology for diagnostic equipment for automated and robotic systems.

## References

Ayed, M. Ben, Zouari, L., & Abid, M. (2017). Software In the Loop Simulation for Robot Manipulators. *Engineering, Technology & Applied Science Research*, *7*(5), 2017–2021. https://doi.org/10.48084/ETASR.1285

Ben Ayed, M., Zouari, L., & Abid, M. (2017). Software In the Loop Simulation for Robot Manipulators. *Engineering, Technology & Applied Science Research*, *7*(5), 2017–2021. https://doi.org/10.48084/ETASR.1285

Bilancia, P., Schmidt, J., Raffaeli, R., Peruzzini, M., & Pellicciari, M. (2023). An Overview of Industrial Robots Control and Programming Approaches. *Applied Sciences 2023, Vol. 13, Page 2582*, *13*(4), 2582. https://doi.org/10.3390/APP13042582

Bolton, W. (2006). Programmable Logic Controllers. *Programmable Logic Controllers, Fourth Edition*, 1–292. https://doi.org/10.1016/B978-0-7506-8112-4.X5018-9

Fraunhofer Institute for Production Systems and Design Technology. (2024). *Realistic Robot Simulation (RRS) - Fraunhofer IPK*. https://www.ipk.fraunhofer.de/en/expertise-and-technologies/automation/realistic-robot-simulation-rrs.html

Gašpar, G., Dudak, J., Sladek, I., Zuziakova, I., & Meravy, T. (2023). Utilization of Open-Source in Upgrading Technological Devices for Industry 4.0 readiness. *Central European Conference on Information and Intelligent Systems; Varazdin*. https://www.proquest.com/docview/2876106571?pq-origsite=gscholar&fromopenview=true&sourcetype=Conference%20Papers%20&%20Proceedings

Gomes, C., Thule, C., Broman, D., Larsen, P. G., & Vangheluwe, H. (2018). Co-Simulation. *ACM Computing Surveys (CSUR)*, *51*(3). https://doi.org/10.1145/3179993

Hasmukhbhai, S., Le Henaff, P., Schiffer, C., Krammer, M., & Benedikt, M. (2021). View of Accurate Robot Simulation for Industrial Manufacturing Processes using FMI and DCP Standards. *Proceedings of the 14thInternational Modelica ConferenceSeptember 20-24, 2021*. https://www.ecp.ep.liu.se/index.php/modelica/article/view/247/207

Jazdi, N. (2014). Cyber physical systems in the context of Industry 4.0. *Proceedings of 2014 IEEE International Conference on Automation, Quality and Testing, Robotics, AQTR 2014*. https://doi.org/10.1109/AQTR.2014.6857843

Kelemen, M., Virgala, I., Lipták, T., Miková, L., Filakovský, F., & Bulej, V. (2018). A Novel Approach for a Inverse Kinematics Solution of a Redundant Manipulator. *Applied Sciences 2018, Vol. 8, Page 2229*, *8*(11), 2229. https://doi.org/10.3390/APP8112229

Khatib, Oussama., & Siciliano, Bruno. (2008). *Springer Handbook of Robotics*.

Kiefer, J., Ollinger, L., & Bergert, M. (2009). Virtuelle Inbetriebnahme - Standardisierte Verhaltensmodellierung mechatronischer Betriebsmittel im automobilen Karosserierohbau.

*Atp Magazin*, *51*(07), 40–46. https://doi.org/10.17560/ATP.V51I07.2000

Kohrt, C., Stamp, R., Pipe, A. G., Kiely, J., & Schiedermeier, G. (2013). An online robot trajectory planning and programming support system for industrial use. *Robotics and Computer-Integrated Manufacturing*, *29*(1), 71–79. https://doi.org/10.1016/J.RCIM.2012.07.010

Li, W., Zhang, X., & Li, H. (2014). Co-simulation platforms for co-design of networked control systems: An overview. *Control Engineering Practice*, *23*(1), 44–56. https://doi.org/10.1016/J.CONENGPRAC.2013.10.010

Marcinko, P., Semjon, J., Jánoš, R., Svetlík, J., Sukop, M., & Ondočko, Š. (2024). Analysis of the Methodology for Experimental Measuring of the Performance Criteria of the Laser-Using Collaborative Robot's Path Accuracy. *Applied Sciences 2024, Vol. 14, Page 1414*, *14*(4), 1414. https://doi.org/10.3390/APP14041414

McGregor, I. (2002). The relationship between simulation and emulation. *Winter Simulation Conference Proceedings*, *2*, 1683–1688. https://doi.org/10.1109/WSC.2002.1166451

Mykoniatis, K., & Harris, G. A. (2021). A digital twin emulator of a modular production system using a data-driven hybrid modeling and simulation approach. *Journal of Intelligent Manufacturing*, *32*(7), 1899–1911. https://doi.org/10.1007/S10845-020-01724-5/TABLES/3

Pullnig, M. (2023). *Virtual Comissioning: Process Optimization*. https://www.3dfindit.com/en/engiclopedia/virtual-commissioning

ROZITEK. (2023). *The ultimate guide to Robot Control System (RCS) - Rozitek*. https://rozitek.com/the-ultimate-guide-to-robot-control-system-rcs/

Sadjina, S., Kyllingstad, L. T., Rindarøy, M., Skjong, S., Esøy, V., & Pedersen, E. (2019). Distributed co-simulation of maritime systems and operations. *Journal of Offshore Mechanics and Arctic Engineering*, *141*(1). https://doi.org/10.1115/1.4040473/366241

Sharma, M., Raut, R. D., Sehrawat, R., & Ishizaka, A. (2023). Digitalisation of manufacturing operations: The influential role of organisational, social, environmental, and technological impediments. *Expert Systems with Applications*, *211*, 118501. https://doi.org/10.1016/J.ESWA.2022.118501

Shylla, D., Jain, A., Shah, P., & Sekhar, R. (2023). Model in Loop (MIL), Software in Loop (SIL) and Hardware in Loop (HIL) Testing in MBD. *2023 4th IEEE Global Conference for Advancement in Technology, GCAT 2023*. https://doi.org/10.1109/GCAT59970.2023.10353323

Skýpala, R., & Ružarovský, R. (2022). The Role of a Behavioural Model for the Virtual Commissioning of Robotic Manufacturing Systems. *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, *30*(50), 45–52. https://doi.org/10.2478/RPUT-2022-0006

Stegmaier, V., Eberhardt, T., Schaaf, W., Jazdi, N., & Weyrich, M. (2023). A behavior model for Digital Twins of vacuum suction cups. *Procedia CIRP*, *118*, 958–963. https://doi.org/10.1016/J.PROCIR.2023.06.165

Strahilov, A., & Damrath, F. (2015). Simulation of the behavior of pneumatic drives for virtual commissioning of automated assembly systems. *Robotics and Computer-Integrated Manufacturing*, *36*, 101–108. https://doi.org/10.1016/J.RCIM.2015.01.001

Striffler, N., & Voigt, T. (2023). Concepts and trends of virtual commissioning – A comprehensive review. *Journal of Manufacturing Systems*, *71*, 664–680. https://doi.org/10.1016/J.JMSY.2023.10.013

Süß, S., Hauf, D., Strahilov, A., & Diedrich, C. (2016). Standardized Classification and Interfaces of Complex Behaviour Models in Virtual Commissioning. *Procedia CIRP*, *52*, 24–29. https://doi.org/10.1016/J.PROCIR.2016.07.049

Ugarte, M., Etxeberria, L., Unamuno, G., Bellanco, J. L., & Ugalde, E. (2022). Implementation of Digital Twin-based Virtual Commissioning in Machine Tool Manufacturing. *Procedia Computer Science*, *200*, 527–536. https://doi.org/10.1016/J.PROCS.2022.01.250

Vagas, M., & Romancik, J. (2023). Calibration of an intuitive machine vision system based on an external high-performance image processing unit. *Proceedings of the 2023 24th International Conference on Process Control, PC 2023*, 186–191. https://doi.org/10.1109/PC58330.2023.10217606

Vocetka, M., Huňady, R., Hagara, M., Bobovský, Z., Kot, T., & Krys, V. (2020). Influence of the Approach Direction on the Repeatability of an Industrial Robot. *Applied Sciences 2020, Vol. 10, Page 8714*, *10*(23), 8714. https://doi.org/10.3390/APP10238714

Wallner, B., Trautner, T., & Bleicher, F. (2023). Functionality test methodology for virtual commissioning of reconfigurable manufacturing systems. *Procedia CIRP*, *118*, 56–61. https://doi.org/10.1016/J.PROCIR.2023.06.011

Wang, J., Niu, X., Gao, R. X., Huang, Z., & Xue, R. (2023). Digital twin-driven virtual commissioning of machine tool. *Robotics and Computer-Integrated Manufacturing*, *81*, 102499. https://doi.org/10.1016/J.RCIM.2022.102499

Wang, L., Wang, Z., Kendall, P., Gumma, K., Turner, A., & Ratchev, S. (2023). Deep Dynamic Layout Optimization of Photogrammetry Camera Position

Based on Digital Twin. *IEEE Transactions on Automation Science and Engineering*. https://doi.org/10.1109/TASE.2023.3323088

# The Role of Digital Twins in Future Skill Development

**Marija Kuštelega, Ahmed Shareef, Renata Mekovec**
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
{marija.kustelega, ahmed.shareef, renata.mekovec}@foi.unizg.hr

***Abstract****. A digital twin is a virtual system designed to replicate the functionality of a physical object. The concept originating from the industrial sector offers promising avenues for creating immersive, interactive, and personalized learning experiences. This paper explores the potential applications of digital twin technology in the field of education. In order to ascertain the context and methodology of the use of digital twins in education, we present the findings of our literature analysis prepared using the PRISMA approach. Furthermore, we explore the opportunities associated with integrating digital twin technology into educational environments and propose future directions for research and development.*

**Keywords.** digital twin, education, enhancement

## 1 Introduction

Digital twins (DTs) have already been used for distance learning in several fields, including healthcare, tourism, and engineering education (Hawkinson, 2022). It is emphasized that COVID 19 technology has accelerated the adoption of digital twin technology (Hazrat et al., 2023). Namely, the emergence of the coronavirus pandemic forced professors and students to use distance learning methods to ensure continuous education (Itani et al., 2022).

The rise of the virtual world, or metaverse technology, has significantly enhanced pedagogical and technical support for education, expanding learning opportunities and demonstrating numerous benefits in distance learning (Tlili et al., 2022). According to EU Business School (2022), one of the most promising applications of the metaverse in education is the "gamification" of learning. This approach turns the classroom into a virtual world and encourages students to finish their assignments.

Digital twin technology can help incorporate the ideas and concepts of the natural or physical world in metaverse-based education, making education more realistic and user-friendly (Mitra, 2023).

Some of the benefits of introducing DT into the educational context are personalized learning, better collaboration possibilities and content accessibility (Hawkinson, 2022). It has been shown that blended learning can significantly improve students' motivation, interest and learning approach (Kartashova et al., 2024).

At the same time, it is emphasized that such education requires investment in professional development of educators to increase their ability to teach students (Hazrat et al., 2023).

The statistics of DT in education show that they will continue to gain popularity due to the shift to online learning that has occurred. According to the National Center for Education Statistics (NCES), the pandemic affected public colleges to enroll approximately 7.5 million online students in 2022 (U.S. Department of Education, 2023). In 2021, as in-person learning slowly returned, distance learning dropped to 59%. Even though online learning rates decreased slightly in 2022, they were still significantly higher than in the years prior to the pandemic (Hamilton and Beagle, 2024).

Moreover, it is also clear that investments in educational technologies have significantly increased. Between 2017 and 2021, US educational technology venture funding jumped from $1 billion to $8 billion, and this growth is anticipated to continue as more universities adopt blended learning and advanced technologies (HolonIQ's Education Intelligence Unit, 2022).

The metaverse is expected to transform cities, schools, and factories, resulting in a digital twin of everything. As a result, the impact of these changes on higher education could be fatal. The Pew Research Center (2023) predicts that 50% of US colleges could close due to outdated facilities and inability to adapt to new technologies. Though some faculties actively use DT in their teaching.

For example, Quanser's QLabs software has been used by 2500 academic institutions worldwide to allow engineering students to manipulate and test 3D models in a virtual environment (Quasner.com, n.d.). Such statistics show that digital technology is slowly being applied in education.

The aim of this paper is to discover the possible applications and advantages of using DT in an educational context.

To achieve this aim, two guiding questions drive this study:

- RQ1: What are domains/applications of DT usage in education?
- RQ2: What are the possible enhancements of introducing DT in education?

The rest of the paper is structured as follows: Section 2 explains the theoretical framework of DT; Section 3 outlines utilized research methodology; Section 4 presents the results about DT applications in education; Section 5 discusses findings about possible DT enhancements in learning environment; Section 6 outlines research gaps and future directions to address the research gaps; and Section 7 concludes the work.

# 2 Theoretical Framework

## 2.1 Definition and Conceptualization of Digital Twins

DTs are virtual representations of physical objects or systems that simulate their behavior and characteristics. The concept originated from NASA's Apollo program, where identical space vehicles were built for testing purposes (Liljaniemi & Paavilainen, 2020; Andreasyan and Balyakin, 2022). DT consist of two systems: a physical entity and a virtual model containing all relevant information about the physical system (Nikolaev et al., 2018). They are used in various sectors, offering benefits such as cost reduction, predictive maintenance, and improved time-to-market (Tjahyadi et al., 2023; Flaga & Pacholczak, 2022). DT leverage technologies like IoT, machine learning, and big data to enable seamless interaction between physical and digital domains, providing detailed insights for optimization and innovation (Andreasyan and Balyakin, 2022). In education, DT can include human factors, such as students and teachers, requiring specific considerations during development (Terkaj et al., 2024).

## 2.2 Key Components and Characteristics

DTs in education involve creating digital replicas of educational entities like students, teachers, and institutions, integrating human factors. Key components include generating digital copies of educational entities, data collection, and real-time data transmission (Liljaniemi & Paavilainen, 2020; Chamorro-Atalaya et al., 2024). Characteristics encompass the synchronization between physical and digital systems, visualization of system information, and integration of technologies like IoT and AI (Andreasyan and Balyakin, 2022). DTs enable simulations, virtual environments for testing, and data-driven applications, enhancing learning experiences by providing hands-on applications of ICT skills and exposure to Industry 4.0 technologies (Geuer et al., 2023; Acker et al., 2023). They offer opportunities for virtual experiments, preparation for real-world scenarios, and continuous data flow between physical objects and DT (Raudmäe et al., 2023; Orsolits et al., 2022), fostering a deeper understanding of complex concepts in a safe and interactive manner (Tjahyadi et al., 2023).

## 2.3 Types of Digital Twins

DT can be categorized into three main types: Digital Model, Digital Shadow, and Hybrid Digital Twins. The Digital Model represents a digital replica of a physical object without automated data exchange (Kim et al., 2018). On the other hand, the Digital Shadow involves an automated one-way data flow between physical and digital objects (Liljaniemi & Paavilainen, 2020). Hybrid DTs combine elements of both physical and cyber representations, allowing for bidirectional data exchange and synchronization between the physical entity and its digital counterpart (Andreasyan and Balyakin, 2022). These different types of DTs play crucial roles in various industries such as manufacturing, smart city development, and healthcare, offering benefits like predictive maintenance, improved efficiency, and enhanced decision-making capabilities (Lee et al., 2023).

# 3 Methodology

For a literature review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) technique was utilized with defined eligibility criteria (Moher et al., 2009). The PRISMA procedure was used to identify publications related to DTs in education. The search was conducted using "digital twin" in titles, abstracts, and keywords and "education" included only in the title of the paper to narrow the search. All posters, book chapters, reports and reviews are excluded from the search, also excluding all publications not written in the English language. Final search results were obtained on April 10, 2024. A total of 116 papers were found in the initial search using the Scopus and WoS databases (69 from Wos and 47 from Scopus). Following the elimination of duplicates, a brief review of the titles and abstracts was conducted. As a result, 66 articles (29 from Wos and 37 from Scopus) that were relevant to the field of DT in education were chosen. After reading the paper in its entirety, 37 publications (19 from Wos and 18 from Scopus) that provided examples of usage DT in education with defined benefits were selected for analysis. To get the most recent data on their applications, the investigation only included English language publications that were released within the last three years, from 2022 to 2024.

# 4 Applications of DT in Education

Table 1 shows the areas and applications of DT, where most of them were observed specifically in STEM education (21 articles), online education (9 articles), vocational education (2 articles) and 1 article each for special cases were found in areas such as environmental education, sustainable education, IoT education, vocal music, and physical education.

**Table 1.** Main areas and applications of DT in education

| No. | Author | Area | Application |
|---|---|---|---|
| 1. | Chamorro-Atalaya et al. (2024) | STEM education | DT in higher education for engineering training and simulation-based training. |
| 2. | Hagedorn et al. (2023) | STEM education | DT for project-based learning in engineering. |
| 3. | Acker et al. (2023) | STEM education | Low-cost DT for teaching robotics (Industry 4.0). |
| 4. | Raudmäe et al. (2023) | STEM education | DT for omnidirectional mobile robot platform. |
| 5. | Zangl et al. (2023) | STEM education | DT for measurements in robotics and AI. |
| 6. | Čech and Vosáhlo (2022) | STEM education | DT to enhance control education affordability. |
| 7. | Tjahyadi et al. (2023) | STEM education | Digital twin-based laboratory for control engineering. |
| 8. | Bunse et al. (2022) | STEM education | Distance laboratory courses in engineering education. |
| 9. | Boettcher et al. (2023) | STEM education | DT for laboratory experiments in engineering. |
| 10. | Christopoulos et al. (2022) | STEM education | DT and virtual reality in robotics education. |
| 11. | Kandasamy et al. (2022) | STEM education | DT for cyber security testing, research and education. |
| 12. | Orsolits et al. (2022) | STEM education | DT to facilitate mixed reality-based robotics. |
| 13. | Samak et al. (2023) | STEM education | DT to support simulation for autonomous vehicles. |
| 14. | Hazrat et al. (2023) | STEM education | Enhance engineering education through Industry 4.0. |
| 15. | Mathur (2023) | STEM education | DT to support cybersecurity research, education, and training for preserving critical infrastructure. |
| 16. | Peshkova et al. (2023) | STEM education | Digital twin usage in medical education. |
| 17. | Goppold et al. (2022) | STEM education | DT for visualizing error consequences in health. |
| 18. | Rovati et al. (2024) | STEM education | Developing a patient DT for critical care education. |
| 19. | Gary et al. (2023) | STEM education | DT in critical care training. |
| 20. | Geuer et al. (2023) | STEM education | Designing smart photometry in Education 4.0. |
| 21. | Lee et al. (2023) | STEM education | DT in math education (enhance gamified learning). |
| 22. | She et al. (2023) | Online education | DT for enhancing credit management in education. |
| 23. | Mourtzis et al. (2023) | Online education | DT hybrid model for personalized education 4.0. |
| 24. | Andreasyan and Balyakin (2022) | Online education | Implementing DT technology in educational practices (application of education digitalization). |
| 25. | Hsiao et al. (2022) | Online education | DT to support real-time expert feedback. |
| 26. | Rubtsova et al. (2021) | Online education | DT for optimizing learning effectiveness. |
| 27. | Dhananjaya et al. (2024) | Online education | Digital recommendation system for personalized learning in online education. |
| 28. | Mitra (2023) | Online education | DT for innovative metaverse-based education. |
| 29. | Xi and Cong (2022) | Online education | DT like a virtual tour in survey education. |
| 30. | Flaga and Pacholczak (2022) | Online education | DT for education and training purposes (bridge real and virtual worlds). |
| 31. | Eriksson et al. (2022) | Vocational education | Applying DT in higher education (example of industrial-like laboratories). |
| 32. | Vrysouli et al. (2022) | Vocational education | DT to enhance vocational education in construction (sustainability and architectural design). |
| 33. | Ruppert et al. (2022) | IoT education | DT of the laboratory (education for Industry 5.0). |
| 34. | Komninos and Tsigkas (2022) | Environmental education | Prototyping DT system (combining smart birdhouse and an electronic DT). |
| 35. | Georgakopoulos et al. (2023) | Sustainable education | Remote lab courses for distance education (designed on the sustainable development principles). |
| 36. | Intelligence (2023) | Vocal Music Education | Application of the DTs platform for music education in higher institutions. |
| 37. | Liu and Jiang (2022) | Physical Education | DT to improve somatosensory recognition efficiency (visual sensing training system). |

# 5 Enhancements of Introducing DT

This chapter will highlight five categories in which DT is believed to contribute to enhancement when utilized in the educational process.

## 5.1 STEM Education

By 2030, higher education is expected to undergo substantial changes. Launching DT will improve learning effectiveness and open up new research options (Rubtsova et al., 2022). Increased robot use in industrial production necessitates adaptation of education programs. Combining immersive technologies like mixed reality with robotics education offers promising opportunities for expanding knowledge (Orsolits et al., 2022). The use of digital twin technologies can motivate students to study and improve learning outcomes (She et al., 2023). In STEM education, a portable photometric measurement device called the Smart Education Photometer (SmaEPho) designed for inquiry-based learning, overcomes the challenges of limited budgets in purchasing professional photometric measurement systems (Geuer et al., 2023). This makes it easier for students to learn about electrical circuits and photometry, allowing them to test and build circuits while lecturers use DT as a demonstrator.

Furthermore, online learning feature analysis technology improves online education management by integrating students' cognitive data and teachers' resources (Intelligence, 2023). Christopoulos et al. (2022) are proposing an approach which can be utilized to teach robotics in blended learning scenarios. The networked virtual world accurately represents laboratory robots and tools, whereas the digital twin system simulates actions that occur in the physical laboratory.

## 5.2 Virtual Laboratories and Simulations

Virtual labs are extensively used in chemistry, physics, biology, environmental science, and medical education to simulate experiments that may be too dangerous, expensive, or time-consuming to perform in a traditional lab setting (Geuer et al., 2023; Mourtzis et al., 2023; Gary et al., 2023). In medical education, DT can be used to create a database of data that will help in further research of diseases analysis (Peshkova et al., 2023). The use of digital twins in engineering education provides a safe environment for students to explore the behavior of complex machinery and control systems without the need for physical prototypes (Chamorro-Atalaya et al., 2024; Acker et al., 2023; Andreasyan and Balyakin, 2022; Dhananjaya et al., 2024; Samak et al., 2023; Terkaj et al., 2024; Tjahyadi et al., 2023). Chamorro-Atalaya et al (2024) further described simulations based on DT as advanced interactive learning environments in engineering disciplines allowing students to virtually replicate laboratory equipment or practical teaching stations (Mourtzis et al., 2023; Samak et al., 2023; Andreasyan and Balyakin, 2022; Dhananjaya et al., 2024). Virtual simulations are crucial in medical education (Chamorro-Atalaya et al., 2024) for providing a risk-free environment for students to practice surgical procedures, patient care, and diagnostic skills (Hagedorn et al., 2023; Rovati et al., 2024). Modeling using tools like Simscape in a DT based laboratory can aid in visualizing mathematical models and concepts, helping students understand abstract concepts through interactive and engaging visual representations (Tjahyadi et al., 2023). Such visualizations could also be used for describing the process of the DT to better understand the real-time reflection of environmental changes that can be further used in various environmental studies.

Hsiao et al. (2022) introduced a cyber-physical co-existence environment for practical distance learning activities - CPE: "Co-existing Practical Environment" which contains a "Holographic Wearable Device" for behavior sensing and vision sharing, a cloud "Digital Twin Model" database for expert variable correction in real time, and a discussion interface.

Georgakopoulos et al. (2023) proposes remote laboratory courses based on the educational principles for sustainable development. The benefits of this strategy include helping students acquire critical thinking, problem solving, and collaborative competencies. DT leads to increase student motivation (Terkaj et al., 2024; Acker et al., 2023; Chamorro-Atalaya et al., 2024), encourage self-responsibility for learning (Bunse et al., 2022; Chamorro-Atalaya et al., 2024), facilitate peer learning (Bunse et al., 2022), and improve content delivery (Terkaj et al., 2024; Acker et al., 2023), demonstration ease (Lee et al., 2023), and student assessment (Lee et al., 2023; Mourtzis et al., 2023).

## 5.3 Personalized Learning Environments

Personalized learning is a process in which the method of learning and instructions are adjusted or personalized to the learner's requirements (Mitra, 2023). AI-driven technology has the potential to revolutionize education by providing personalized learning experiences (Dhananjaya et al., 2024). Digital twins could contribute to medical education and improve diagnostics of diverse forms of disease where there is a need for personalized treatment strategies in healthcare (Peshkova et al., 2023).

Intelligence (2023) explores the use of DTs technology in music education, focusing on improving teaching quality through the integration of real and virtual teaching spaces and online learning data analytics. It further suggests that a well-designed DT system can enhance students' interest in music learning by facilitating interaction and providing real-time feedback. Chatbots and virtual assistants can enhance student education by providing personalized

assistance, answering questions, offering feedback, and suggesting learning resources (Dhananjaya et al., 2024). This demonstrates how DT can be tailored to individual students, maximizing their potential.

### 5.4 Economic Alternatives

In robotic education, the open-source omnidirectional mobile robot platform (ROBOTONT) can offer professional tools for robotics education and provide researchers a portable platform for validating scientific results (Raudmäe et al., 2023), providing a DT based economical alternative for conducting research when access to a physical platform is unavailable (Acker et al., 2023; Eriksson et al., 2022; Flaga et al., 2022).

Acker et al. (2023) offer a way for secondary school students to study Industry 4.0 technologies, build ICT skills, and establish a practical learning environment by utilizing a low-cost digital twin prototype. The prototype enables STEM skills development and critical thinking through simulations, testing, and implementation of simple algorithms, facilitating real-time data exchange between simulation environments and robotic systems.

Zangl et al. (2023) are introducing a low-cost robot platform that can be used to handle a wide range of measurement science and sensor subjects, as well as machine learning, actuators, and mechanics. The 3D printed chassis may be outfitted with various sensors for environmental perception while being adaptable to a variety of embedded computer platforms.

To present principles of robot simulation and realization approaches in a hands-on manner, the table-top robot is also available as a digital twin. Xi and Cong (2022) describe how Tianjin University's field trip for a measured survey of built heritage was replaced with remote solutions that included on-site data gathering, post-processing, online instruction, observation, modeling, and delivery. Čech and Vosáhlo (2022) present a cost-effective approach for teaching digital twins in control engineering education, demonstrating their application using a gantry crane simulator.

Komninos and Tsigkas (2022) describe usage of DT to give children the ability to interact with the natural environment from the classrooms, which consists of a smart birdhouse and an electronic DT where the smart birdhouse is placed in a natural setting to collect atmospheric data, such as humidity and temperature measurements, as well as record sounds and images of potential birds. At the same time, the collected data will be available to the end user via a dashboard, which plays the acquired music or images.

### 5.5 Develop Skills

Digital twin technology can promote development of multidisciplinary skills (Acker et al., 2023; Bunse et al., 2022; Flaga et al., 2022; Lee et al., 2023; Samak et al., 2023), thus preparing students to cope with the future demands of the industry (Hazrat et al., 2023). Some authors propose a framework to prepare students for the development and application of Industry 5.0 technologies (Ruppert et al., 2022; Acker et al., 2023; Eriksson et al., 2022; Lee et al., 2023; Tjahyadi et al., 2023). To better incorporate the increasing competency requirements of the working world 4.0 Boettcher et al. (2023) employed a real-world scenario (RWS) to address the ambiguous problem-solving assignments in W4.0. DT concept could also be used for smart grid security studies, allowing users to gain experience testing attacks and countermeasures in a controlled setting (Kandasamy et al., 2022). Mathur (2023) highlights its utility in research, education, and training in critical infrastructure defense. It could be used in various fields to improve not only the mental but also physical abilities of students. For example, Liu and Jiang (2022) explore the application of DT in physical education teaching practice, involving 25 participants in an exercise detection and analysis experiment, aiming to design a platform for efficient information collection and processing.

On the other hand, Vrysouli et al. (2023) is introducing ideas of DT and sustainability in the vocational program in the sector of construction works, structured environment and architectural design. Goppold et al. (2022) presents a learning system based on learning from errors. In a technical proof-of-concept, DTs are used to simulate and see the harmful consequences of erroneous acts.

## 6 Future Directions

While existing studies discuss the benefits of digital twins in various educational contexts, there is a gap in a comprehensive, comparative analysis of how these benefits vary across different educational levels (elementary, secondary, and higher education). In addition, ethical implications and privacy concerns associated with the use of digital twins in education are not thoroughly explored. This includes issues related to data security, consent, and the potential misuse of personal data. There is a lack of adequate education about the digital twin concept, which is why efforts should be made to develop cost-effective teaching strategies to bridge the gap between education, research, and industrial practice (Čech and Vosáhlo, 2022). Acker et al. (2023) and Orsolits et al. (2022) also agree that cost-effective solutions should be pursued to improve robotics education, which will be critical in meeting future industry demand. Not only for its cost-effectiveness, it is also necessary for simulations that otherwise would not be possible to perform like offensive security testing (Kandasamy et al., 2022). DT has often been associated with the IoT, but visual sensor technology is now increasingly being explored to collect data without need for wearable devices (Liu and Jiang, 2022). This new perspective on the digital twin will raise various open questions, including the privacy threat it might bring.

# 7 Conclusions

The use of digital twin technology has yielded several benefits in a wide range of fields, including STEM education, distance learning, vocational training, and other fields including music, physical education, internet of things and environmental education. The advantages of DT are most evident in STEM education, particularly in the areas of robotics and engineering, where it acts as a mold for conducting a variety of experiments. As an applicable economic alternative for developing multidisciplinary competencies, it has gained popularity and is employed for both virtual laboratory enhancement and diverse simulation purposes. It helps to create personalized learning environments in which the instructions and learning strategy are customized to meet the needs of each individual student.

While there are educational benefits of digital twins, they should be taken with caution. The existing studies do not fully address ethical implications and privacy concerns, such as data security, permission, and the use of personal data. Future studies should therefore examine the responsible use of digital twin technology, particularly when combined with other emerging technologies like augmented reality and artificial intelligence tools. The digital twin could be a useful tool for creating future learning environments with a focus on bridging the gap between academics and industry because of the advantages it provides. For this reason, digital twin-based virtual environments may have an impact on future directions for educational practice and policy.

# References

Acker, J., Rogers, I., Guerra-Zubiaga, D., Tanveer, M. H., & Moghadam, A. A. A. (2023). Low-Cost Digital Twin Approach and Tools to Support Industry and Academia: A Case Study Connecting High-Schools with High Degree Education. *Machines*, *11*(9), 860.

Andreasyan, A., & Balyakin, A. (2022). Transformation of education through Big Data: Digital twins case study. Journal of Physics: Conference Series, 2210(1), 012003. https://doi.org/10.1088/1742-6596/2210/1/012003

Boettcher, K., Terkowsky, C., Schade, M., Brandner, D., Grünendahl, S., & Pasaliu, B. (2023). Developing a real-world scenario to foster learning and working 4.0–on using a digital twin of a jet pump experiment in process engineering laboratory education. *European Journal of Engineering Education*, *48*(5), 949-971.

Bunse, C., Kennes, L., & Kuhr, J.-C. (2022). On using distance labs for engineering education. Proceedings of the 4th International Workshop on Software Engineering Education for the Next Generation, 5–11. https://doi.org/10.1145/3528231.3528355

Čech, M., & Vosáhlo, M. (2022). Digital twins and hil simulators in control education–industrial perspective. *IFAC-PapersOnLine*, *55*(17), 67-72.

Chamorro-Atalaya, O., Flores-Velásquez, C. H., Flores-Cáceres, R., Arévalo-Tuesta, J. A., Zevallos-Castañeda, M., Tomás-Quispe, G., ... & Alarcón-Anco, R. (2024). Use of Digital Twin Technology in the Teaching-Learning Process, in the field of University Education: A Bibliometric Review. *International Journal of Learning, Teaching and Educational Research*, *23*(2).

Christopoulos, A., Coppo, G., Andolina, S., Priore, S. L., Antonelli, D., Salmas, D., ... & Laakso, M. J. (2022). Transformation of Robotics Education in the Era of Covid-19: Challenges and Opportunities. *Ifac-papersonline*, *55*(10), 2908-2913.

Dhananjaya, G. M., Goudar, R. H., Kulkarni, A., Rathod, V. N., & Hukkeri, G. S. (2024). A Digital Recommendation System for Personalized Learning to Enhance Online Education: A Review. *IEEE Access*.

Eriksson, K., Alsaleh, A., Behzad Far, S., & Stjern, D. (2022). Applying Digital Twin Technology in Higher Education: An Automation Line Case Study. In A. H. C. Ng, A. Syberfeldt, D. Högberg, & M. Holm (Eds.), Advances in Transdisciplinary Engineering. IOS Press. https://doi.org/10.3233/ATDE220165

EU Business School (2022). *How can the Metaverse be used in education*? Retrieved April 10, 2024, from: https://www.euruni.edu/blog/how-can-the-Metaverse-be-used-in-education/

Flaga, S., & Pacholczak, K. (2022). Demonstrator of a Digital Twin for Education and Training Purposes as a Web Application. Advances in Science and Technology Research Journal, 16(5), 110–119. https://doi.org/10.12913/22998624/152927

Gary, P. J., Rovati, L., Dong, Y., Lal, A., Cubro, E., Wörster, M., ... & Niven, A. S. (2023). Use of a Digital Twin Virtual Patient Simulator in Critical Care Education: A Pilot Study. In *A45. ICU PRACTICES, QUALITY IMPROVEMENT, AND MEDICAL EDUCATION* (pp. A1681-A1681). American Thoracic Society.

Georgakopoulos, I., Piromalis, D., Ntanos, S., Zakopoulos, V., & Makrygiannis, P. (2023). A Prediction Model for Remote Lab Courses Designed upon the Principles of Education for Sustainable Development. *Sustainability*, *15*(6), 5473.

Geuer, L., Lauer, F., Kuhn, J., Wehn, N., & Ulber, R. (2023). SmaEPho–Smart Photometry in Education 4.0. *Education Sciences*, *13*(2), 136.

Goppold, M., Herrmann, J. P., & Tackenberg, S. (2022). An error-based augmented reality learning system for work-based occupational safety and health education. *Work*, *72*(4), 1563-1575.

Hagedorn, L., Riedelsheimer, T., & Stark, R. (2023). Project-Based Learning in Engineering Education–Developing Digital Twins in A Case Study. *Proceedings of the Design Society*, *3*, 2975-2984.

Hamilton, I. and Beagle, V. (2024). *By The Numbers: The Rise Of Online Learning In The U.S. )*. Retrieved April 10, 2024, from: https://www.forbes.com/advisor/education/online-colleges/online-learning-stats/

Hawkinson, E. (2022). Automation in Education with Digital Twins: Trends and Issues. *International Journal on Open and Distance e-Learning*, *8*(2).

Hazrat, M. A., Hassan, N. M. S., Chowdhury, A. A., Rasul, M. G., & Taylor, B. A. (2023). Developing a Skilled Workforce for Future Industry Demand: The Potential of Digital Twin-Based Teaching and Learning Practices in Engineering Education. *Sustainability*, *15*(23), 16433.

HolonIQ's Education Intelligence Unit (2022). Global EdTech Venture Capital Report - Full Year 2021. Retrieved April 10, 2024, from: https://www.holoniq.com/notes/global-edtech-venture-capital-report-full-year-2021

Hsiao, C. F., Lee, C. H., Chen, C. Y., & Chang, T. W. (2022, June). An Approach of Holographic Technology for the Practical Distance Education. In *International Conference on Human-Computer Interaction* (pp. 61-70). Cham: Springer International Publishing.

Intelligence, C. (2023). Retracted: Application of Lightweight Deep Learning Model in Vocal Music Education in Higher Institutions. *Computational Intelligence and Neuroscience*, *2023*.

Itani, M., Itani, M., Kaddoura, S., & Al Husseiny, F. (2022). The impact of the Covid-19 pandemic on on-line examination: challenges and opportunities. *Global Journal of Engineering Education*, *24*(2), 105-120.

Kandasamy, N. K., Venugopalan, S., Wong, T. K., & Leu, N. J. (2022). An electric power digital twin for cyber security testing, research and education. *Computers and Electrical Engineering*, *101*, 108061.

Kartashova, L. A., Gurzhii, A. M., Zaichuk, V. O., & Sorochan, T. M. (2024, March). Digital twin technology for blended learning in educational institutions during COVID-19 pandemic. In *CTE Workshop Proceedings* (Vol. 11, pp. 411-426).

Kim, H., Shin, H., Kim, H., & Kim, W.-T. (2018). VR-CPES: A Novel Cyber-Physical Education Systems for Interactive VR Services Based on a Mobile Platform. Mobile Information Systems, 2018, 1–10. https://doi.org/10.1155/2018/8941241

Komninos, A., & Tsigkas, G. (2022, November). Prototyping a digital twin system for environmental education. In *Proceedings of the 26th Pan-Hellenic Conference on Informatics* (pp. 361-366).

Lee, J. Y., Pyon, C. U., & Woo, J. (2023). Digital Twin for Math Education: A Study on the Utilization of Games and Gamification for University Mathematics Education. Electronics, 12(15), 3207. https://doi.org/10.3390/electronics12153207

Liljaniemi, A., & Paavilainen, H. (2020). Using Digital Twin Technology in Engineering Education – Course Concept to Explore Benefits and Barriers. Open Engineering, 10(1), 377–385. https://doi.org/10.1515/eng-2020-0040

Liu, X., & Jiang, J. (2022). Digital twins by physical education teaching practice in visual sensing training system. *Advances in Civil Engineering*, *2022*, 1-12.

Mathur, A. P. (2023). Reconfigurable Digital Twin to Support Research, Education, and Training in the Defense of Critical Infrastructure. *IEEE Security & Privacy*.

Mitra, S. (2023). Metaverse: A potential virtual-physical ecosystem for innovative blended education and training. *Journal of Metaverse*, *3*(1), 66-72.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269.

Mourtzis, D., Panopoulos, N., & Angelopoulos, J. (2023). A hybrid teaching factory model towards personalized education 4.0. International Journal of Computer Integrated Manufacturing, 36(12), 1739–1759. https://doi.org/10.1080/0951192X.2022.2145025

Nikolaev, S., Gusev, M., Padalitsa, D., Mozhenkov, E., Mishin, S., & Uzhinsky, I. (2018). Implementation of “Digital Twin” Concept for Modern Project-Based Engineering Education. In P. Chiabert, A. Bouras, F. Noël, & J. Ríos (Eds.), Product Lifecycle Management to Support Industry 4.0 (Vol. 540, pp. 193–203). Springer International Publishing. https://doi.org/10.1007/978-3-030-01614-2_18

Orsolits, H., Rauh, S. F., & Estrada, J. G. (2022, October). Using mixed reality based digital twins for robotics education. In *2022 IEEE International*

*Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 56-59). IEEE.

Peshkova, M., Yumasheva, V., Rudenko, E., Kretova, N., Timashev, P., & Demura, T. (2023). Digital twin concept: Healthcare, education, research. *Journal of Pathology Informatics*, *14*, 100313.

Pew Research Center (2023). As AI Spreads, Experts Predict the Best and Worst Changes in Digital Life by 2035. Retrieved April 10, 2024, from: https://www.pewresearch.org/internet/2023/06/21/as-ai-spreads-experts-predict-the-best-and-worst-changes-in-digital-life-by-2035/

Quasner.com (n.d.). *Case studies.* Retrieved April 10, 2024, from: https://www.quanser.com/community/case-studies/

Raudmäe, R., Schumann, S., Vunder, V., Oidekivi, M., Nigol, M. K., Valner, R., ... & Kruusamäe, K. (2023). ROBOTONT–Open-source and ROS-supported omnidirectional mobile robot for education and research. *HardwareX*, *14*, e00436.

Rovati, L., Gary, P. J., Cubro, E., Dong, Y., Kilickaya, O., Schulte, P. J., Zhong, X., Wörster, M., Kelm, D. J., Gajic, O., Niven, A. S., & Lal, A. (2024). Development and usability testing of a patient digital twin for critical care education: A mixed methods study. Frontiers in Medicine, 10, 1336897. https://doi.org/10.3389/fmed.2023.1336897

Rubtsova, A. V., Anastasiia, T. V., Tikhonov, D. V., Snegirev, N. I., Bolsunovskaya, M. V., Almazova, N. I., ... & Kats, N. G. (2021, September). The Model of Digital Lifelong Education System in the Era of Grand Challenges: The Case of Multidisciplinary University. In *International Conference on Interactive Collaborative Learning* (pp. 835-843). Cham: Springer International Publishing.

Ruppert, T., Darányi, A., Medvegy, T., Csereklei, D., & Abonyi, J. (2022). Demonstration Laboratory of Industry 4.0 Retrofitting and Operator 4.0 Solutions: Education towards Industry 5.0. *Sensors*, *23*(1), 283.

Samak, T., Samak, C., Kandhasamy, S., Krovi, V., & Xie, M. (2023). AutoDRIVE: A Comprehensive, Flexible and Integrated Digital Twin Ecosystem for Autonomous Driving Research & Education. Robotics, 12(3), 77. https://doi.org/10.3390/robotics12030077

She, M., Xiao, M., & Zhao, Y. (2023). Technological implication of the digital twin approach on the intelligent education system. *International Journal of Humanoid Robotics*, *20*(02n03), 2250005.

Terkaj, W., Pessot, E., Kuts, V., Bondarenko, Y., Pizzagalli, S. L., & Kleine, K. (2024). A framework for the design and use of virtual labs in digital engineering education. 030003. https://doi.org/10.1063/5.0189669

Tjahyadi, H., Prasetya, K., & Murwantara, I. M. (2023). Digital Twin Based Laboratory for Control Engineering Education. International Journal of Information and Education Technology, 13(4), 704–711. https://doi.org/10.18178/ijiet.2023.13.4.1856

Tlili, A., Huang, R., Shehata, B., Liu, D., Zhao, J., Metwally, A. H. S., ... & Burgos, D. (2022). Is Metaverse in education a blessing or a curse: a combined content and bibliometric analysis. *Smart Learning Environments*, *9*(1), 1-31.

U.S. Department of Education (2023). *Fall Enrollment component, Spring 2022 (final data) and Spring 2023 (provisional data)*. Retrieved April 10, 2024, from: https://nces.ed.gov/programs/digest/d23/tables/dt23_311.15.asp

Vrysouli, N., Kotsifakos, D., & Douligeris, C. (2022, September). Digital Twins and Sustainability in Vocational Education and Training: The Case of Structural Environment and Architectural Design in Vocational High Schools. In *International Conference on Interactive Collaborative Learning* (pp. 220-230). Cham: Springer International Publishing.

Xi, W., & Cong, W. (2022). Remote practice methods of survey education for HBIM in the post-pandemic era: Case study of kuiwen pavilion in the temple of confucius (qufu, China). *Applied Sciences*, *12*(2), 708.

Zangl, H., Anandan, N., & Kafrana, A. (2023). A low-cost table-top robot platform for measurement science education in robotics and artificial intelligence. *Acta IMEKO*, *12*(2), 1-5.

# From Classroom to Real-World: University Students' Experience in Learning Solidity and Blockchain

**Snježana Križanić, Neven Vrček, Matija Habuš**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin
{skrizanic, nvrcek}@foi.unizg.hr, mhabus@student.foi.hr

***Abstract***. *Blockchain technology has taken over many business processes today. We find its influence in many real-world use cases such as energy management, healthcare, supply chains, waste management, shipping, insurance, tourism, education and others. Solidity is an object-oriented programming language used to write smart contracts in the Ethereum environment, often utilizing the Remix IDE. Learning how to write smart contracts in Solidity was included in the teaching process at one of the faculties in Croatia, and a survey was conducted among students on the success of knowledge acquisition in the areas of blockchain and Solidity.*

**Keywords.** Blockchain, Solidity, university, practice.

## 1 Introduction

The use of blockchain in education has become one of the trending topics in educational technology research. However, few blockchain solutions for education have provided a measurement of impact on student learning outcomes, teaching or administrative processes (Ocheja et al., 2023). The use of blockchain, particularly in the areas of teaching and learning and academic administration, has created space for a new business model in higher education (Brown, 2022). The advantages of blockchain technology, such as decentralization, tamper-proofing, anonymity, traceability, etc., offer the possibility of improving the security of the management of teaching informatization in higher education (Zhang et al., 2020). "Blockchain technology is conducive to protecting online educational resources, strengthening the transparency of online teaching resources, simplifying the copyright transaction process, and thus improving the scientific and technological innovation ability of teachers." (Hao et al., 2023)

Normally, audit authorities and central institutions need more time to carry out operations, which is a time-consuming process. With blockchain technology, we can eliminate the risk of central server failure. The data stored in the blockchain is tamper-proof and cannot be modified once deployed on the blockchain network (Karale & Khanuja, 2019).

The motivation for writing this paper came from the curiosity to evaluate the experiences of university students learning to write smart contracts using the Solidity programming language. The Solidity programming language is used to implement smart contracts on various blockchain platforms, mainly Ethereum. The students wrote smart contracts in the Remix Ethereum IDE development environment. The goal of this research is to investigate students' experiences of learning blockchain concepts and writing smart contracts in the Solidity programming language at a faculty.

The research questions are:

1. How can blockchain technology and Solidity be used in higher education and teaching according to the existing literature?
2. What are the limitations of learning blockchain and Solidity in higher education?
3. What would be the recommendations for improving the quality of blockchain and Solidity teaching according to the research findings?

## 2 Literature Review

For the literature review, articles were selected from the scientific databases Scopus and Web of Science. Recent articles were selected based on the relevance of the research area using the keywords ("blockchain" OR "Solidity") and "education".

Distributed ledgers and smart contracts are among the blockchain features that could help improve efficiency, security and privacy in the execution of academic administration, particularly in student identification, admissions application processing, degree certification, document authentication and quality assurance (Brown, 2022). The aim of the chapter (Brown, 2022) was to introduce blockchain technology as a framework and show how it can be used in higher education teaching, learning and

administration in times of crisis to effect and manage practical change.

The authors (Nguyen et al., 2023) propose a system for ticket issuance and ticket traceability by applying smart contract technology on the blockchain platform to store, manage and trace the origin of tickets. The system was developed on the Ethereum platform to implement smart contracts and blockchain technologies using the Solidity language. The advantage of blockchain is that "it offers great security by preventing information theft and manipulation." (Nguyen et al., 2023)

In the study (Rosasooria et al., 2020), a blockchain system for electronic voting was developed to detect fraud in voting using the Solidity language. A blockchain e-voting system was proposed that has secure and transparent features for use during the Student Representative Council (SRC) election.

In the paper (Ocheja et al., 2023), the authors present the results of the review of existing visualizations in blockchain and non-blockchain educational technologies and suggest how blockchain tools for education can provide better visualizations to support learning objectives. They also report the results of an initial validation of their proposal with teachers through a qualitative method.

The paper (An et al., 2021) addresses the shortcomings of talent training modes in colleges and universities and proposes a blockchain-based intelligent teaching system for smart education. This system proposes to improve students' practical innovation ability through the comprehensive participation of tutors from school and enterprise in the students' education process.

The authors of the study (Meyliana et al., 2020) attempted to create a process model of the university teaching and learning process. They tried to answer how teaching and learning activities can be supported by blockchain technology. To do this, they used qualitative methods and focus group discussions with 9 universities.

The paper (Chan & Shan, 2022) uses the literature research method, by combining the blockchain education application related literature, summarizes and finds that the current blockchain application mainly focuses on the following aspects of learning process: tracking and analysis, education management and decision-making, open resource sharing, copyright authentication and protection, credit bank and education service supported by blockchain technology, and the future development of blockchain.

The authors (Zhang et al., 2020) sought to explore the application of blockchain in education to improve the practice of the system of teaching informatization management in higher education. They supposed that the security of informatization storage would be ensured, the credibility of data monitoring would be increased, administrative costs would be reduced, and the efficiency of higher education management would be greatly improved if blockchain technology would be used in the system of teaching informatization management.

The research (Liu, 2021) mainly explores and studies the distance education system based on blockchain technology according to the characteristics of blockchain technology. A multi-layer logical structure was designed, and the generation steps of the blockchain in the system and the operation process of the consensus mechanism were analyzed.

The authors (Thompson et al., 2023) present NFTrig, a web-based application developed as an educational tool for teaching trigonometry and blockchain technology. The creation of the application includes front-end and back-end development as well as integration with other external sources such as MetaMask and OpenSea. Key development languages include HTML, CSS (Bootstrap 5) and JavaScript, as well as Solidity for the creation of smart contracts. The technical report describes how the application was created, the requirements of the application and how the smart contract was designed with security aspects in mind.

Studies on the application of blockchain technology in education were analyzed in (Hao et al., 2023), and the influences of the aspects of the application of blockchain technology in education on the sharing of online teaching resources were investigated. Based on the research on blockchain technology, educational applications and the sharing of online teaching resources, the study developed a questionnaire entitled “The influence of blockchain technology application in education on online teaching resource sharing”, which mainly included the following aspects. The first aspect was to measure the educational application of blockchain technology. The second aspect was to measure the online sharing of teaching resources. The third aspect, the descriptive statistics of respondents, mainly included 5 questions: respondents' gender, title, age and colleges/universities, and their familiarity with blockchain technology.

The research (Swati & Nitin, 2023) presents an innovative smart scholarship management system that utilizes blockchain technology. The research presents an Ethereum-based implementation that uses Solidity for backend smart contracts and ReactJS for the frontend. The experimental evaluation validates the gas cost for transaction execution and deployment cost.

The article (Liu & Zhu, 2021) analyses the technical foundations of blockchain and the shortcomings of online teaching and cross-regional teaching in cultural and creative design education. Blockchain technology is used in the management of learning resources, the tracking of learning processes, learning evaluation, the design of learning pathways and teaching management and assessment. The conclusion is that "blockchain technology can effectively optimize the teaching structure and management in cultural and creative design education" (Liu & Zhu, 2021) and promote the adaptation of design education to contemporary development.

The authors (Li & Zhao, 2021) investigated the application of blockchain technology in the Open University's distance learning sector. Their research adopts a hierarchical network architecture and hierarchical storage structure, which has good scalability and node transparency, and supports the user service strategy in large-scale applications.

The article (Li & Wu, 2019) analyses the core technology and the principle of blockchain and presents its advantages for use in the classroom. In the context of educational practise, the authors explain the structure and function of the framework, core components and application modules of the blockchain teaching platform. The research also explains the method of building the blockchain teaching platform and introduces problems and thoughts on blockchain application in school teaching.

The research (Widjaja et al., 2020) attempts to look for outstanding issues, current trends, benefits and challenges of blockchain technology in the education sector and the impact of blockchain technology on the important elements in universities such as people, processes and technology. The research is conducted through a systematic literature review to get a complete picture of blockchain technology in universities. The results show that the people, processes and technology aspect of the university has a direct impact, but the impact of blockchain technology brings many benefits not only to the students but also to the institution. Efficient processes, secure data, reliable and trustworthy information could be another value for the university.

The advantages of using blockchain in education and administration are seen in the existing literature primarily in the security of the system against unauthorized manipulation and data theft. In addition, the use of smart contract technology for data storage, management and monitoring was noticed. Finally, the studies attempted to answer how teaching and learning activities can be supported by blockchain technology.

# 3 Methodology

In 2024, writing smart contracts in the Solidity programming language was taught for the first time at one of the faculties in Croatia. Writing smart contracts was taught by students in two related courses at the same faculty. These two courses were related to the digital business. One of the courses was conducted as part of the undergraduate vocational study programme and the other as part of the graduate study programme.

During the semester, students had the opportunity to go through the material on writing smart contracts in Solidity in about ten lessons, starting from learning the basics of the programming language to more complex tasks that occur in the real world of blockchain. The lessons were held in the form of exercises on the computers and the lectures. The basic concepts of blockchain and digital business were taught in the lectures, and the programming of smart contracts in Solidity was taught in the exercises on the computers. Students first learned the basics of programming in the Solidity programming language and then worked on real examples of transactions that occur in the blockchain. In addition, the students had the task of independently creating a Solidity project by researching specific real-world examples and programming them independently.

For research purposes, a survey was conducted among students who were studying the programming of smart contracts in the Solidity. From 40 full-time students, attended the undergraduate vocational study (UVS) program, 36 of them participated in the survey. From 114 full-time students, attended the graduate study (GS) program, 85 of them participated in the survey. Participation in the survey was anonymous and voluntary. The respondents of the survey were IT students.

The survey was conducted using Google forms and the questions were mainly aimed at assessing the satisfaction or success of the materials adopted and the application of the knowledge acquired. Responses were recorded via online Google forms on a scale of 1 to 5, with 1 representing "least satisfaction" and 5 representing "greatest satisfaction", i.e., "excellent performance". After completing and submitting the surveys, the Google forms summarized the results in columns for each question. More specifically, the percentages for each grade were calculated for each question. The students filled out the survey at the end of the semester. The survey questions were structured as follows: First, demographic data such as age of the respondents were examined. Then, questions about previous experience with programming were asked. This was followed by questions about knowledge of blockchain technology and the Solidity programming language, examining respondents' knowledge of these topics before the course and their understanding of blockchain and Solidity concepts after the course they had attended. Questions were then asked about the learning experience, in which respondents were asked about learning challenges. After the learning experience questions, respondents were asked questions about the practical application of Solidity. The last group of questions related to the respondents' future interest in the further application of knowledge in the field of blockchain and Solidity.

# 4 Results of the survey

This chapter describes the results of an online survey conducted among students participating in courses about blockchain and writing Solidity smart contracts. The results of the survey showed that the majority of respondents, 97% from the UVS program and 98% from the GS program, were between 18 and 25 years old. The remaining respondents were between 25 and 30 years old.

According to the results, the respondents of both study programs (UVS and GS) stated that they already had experience with programming in the following languages: C++, C, C#, Python, Java and JavaScript. Considering the similarity of Solidity with the listed programming languages, this result can be considered very significant as it means that it was easier for the students to acquire the knowledge of programming smart contracts in Solidity. The comparison of the results of knowledge about blockchain technology and the Solidity programming language from two study programs can be seen in Table 1.

**Table 1** Knowledge of blockchain technology and the Solidity programming language: a comparison of the results of two study programs

| Question | Undergraduate vocational study (UVS) programme | Graduate study (GS) programme |
|---|---|---|
| Familiarity with blockchain technology before the class? | Insufficiently- 22,2%; sufficiently – 44,4%; good – 19,4%; very good – 13,9%; excellent – 0%. | Insufficiently- 27,1%; sufficiently – 43,5%; good – 21,2%; very good – 7,1%; excellent – 1,2%. |
| Familiarity with Solidity programming language before the class? | Insufficiently- 86,1%; sufficiently – 11,1%; good – 0%; very good – 2,8%; excellent – 0%. | Insufficiently- 92,9%; sufficiently – 3,5%; good – 1,2%; very good – 1,2%; excellent – 1,2%. |
| Level of understanding of blockchain concepts after attending the class on this topic? | Insufficiently- 0%; sufficiently – 2,8%; good – 27,8%; very good – 61,1%; excellent – 8,3%. | Insufficiently- 1,2%; sufficiently – 2,4%; good – 30,6%; very good – 61,2%; excellent – 4,7%. |
| Level of understanding of writing smart contracts in Solidity after attending the class on this topic? | Insufficiently- 0%; sufficiently – 13,9%; good – 33,3%; very good – 47,2%; excellent – 5,6%. | Insufficiently- 2,4%; sufficiently – 3,5%; good – 44,7%; very good – 40%; excellent – 9,4%. |

We knew that the responders were familiar with the C programming languages, Python, Java and JavaScript before this course, but what about blockchain technology and the Solidity programming language? According to Table 1, we can notice how the understanding of blockchain concepts increased among students in both study programs after participating in the courses. In addition, students had little or no familiarity with the Solidity programming language before the course. After the course, the majority of respondents in both study programs feel that they have a good or very good understanding of writing smart contracts in Solidity. It is interesting to note that 5.6% of students on the UVS program and 9.4% of students on the GS program believe that they have an excellent command of writing smart contracts. It can be assumed that these are students who may be interested in applying programming using Solidity in their future careers or simply continuing their personal development in this direction.

**Table 2** A learning experience: a comparison of the results of two study programs

| Question | Undergraduate vocational study (UVS) programme | Graduate study (GS) programme |
|---|---|---|
| Were there any aspects of learning the Solidity language that you found challenging? | Yes – 19,4%<br>No – 80,6% | Yes – 17,6%<br>No – 82,4% |
| What resources have helped you the most in mastering the Solidity language? | Teaching material from the course page + exercises on the computer: 77,8%, Online content: 22,2%. | Teaching material from the course page + exercises on the computer: 64,7%, Online content: 32,9%, Something else: 2,4%. |
| Evaluate the effectiveness of the teaching methods and materials used in the course (exercises on the computer, project assignment, etc.) in mastering Solidity and blockchain: | Insufficiently- 0%; sufficiently – 0%; good – 16,7%; very good – 50,0%; excellent – 33,3%. | Insufficiently- 2,4%; sufficiently – 7,1%; good – 21,2%; very good – 40,0%; excellent – 29,4%. |

In the survey, respondents were asked if there were any aspects of writing smart contracts in Solidity that they found challenging. If so, respondents were asked to explain these aspects. Most responded that it took time to understand the differences of the Solidity language compared to other languages. They responded that once you understand how the logic of smart contracts works conceptually, "everything works smoothly". They also responded that understanding some of the specifics of the language took a little more time. In addition, some respondents stated that they were generally less good at programming or simply did not have enough experience with programming. In general, the answers were along the direction that you first have to catch the differences compared to other programming languages. According to Table 2, the majority of respondents, 77.8% in the UVS program and 64.7% of respondents in the GS program, declared that the teaching material from the course page, including the exercises on the computer, helped them the most in learning the Solidity language, and then the online content. To learn Solidity, 22.2% of responders in the UVS program and 32.9% of responders in the GS program used online content. In addition, respondents rated the effectiveness of the teaching methods and materials used in the course for mastering Solidity and blockchain very highly. As many as 50% of respondents of the UVS program consider the methods and materials are very good and 33.3% consider the methods and materials are excellent. For the GS program, 40.0% of respondents believe that the methods and materials are very good and 29.4% consider them to be excellent.

**Table 3** Practical use of Solidity and blockchain knowledge: a comparison of the results of two study programs

| Question | Undergraduate vocational study (UVS) programme | Graduate study (GS) programme |
|---|---|---|
| Have you had the opportunity to use your Solidity skills outside the classroom (e.g., on your own projects)? | Yes - 19,4%,<br>No – 80,6% | Yes – 20,0%,<br>No – 80,0% |
| How ready do you feel to apply Solidity in the real sector? | Insufficiently- 0%; sufficiently – 27,8%;<br>good – 44,4%; very good – 27,8%; excellent – 0,0%. | Insufficiently- 9,4%; sufficiently – 29,4%;<br>good – 47,1%; very good – 12,9%; excellent – 1,2%. |

Regarding the practical use of Solidity and blockchain knowledge, the majority of responders on both study programs stated that they did not have the opportunity to apply their Solidity skills outside of the classroom, e.g., on their own projects or in a real sector. However, according to the results from Table 3, the majority of responders in both study programs feel sufficiently prepared or ready to apply their knowledge and skills in the real sector. It is assumed that these are students who are willing to work in the real sector, where applications are developed in blockchain, and who are familiar with programming in Solidity after the course.

**Table 4** Future interests for using Solidity and blockchain: a comparison of the results of two study programs

| Question | Undergraduate vocational study (UVS) programme | Graduate study (GS) programme |
|---|---|---|
| How interested are you in further projects related to blockchain technology and Solidity? | Insufficiently- 2,8%; sufficiently – 16,7%;<br>much – 38,9%; very much – 27,8%; the most – 13,9%. | Insufficiently- 17,6%; sufficiently – 21,2%;<br>much – 38,8%; very much – 18,8%; the most – 3,5%. |
| Assess how knowledge of blockchain technology and Solidity will benefit your future career or profession. | Insufficiently- 0%; sufficiently – 11,1%;<br>good – 47,2%; very good – 38,9%; excellent – 2,8%. | Insufficiently- 11,8%; sufficiently – 29,4%;<br>good – 36,5%; very good – 15,3%; excellent – 7,1%. |

Table 4 shows that respondents in the UVS program are more interested in participating in future projects related to Solidity and blockchain than students in the GS program. Furthermore, it was found that the same respondents were more likely to see the benefit of knowing a new programming language for their future career and/or profession. The reason for this could be that students in the GS program have already found an

interest in another field during their studies or do not believe that they will need programming skills in the future.

Most respondents in both programs believe that knowledge of blockchain technology and Solidity will benefit them in their future career or profession. As many as 47.2% of respondents in the UVS program believe this knowledge will benefit them good, and 38.9% of respondents believe it will benefit them very good. 36.5% of respondents in the GS program believe that this knowledge will be of good use to them, 15.3% believe that it will be of very good use to them, and 7.1% believe that knowledge about blockchain and Solidity will be of excellent use to them.

# 5 Limitations of learning blockchain and Solidity in higher education

The limitations of the learning outcomes may be hidden in the fact that writing smart contracts in Solidity has not yet reached its full potential as it is a new technology, so students do not yet realise the importance of knowing the new technology. Students are used to the knowledge taught at university, so it can be more difficult for them to perceive new knowledge and new technologies with which they were previously unfamiliar. It is assumed that as writing smart contracts in Solidity and knowing how transactions in blockchain work become more dominant in the market, students will also feel the need to keep up with the times and technologies.

It should also be noted that many students do not have an affinity for programming, so learning experience surveys may show weaker results when respondents are asked about their future interests. Most students may not want to programme in the future and see their professional career in other fields.

The limitations of learning blockchain and Solidity are seen in several aspects: students should have developed the basic concepts of object-oriented programming. They should also be more motivated to learn a new programming language. Prior knowledge about the blockchain technology would also be helpful. It would be scientific interesting to repeat this type of survey in the future with other students to see if students have recognized the importance of knowing another programming language over time and how blockchain transactions work.

# 6 Recommendations for improving the quality of blockchain and Solidity teaching

In order to successfully teach blockchain and Solidity, it is necessary to keep up with the times and monitor the projects appearing on the market and the private sector's need for people who have a basic skill of blockchain and are able to independently create smaller applications such as smart contracts in Solidity.

We believe that we offered a lot to the students during the course, because we taught them the basics of the programming language, we taught them to create real smart contracts for sending transactions on the blockchain as they usually occur in real sector, and the students created independent projects in which they were to show their autonomy, the ability to create "real world" smart contracts and present what they had learned. The experiences we gave them were theoretical and practical in nature.

The recommendations that we would give are in the direction of practical examples and independent work of the students in the form of projects, with a previous good theoretical background and processed examples from “real world”, which the students go through together with the teacher. In addition, students can be involved in projects in which the university works together with the real sector or market. This allows students to get to know the experience of working in private companies during their education and establish contacts with individuals who conduct blockchain transactions as part of their daily business.

# 7 Conclusion

Understanding how to write smart contracts in Solidity is significant for university students due to its relevance in emerging technologies, innovation, and practical application. Blockchain technology, particularly the Ethereum platform, is rapidly gaining traction across various industries, and Solidity is the primary language for developing smart contracts on this platform. Mastery of Solidity ensures that students are well-versed in cutting-edge technology, enhancing their competitiveness in the job market. Thus, proficiency in Solidity and smart contract development equips students with valuable skills that are essential for their future careers in a technology-driven society.

The goal of this research was to investigate students' experiences of learning blockchain concepts and writing smart contracts in the Solidity programming language. Furthermore, based on the research findings, the limitations of learning blockchain and Solidity in higher education were identified and recommendations to improve the quality of blockchain and Solidity teaching were proposed. The potential impact of this research in the field of educational technology is to improve and, if necessary, adapt the teaching material to the needs of the students and to recognize the purpose and importance of the course in today's era of rapid changes in the IT sector and ICT. In the course we try to meet not only the needs of students but also those of employers who are looking for workers with specific, possibly rare, skills. Our task is to identify such employers and provide them with a workforce that meets their business needs.

# References


An, L., Yue, J., Zhang, G., & Wang, Q. (2021). BITS: A Blockchain-Based Intelligent Teaching System for Smart Education. *Proceedings - 2021 International Conference on Education, Information Management and Service Science, EIMSS 2021*, 159–162.

Brown, B.A. (2022). Blockchain Technology in Teaching, Learning, and Academic Administration as a Long-term Solution to Emergencies in Higher Education. In *The Emerald Handbook of Higher Education in a Post-Covid World: New Approaches and Technologies for Teaching and Learning* (pp. 173–194).

Chan, H. & Shan, J. (2022). The Application and Future Development Trend of Blockchain Technology in Education. *Proceedings - 2022 11th International Conference of Educational Innovation through Technology, EITT 2022*, 151–157.

Hao, Z., Bin Yahya, M.Y., & Lu, J. (2023). Influence of Blockchain Technology Application in Education on Online Teaching Resources Sharing. *International Journal of Emerging Technologies in Learning , 18(11)*, 25–37.

Karale, A.S. & Khanuja, H. (2019). Implementation of Blockchain Technology in Education System. *International Journal of Recent Technology and Engineering, 8(2)*, 3823–3828.

Li, L. & Wu, X. (2019). Research on school teaching platform based on blockchain technology. *14th International Conference on Computer Science and Education, ICCSE 2019*, *8845353*, 38–43.

Li, Z. & Zhao, Q. (2021). Blockchain Technology in Open University Distance Open Education. *Journal of Physics: Conference Series , 1992(4)*, *042043*.

Liu, J. & Zhu, T. (2021). Application of Blockchain Technology in Cultural and Creative Design Education. *International Journal of Emerging Technologies in Learning , 16(4)*, 228–239.

Liu, X. (2021). Exploration & Research on Distance Education System Based on Blockchain Technology. *Journal of Physics: Conference Series , 1769(1)*, *012041*.

Meyliana, Cassandra, C., Surjandy, Widjaja, H.A.E., Prabowo, H., Fernando, E., & Chandra, Y.U. (2020). A blockchain technology-based for university teaching and learning processes. *Proceedings of 2020 International Conference on Information Management and Technology, ICIMTech 2020*, *9211209*, 244–247.

Nguyen, H.-Q., Nguyen, H.T., & Pham, T.-T. (2023). Applying Smart Contract in Blockchain Technology to Manage the Ticketing Issuance and Ticketing Traceability. *Nguyen, H.T. , Pham, T.-T. 2023 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2023*, 160–164.

Ocheja, P., Flanagan, B., Ogata, H., & Oyelere, S.S. (2023). Visualization of education blockchain data: Trends and challenges. *Interactive Learning Environments, 31(9)*, 5970–5994.

Rosasooria, Y., Mahamad, A.K., Saon, S., Isa, M.A.M, Yamaguchi, S., & Ahmadon, M.A. (2020). E-Voting on Blockchain using Solidity Language. *Proceeding - 2020 3rd International Conference on Vocational Education and Electrical Engineering: Strengthening the Framework of Society 5.0 through Innovations in Education, Electrical, Engineering and Informatics Engineering, ICVEE 2020, 9243267*.

Swati, J. & Nitin, P. (2023). CryptoScholarChain: Revolutionizing Scholarship Management Framework with Blockchain Technology. *International Journal of Advanced Computer Science and Applications , 14(8)*, 652–659.

Thompson, J., Benac, R., Olana, K., Hassan, T., Sward, A., & Mohd, T.K. (2023). NFTrig: Using Blockchain Technologies for Math Education. *Lecture Notes in Networks and Systems , 665 LNNS*, 609–624.

Widjaja, M., Surjandy, S., Fernando, E., & Chandra, Y.U. (2020). The Recent Open Problem, Benefit and Challenge of Blockchain Technology in Education. *International Journal of Recent Technology and Engineering (IJRTE)*.

Zhang, L., Ma, Z., Ji, X., & Wang, C. (2020). Blockchain: Application in the System of Teaching Informatization Management of Higher Education. *Proceedings - 2020 3rd International Conference on Smart BlockChain, SmartBlock 2020*, *9415687*, 185–190.

# Optimization of urban road network and digital twins: a literature review

**Matija Habuš, Neven Vrček, Snježana Križanić**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin
mhabus@student.foi.hr, {nvrcek, skrizanic }@foi.unizg.hr

***Abstract**. The optimization of road traffic in urban areas has been researched for many years, with various attempts made to avoid waiting times and congestion, and to reduce air pollution and noise levels. Given the dynamic nature of traffic, it is crucial to optimize it to increase safety and reduce greenhouse gas emissions. In most cases, solving this problem involves using simulations in specialized programs that are calibrated with historical data to obtain the most accurate representation of real traffic in the network. Digital twin (DT) systems aim to create virtual replicas of physical objects that update in real-time with their physical counterparts and evolve alongside those assets throughout their life cycle. By using digital twin technology in traffic systems, it is possible to collect and use real-time traffic data and import the data into the SUMO simulation program, which can create a virtual replica of the physical traffic. This type of data usage and presentation can help traffic planners more accurately predict congestion and prescribe optimized solutions based on those predictions.*

**Keywords.** optimization of the urban road network, digital twin, simulation, traffic network, greenhouse gas emissions (GHG), intelligent transport systems (ITS).

## 1 Introduction

This article provides an overview of the interdependence between transport systems, traffic management, ITS (Intelligent Transport Systems), digital twin, urban network optimization and greenhouse gas emissions reduction. To function smoothly, efficient management of traffic flows is essential for the transport system. The transport system is directly integrated into the concept of smart city development and it is based on the applications of digital and communication technology (ICT) and Internet of Things (IoT) to achieve better connectivity, efficiency, functionality and sustainability. Traffic flows are organised hierarchically and their management is enhanced by intelligent transport systems. ITS work on the principle of automatic or adaptive systems (using one of the methods of computer science: deep learning, machine learning, artificial intelligence, neural networks). The input data used for ITS is collected from traffic counters at road junctions or intersections. The data collected consist of the number of vehicles from different categories. The function of ITS is to provide feedback by sending, receiving, and exchanging information between different entities (such as drivers, passengers, road infrastructure, construction facilities, equipment, and the environment) in order to increase road safety, reduce pollution, improve traffic flows, direct vehicles on roads and intersections both in and outside cities, increase mobility, reduce travel time, and lower the proportion of costs generated by traffic. With DIRECTIVE 2010/40/EU, the European Parliament and the Council of the European Union have adopted a legal framework that describes the obligation and importance of introducing an ITS in the road transport system. In the literature review, the concept of digital twin was considered as a technology that enables a different dimension of exploring traffic flow optimization. The basic definition of the digital twin represents a model of an object or system that continuously uses data collected from sensors, network-connected devices and other data sources to present a realistic and dynamic physical model or system. According to Kritzinger and co-authors (2018.), the digital twin is a digital representation of a physical entity with a fully integrated and automated bidirectional data flow to the physical entity. Wu and co-authors (2023) divide the digital twin into three parts: a physical entity, its virtual counterpart, and a mapping between the actual object and the virtual twin, allowing the physical and virtual sides to evolve together. These different definitions offer multiple perspectives on digital twins and show that digital twins play a major role in research and practice. According to Zirin Wang and co-authors (2022), in a paper describing the concept, architecture, case study, and future challenges in the transformation of digital

twins in urban mobility, a digital twin is defined as a digital image of a living or non-living physical entity. They note that this still-emerging technology is attracting significant attention across various industries this decade. However, while some studies have investigated digital twins in transportation, existing research lacks systematic connections between different entities: people, vehicles, and the transportation environment as a whole. A theoretical framework has been developed based on the application of artificial intelligence that utilises data from the database, cloud, edge computing, and smart devices for mobility services. Based on the database queries 'traffic flow management' and 'traffic flow management at intersections,' proposals were developed that describe solutions and improvements through the application of traffic simulation programs utilizing predictive analytics to optimize traffic flow and the traffic system overall. Predictive analytics in traffic system is applied through machine learning and artificial intelligence, where data is collected through sensors and devices connected via Internet of Things and used based on historical data. Mathematical models of traffic flows are also created using mathematical methods such as graph theory, tail theory, linear programming and the traffic network. The theory of traffic flow is based on various parameters such as vehicle flow, traffic flow density, traffic flow speed, vehicle travel time, vehicle following interval and distance between vehicles. These are all mathematical quantities that are mutually dependent and connected. Since the research area is based on the framework of digital twins, which operate using real-time data, and traffic flows in the urban environment, the need for real-time measures is investigated through a case study. Decision theory has been imposed as important for decision making in order to optimize and reduce as much as possible the above-mentioned deficiencies that occurring in the transport network. Traffic flow management based on predictive analytics provides effective results, but a prescriptive decision-making theory that answers the question, 'How can better decisions be made?' and prescriptive analytics for decision-making show that there is room for improvement in the optimization of traffic flow. Prescriptive analytics and optimization are often used as synonyms and describe a provable mathematical model that is able to find the best possible variant for achieving certain optimization goals.

The aim of this article is to examine the results of previous research in the field of traffic flow optimization in urban networks using the digital twin methodology. The following research questions can therefore be placed: What are the limitations of previous research? Is there a way to solve the perceived shortcomings more successfully in order to achieve some progress in optimizing traffic flows in urban road networks? This article is organized into the following chapters: Chapter 2 explains the methodology used to conduct this study. Chapter 3 presents a simulation of the traffic flow in the urban network and the digital twin. Chapter 4 shows the limitations of previous research and further research. Chapter 5 is the conclusion.

# 2 Methodology

For the purpose of this study, a literature review was conducted in the areas of traffic simulation at signalized and unsignalized intersections, as well as digital twins. The literature search was conducted using Scopus, Web of Science, and IEEE Xplore databases, limited to the period from 2019 to 2024 and focused on the fields of Computer Science and Engineering. Considering the complexity of the research field, the search query in the three databases consisted of several queries: *("road traffic AND "digital twin"), ("road network optimization AND "digital twin"), ("road network modeling and simulation AND "digital twin"),* and resulted:

**Table 1** Citation database and query

| Citation database /query: | Web of Science | Scopus | IEEE Xplore |
|---|---|---|---|
| Road traffic digital twin | 86 | 175 | 240 |
| Road network optimization digital twins | 40 | 31 | 44 |
| Road network modeling and simulation digital twin | 65 | 33 | 93 |

When searching the databases, 807 articles related to the research field were displayed. Given that the authors explore the concept of digital twins in traffic management and the optimization of traffic systems in various ways, it is understandable that a large number of articles were retrieved. Also, many of articles were found in all three databases, so the total number of articles that could be analyzed in the field of interest is greatly reduced. The topics that researchers address and investigate in the field of digital twins are connected with the following areas:

1. The maintenance of road infrastructure, bridges, and tunnels
2. The relationship between road conditions, tire condition, and vehicle controllability
3. Methods for collecting, analyzing, and visualizing traffic data
4. The placement of sensors on specific roads to gather information about the traffic situation
5. The relationship between autonomous vehicles and human-driven vehicles
6. The impact of lane changes on traffic accidents

7. Communication between autonomous vehicles, other road users, Intelligent Transportation Systems (ITS), and the Internet of Things (IoT)
8. The impact of road infrastructure lighting
9. The role of artificial intelligence in traffic management
10. The application of agent-based traffic management and deep learning methods, including Deep Reinforcement Learning (DRL), in adaptive traffic signal management
11. The use of blockchain methodology in traffic management
12. The cybersecurity of collected traffic data
13. The optimization of traffic to reduce exhaust emissions and improve signal timing at intersections
14. The dynamic management of signal phase duration at traffic light intersections and on freeways
15. The synergy between real-time collected traffic data and simulations for a virtual representation of traffic on a highway section.

After analyzing the abstracts of the articles retrieved from the database search, 10 articles were selected to represent the research area and to serve as a framework for future studies.

# 3 Simulation of traffic flow in urban network and digital twins

This chapter presents the results of previous research in the field of road traffic optimization using the digital twin methodology. The research by Kušić and co-authors (2022) describes an application that performs microsimulations based on carefully divided traffic data. This data is continuously collected and stored in a database according to traffic intensity, which is monitored by traffic counters positioned on a section of the motorway in Geneva. These counters provide real-time insights into traffic flow, allowing the researchers to simulate and analyze the dynamics of vehicle groups with a high degree of accuracy. This continuous data collection is crucial for ensuring that the microsimulations reflect the actual traffic conditions, thereby enhancing the reliability and applicability of the simulation results for traffic management and planning. This synchronization ensures that the simulations are as accurate and reflective of real-world conditions as possible. However, they noted a significant gap in the existing literature: there was no detailed methodology or set of steps available for dynamically generating and calibrating traffic flows by creating scenarios and microsimulations within the SUMO (Simulation of Urban Mobility) traffic simulation program. Recognizing the absence of detailed methodologies for dynamically generating and calibrating traffic flows in SUMO, they proposed a comprehensive methodology for creating a digital twin of the Geneva Motorway in their paper. This methodology focuses on the dynamic calibration of traffic flow and the dynamic rerouting of vehicles, which are essential mechanisms within the SUMO application. These mechanisms are based on application objects that can be calibrated, ensuring that the digital twin accurately mirrors real-world traffic conditions and responds to changes in real-time. They highlight that the dynamic calibration of traffic flow and rerouting of vehicles can be controlled during the simulation by interacting with the SUMO and TraCI (Traffic Control Interface) interfaces. These interfaces are essential for dynamically generating the desired traffic volume and executing simulations based on specific traffic scenarios. Importantly, the authors note that these elements can be continuously calibrated to maintain the accuracy of the simulations. The data used in the simulations includes the actual speed of vehicles, with the number of vehicles distributed according to groups detected by the traffic counters. This data is entered into the continuously running program every minute. In this way, the digital twin of the Geneva Motorway continuously adapts to real traffic in real time by calibrating the input data. This is the main difference between the standard process of model creation and traffic simulation. Kušić and co-authors (2023) also present their research, which follows on previous work, and state that digital twins will change the technology of managing the transportation system. They present a compelling concept for monitoring the entire lifecycle of the transportation system. Shamlitsky and co-authors (2024) investigate the use of digital twins for traffic flow management to increase road safety, optimize vehicle flow and improve the functioning of the entire transport system. They conclude that it is necessary to develop effective algorithms and methods for data analysis that distinguish the collected data and information from the digital twins in order to make effective decisions. They give the methodological framework in which they created a digital twin model of traffic flow on the road network with an intersection using real data in real time using the AnyLogic program, which is intended for creating traffic model simulations. The input data for the model are the intensity of arrival of vehicles from all approaches to the intersection, the distribution of vehicles in each approach and in all directions. They tested the model by applying the Webster algorithm, which is suitable for managing traffic signal intersections, they calculated the saturation (saturation of traffic flow) and calculated the duration of the phases of the traffic signals. After testing the model through simulation, they came to the conclusion that it is possible to reproduce real traffic in real time with the help of simulation. They developed a tool as an upgrade to the AnyLogic program that enables the use of real-time data. However, they cite the inability to access real-time data as a disadvantage, due to the need for permission from the city administration and traffic control. In their work, Kamal and co-authors (2024)

present the effective application of the digital twin to adaptive traffic management at intersections with traffic signals, which has an impact on reducing harmful exhaust emissions in urban environments. From the conducted research, they concluded that artificial intelligence has potential in the application of adaptive control of traffic signalization, but little research has been done on how it affects the reduction and mitigation of harmful exhaust emissions in urban environments. In particular, they highlighted the application of DRL (Deep Learning with Reinforcement) in the adaptive control of traffic signals to reduce travel time. The methodology is described in five phases: The first phase is the collection of traffic data, the second phase is used to create a digital twin model in which models are modelled, analysed, scenarios are created based on multi-criteria decision making, and simulations are performed on a virtual replica of the intersection using the data obtained from the collection phase. The modelling is carried out as an intermediate layer between the physical model and the digital twin, in which the TraCI interface (Traffic Control Interface) and SUMO are used. In SUMO, a model of the traffic network is created, consisting of signals, roads, intersections, vehicles, pedestrians and bicycles. The model can be configured to reflect the traffic system in real time or to create scenarios that can be used to simulate, test, validate and evaluate the functioning of the traffic network under different conditions and traffic loads. The TraCI module allows access to the running simulation in the SUMO traffic network (using the TCP protocol) and with the application of different programming languages (in this case the Python programming language module), it allows users of the SUMO program to access a set of functions that can be called from a Python script using the TraCI protocol. In this way, information from the SUMO simulation can be controlled and retrieved, such as the speed of the vehicle or the position of the vehicle in the traffic network. The middle layer is a traffic network model that represents a real intersection that is monitored in the physical layer. The third phase is analysis and simulation, which is performed using the Multi-Agent Deep Deterministic Policy Gradient algorithm, which learns optimal strategies for managing and controlling traffic flows in the transportation network model. This is followed by the definition of possible actions, in which the possible interactions between all traffic elements represented in the matrix are listed, so that the agents (traffic signalization) can accelerate the training, learning and decision-making process. Then follows the definition of all authorized actions or decisions available to the agent communicating with the elements of the traffic system so that the traffic management system can influence the control of traffic flows. Finally, there is feedback to the agent, from which it learns from the actions performed and implies an evaluation of the effectiveness of the actions performed in the traffic flow control system. The fourth stage is the decision making based on the initial data from the third stage and the fifth stage is the stage of activating the traffic management system. The evaluation in this paper was carried out using simulation as a quantitative method using synthesized and real data obtained by counting the traffic at the intersections and showing the results in the reduction of harmful gasses in the urban environment. In their paper, Dasgupta and co-authors (2021) present a framework for creating digital twins that manage an adaptive traffic control system at traffic signal intersections to reduce waiting times and improve travel time on the transportation network in an urban environment. They conducted a case study by creating a traffic microsimulation in the SUMO program by creating a virtual image of the traffic network with traffic lights in an urban environment where the data used for the simulation was collected in real time. They proved that with the help of a digital twin it is possible to reduce the waiting time at the intersection and thus influence the satisfaction of road users.

In their paper, Soeffker and co-authors (2022) discuss the problem of stochastic dynamic routing of vehicles in the transportation network, where a number of decisions need to be made on how to route (assign) the vehicles in the network. In their literature review, they came to the realization that there are various methods to integrate the collected information and make a prediction based on the information so that appropriate decisions can be made. They explain this as an example of prescriptive analytics being increasingly used in optimization decision making, and in the paper they examine how it relates to stochastic dynamic routing of vehicles in the transportation network. In their paper, Zirin Wang and co-authors (2022) introduce digital twin technology to develop a cooperative driving system at unsignalized intersections, in which connected vehicles cooperate with each other to cross intersections without emergency braking. They developed an algorithm to plan the vehicles and their movement through intersections without stopping, based on communication between the vehicles. This was achieved using agent-based modeling and simulation in the Unity game engine based on a real map of San Francisco. In their paper, Zhao H. and co-authors (2019) present a simulation at signalized intersections using a vehicle tracking model and a vehicle specific power consumption model to estimate vehicle emissions. They also mention that signalized intersections are areas of high fuel consumption and high pollutant emissions, which is related to the speed of vehicles during the different signal phases. Batista and co-authors (2022) present a study in which they investigate the potential of aggregate traffic models for estimating network-wide emissions. The study is based on a simulation environment using the SUMO simulator, focusing on the calibration of aggregated traffic models and routes to investigate overall network emissions. They concluded that real-time monitoring systems can play an important role in reducing traffic

emissions and pollutants in urban networks, but that reliable monitoring systems are needed. In their paper, Wang Han and co-authors (2023) present an optimization method for urban traffic signal control based on digital twin technology. The method consists of the following parameters: digital roads, traffic lights and vehicles, and the combination of the optimization algorithm to take fuel consumption as the optimization target. The algorithm used for the optimization is based on a genetic algorithm whose main objective is to reduce fuel consumption. They performed the optimization according to the digital twin traffic map with different iteration times and performed several iterative optimizations on the four-phase duration of each intersection. After setting the parameters, they run a simulation in the SUMO simulator, where different traffic scenarios are created. At the end of the investigation, they concluded that the modeling process for a digital twin traffic model is relatively easy to accomplish in practice and that this model, combined with optimization algorithms for controlling digital traffic, has advantages such as ease of operation and high efficiency compared to the original model.

# 4 Limitations of previous research and further research

The researched articles show progress in the development of digital twins in road traffic management and optimization. Digital twins are used for traffic optimization in the following research areas: in sections on highways (motorways), at isolated intersections and connected intersections controlled by traffic signalization, to increase traffic safety at intersections and improve traffic flow in general, to reduce the concentration of exhaust gases at intersections with adaptive traffic signalization and phase timing, to facilitate communication between autonomous vehicles and enable information exchange at intersections without traffic signals, to apply the Unity Game Engine, and to reduce fuel consumption. Articles also describe various methods of how and in what way is it possible to perform traffic optimization at intersections. Few examples of methods are: various algorithms, artificial intelligence, machine learning, deep learning, traffic management based on a multi-agent system with the application of DRL etc. In general, it can be concluded that digital twin technology is researched with different methodologies. Also it can be stated that it is necessary to combine real data and virtual environment with the use of specialized simulation programs. In the field of road traffic optimization, mostly used simulation program is SUMO (Rydzewski and co-authors, 2020). Researchers are using different simulators which can be an obstacle in choosing a simulation program for traffic modeling and simulation. The simulation programs have a well-described interface and support digital twin technology but the interfaces of the platforms are not unified. There are numerous ways how to optimize traffic flow at intersection by observing traffic flow parameters, but it is not defined what is the most effective way to optimize traffic in an urban network and when is the right time to apply the appropriate methodology to achieve the aim of the research. Furthermore, there is a lack of research using real-time data, not considering the few papers in which the authors use real-time data. There is no synergy between data collected by the road measuring devices and the data from the simulation and its updates in real time, so the digital twin technology is not really relevant. Case studies that explain different traffic networks, e.g. highways, closed signalized intersections, connected intersections with traffic lights, etc., are usually performed in a simulation environment where the urban network is configured and modelled with different scenarios. The scenarios are created with various input parameters: traffic light duration, waiting times, traffic flow, traffic density, vehicle speed, etc., and it is observed how these parameters are related to each other. Based on these scenarios, it is possible to optimize the traffic flow at intersections by changing the duration of the traffic lights. Greenhouse gas emissions and pollution caused by the exhaust systems from car engines have been mentioned in many studies, but only as a consequence of the side effects of road traffic that occur in the transportation network depending on traffic congestion at different periods of the day and night or during peak hours. There is a lack of studies dealing with the development of methods and algorithms for the optimization of traffic flow at intersections, where the main trigger for the optimization at intersections in the urban traffic network is the critical amount of greenhouse gas emissions generated by engine exhaust systems. This could be a way to further investigate the existing problems in the optimization of the urban transport network with the help of digital twin technology. This leads to the conclusion that there are possibilities for progress and scientific contributions in research. Regarding to the mentioned limitations, in most of the papers whose main goal is to optimize traffic flow and solve traffic congestion, it exists a gap in the exact explanation of how digital twin models are updated and how the models communicate with the digital twin. Križanić and co-author (2023) therefore state in their paper that it is necessary to mention that a digital twin uses data in real time, updates its performance in real time using data, reads from external devices and sensors, and is a more advanced form of digital twin when machine learning is included in its implementation. The idea of machine learning is that the model of the digital twin is updated according to the amount of data collected during its execution, forming a dataset on which artificial intelligence (AI) can be trained so that it can communicate with the digital twin in terms of adjusting execution based on its behavioural patterns.

# 5 Conclusion

Road traffic is a highly complex system. Scientists from various fields of research, along with engineers and experts, have observed and described traffic and transportation in numerous ways. The traffic and transportation is irreplaceable for the economic development of any country and connects all parts of the world. To achieve the uninterrupted movement of vehicles between start and end points, or between origin and destination, the traffic flows of people and goods (freight) must be successfully controlled. Therefore, traffic flows require appropriate management and must be monitored and optimized in real time. The movement of traffic flows can be predicted in real time using simulation programs, with researchers utilizing data collected from sensors, IoT devices, and communication between vehicles. These simulations enable researchers to predict possible scenarios in the traffic network. However, since traffic is a very dynamic system, it is not always possible to optimize traffic flow based on predictions. Therefore, it is necessary to use different methods to model and optimize traffic flow based on prescriptive analysis. Prescriptive analytics can respond to the current demands of congestion and traffic flows, influencing the optimization of traffic flows by applying appropriate mathematical models and methods. Additionally, it is essential to observe how the speed of vehicles affects traffic flow, traffic density, and greenhouse gas emissions. Therefore, researchers have to choose the most relevant and optimal techniques or methods which will describe and solve potential problem.

# References


Affenzeller M., Bogl m., Fischer L., Sobieczky M., Yang K., Zenisek J. (2022). Prescriptive Analytics: When Data- and Simulation-based Models Interact in a Cooperative Way, *2022 24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC),* Page(s):1– 8.

Batista, S.F.A., Tilg, G., Menendez, M. (2022). Exploring the potential of aggregated traffic models for estimating network-wide emissions. In *Transportation Research part D:Transport and Environment, Volume 109, Article 103354*, https://doi.org/10.1016/j.trd.2022.103354

Brendel, A.B., Zapadka, P., Kolbe, L.M. (2018). Design Science Research in Green IS: Analyzing the Past to Guide Future Research, In *Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, UK*

Csikos, A., Varga, I. (2012). Real-time Modeling and Control Objective Analysis of Motorway Emissions. In *Procedia – Social and Behavioral Sciences, Volume 54,* pp. 1027-1036.

Dasgupta, S., Rahman, M., Lidbe, A. D., Lu W., Jones S. (2023). A Transportation Digital-Twin Approach for Adaptive Traffic Control Systems. *Paper submitted for presentation at the Transportation Research Board 100th Annual Meeting and publication in Transportation Research Record, https://doi.org/10.48550/arXiv.2109.10863*

Delen, D. (2020). Prescriptive Analytics, The final frontier for Evidence-based management and optimal decision making, *Pearson Education, Inc.*

Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport Text with EEA relevance

Feng, H., Lv, H., Lv, Z. (2023) Resilience towarded Digital Twins to improve the adaptability of transportation systems, In *Transportation Research Part A: Policy and Practice, Volume 173.*

Florida Department of Transportation, Traffic analysis Handbook, 2021.

Gagliardi, G., Gallelli, V., Violi, A., Lupia, M., Cario, G. (2024). Optimal Placement of Sensors in Traffic Networks Using Global Search Optimization Techniques Oriented towords Traffic Flow Estimation and Pollutant Emission Evaluation, *Research on Sustainable Transportation and Urban Traffic—2nd Edition, in MDPI, Sustainability volume 16 (9), 3530, https://doi.org/10.3390/su16093530*

Kamal, H., Yanez W, Hassan S., Sobhy D. (2024). Digital-Twin-Based Deep Reinforcement Learning Approach for Adaptive Traffic Signal Control, *IEEE Internet of Things Journal, Volume 11, Issue: 12,* pages: 21946 – 21953.

Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. In *IFAC PapersOnLine 51-11,* pp. 1016-1022.

Križanić, S., Vrček, N. (2023). Business process modeling and digital twins: a literature review, *CECIIS 2023., Central European Conference on Information and Intelligent Systems,* p. 145-151.

Kušić, K., Schumann, R., Ivanjko, E. (2022). Building a Motorway Digital Twin in SUMO: Real-Time

Simulation of Continuous Data Stream from Traffic Counters, International Symposium Electronics In *Marine, Volume 2022, 64th International Symposium ELMAR, ELMAR 2022,* pages: 71-76. https://doi:10.1109/ELMAR55880.2022.9899796

Kušić, K., Schumann, R., Ivanjko, E. (2023). A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics, *Advanced Engineering Informatics, Volume 55, Issue C,* https://doi.org/10.1016/j.aei.2022.101858

Lakshmanan, S., Sornam, M., Flores, J. (2020). Demystifying Prescriptive Analytics Frameworks and Techniques, *International Journal of Innovative Technology and Exploring Engineering (IJITEE*), Volume-9 Issue-6, pp: 1422-1427.

Martínez, Víctor, M. G.; Ribeiro, Moisés R. R.; Campelo, Divanilson R. (2022). Intelligent road intersections: A Case for Digital Twins. In *Workshop brasileiro de cidades inteligentes (wbci), 3., Niterói. anais [...]. porto alegre: sociedade brasileira de computação,* p. 151-158.

Menon, D., Anand, B., Chowdhary, C. L. (2023). Digital Twin: Exploring the Intersection of Virtual and Physical Worlds. *IEEE Access, DOI: 10.1109/ACCESS.2023.3294985,* pages 75152-75172.

Ortuzar, J. de D., Willumsen, L. G. (2011). Modeling Transport, *Fourth Edition, 2011 John Wiley & Sons, Ltd,.*

Rydzewski, A., Czarnul, P. (2020). Recent advances in traffic optimisation: systematic literature review of modern models, methods and algorithms, *IET Intelligent Transport System, Volume 14(13),* , *https://doi.org/10.1049/iet-its.2020.0328,* pages: 1740–1758.

Sami Irfan, M., Dasgupta, S. (2024). Rahman, M. Toward Transportation Digital Twin Systems for Traffic Safety and Mobility: A Review, *IEEE Internet of Things Journal 11(14),* pp. 24581-24603.

Shamlitsky, Y., Aleksey O., Morozov, E., Strekaleva, T. (2024). Using digital twins to manage traffic flows, *E3S Web Conferences 471, XIV International Conference on Transport Infrastructure: Territory Development and sustainability, Article 04027, https://doi.org/10.1051/e3sconf/202447104027*

Soeffker, N., Ulmer, M. W., Mattfeld, D. C. (2022). Stochastic dynamic vehicle routing in the light of prescriptive analytics: A review, *European Journal of Operational Research 298,* p. 801-820,

Transportation Research Board; 5th edition. HCM 2010, Highway Capacity Manual, Transportation Research Board, Washington, D.C., 2010.

Wang, H., Guan, Y., Zhao, L., Shi, J., Li, S., So, W., Ma, J., Song, Q., Zhang, Y., Liu, X. (2023). Optimization Method for Urban Traffic Signal Control Based on Digital Twin Technology, *2023 15th International Conference on Communication Software and Networks (ICCSN),* Page(s):444 – 448.

Wang, Z., Han, K., Tiwari, P. (2022). Digital Twin-Assisted Cooperative Driving at Non-Signalized Intersections. *IEEE Transactions on Intelligent Vehicles , Volume: 7, Issue: 2,* pages 198-209.

Wu H., Ji P., Ma H., Xing L. A (2023). Comprehensive Review of Digital Twin from the Perspective of Total Process: Data, Models, Networks and Applications, *Sensors 2023, 23(19), 8306, https://doi.org/10.3390/s23198306*

Zhao, H., He, R., Jia, X. (2019). Estimation and Analysis of Vehicle Exhaust Emissions at Signalized Intersections Using a Car-Following Model. In *Sustainability, MDPI, vol 11(14),* pp. 1-25.

Zhao, H.-X., He, R.-C. (2021). Modeling of vehicle $CO_2$ emissions and signal timing analysis at a signalized intersection considering fuel vehicles and electric vehicles. In *European Transport Research Review 13, Article number 5.*

# Agile Software Development Life Cycle Model for Internet of Things

**Goran Filinić**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Hrvatska
gfilinic22@student.foi.hr

**Ivan Magdalenić**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Hrvatska
ivan.magdalenić@foi.unizg.hr

**Abstract.** *In the fast developing landscape of IoT technology, the development of effective software solutions faces unique challenges due to the integration of diverse hardware components and the need for continuous adaptation to evolving requirements. Traditional software development lifecycle (SDLC) models often struggle to accommodate the iterative nature and rapid pace of IoT projects. This paper proposes an Agile SDLC model tailored specifically for IoT systems, integrating principles from Agile methodologies, Rapid Prototyping, and Model-Driven Development. The model emphasizes continuous iteration, comprehensive documentation, and adaptive design, aiming to enhance development efficiency, mitigate risks, and accelerate time-to-market for IoT solutions. By addressing the complexities of IoT development through structured yet flexible phases—ranging from requirement gathering and design to deployment and maintenance—the proposed model seeks to provide a robust framework capable of meeting the dynamic demands of IoT environments. This model aims to contribute to the advancement and adoption of effective SDLC practices in IoT development.*

**Keywords.** Agile SDLC, IoT systems, Rapid Prototyping, Model-Driven Development, software development lifecycle, embedded systems, embedded programming

## 1 Introduction

The accelerated advancement of the Internet of Things (IoT) technology has brought about significant transformations in various industries, enhancing efficiency, connectivity, and automation. Despite these advancements, there remains a notable gap in structured guidelines for the development of IoT projects (Fahmideh et al., 2022). This absence of a comprehensive framework poses challenges for developers and organizations aiming to implement IoT solutions efficiently and effectively.

Because IoT projects inherently involve unique challenges with the perspective of integrating physical devices with software components, they introduce a layer of complexity not present in traditional software development. Each IoT project is unique due to the specific hardware requirements and the need for seamless interaction between devices and software Singh et al. (2014), Guerrero-Ulloa et al. (2019), Dado et al. (2015), Madakam et al. (2015). This added complexity necessitates a more flexible and adaptive approach to the SDLC, especially when considering the integration of embedded systems which play a critical role in IoT infrastructure.

To bridge this gap, an enhancement of the classic SDLC model is proposed by integrating present development practices such as Agile methodologies Srivastava et al. (2017), Schwaber and Sutherland (2011), Salo and Abrahamsson (2008), Rapid Prototyping Lantz (1986), Fern and Donaldson (2003), and Model-Driven Development Cabot (2012). These modern approaches not only streamline the development process but also foster greater flexibility, iterative improvements, and a higher degree of customization, which are critical in IoT project management.

The goal of this research is to synthesize these insights and propose a refined SDLC model tailored specifically for IoT projects. By leveraging modern practices, this model aims to optimize the development process, ensuring that IoT solutions are not only innovative but also robust, scalable, and aligned with the fast-paced evolution of technology. Through this endeavor, we hope to contribute to the establishment of a more structured and efficient framework for IoT development, ultimately driving further advancements in this transformative field.

The rest of the paper is structured as follows. Section 2 presents a review of relevant related work in IoT development methodologies. Section 3 introduces the proposed Agile SDLC model for IoT, highlighting its fundamental components and principles. Section 4 provides a detailed discussion on the potential benefits and challenges of the model. Finally, Section 5 concludes the paper by summarizing key findings, discussing implications for future research, and outlining potential avenues for further exploration.

## 2 Related Work

The current landscape of IoT systems development has been extensively explored by Fahmideh et al. (2022). Their mixed-methods study identified a significant knowledge gap in the development lifecycle for IoT projects. By synthesizing existing literature, they derived and validated a conceptual framework outlining 27 crucial tasks for integrating IoT systems into development processes Fahmideh et al. (2022). This framework, validated through a global survey of 127 IoT practitioners, forms the cornerstone of this paper's proposed model, which aims to establish effective processes documented in the literature.

Gupta and Gayathri (2022) closely examined the Software Development Lifecycle (SDLC), as well as emphasizing the critical role of early testing Jat and Sharma (2017). Their studies highlighted that early initiation of testing prevents bugs, reduces costs, and accurately determines requirements, ultimately improving software reliability and performance (Gupta and Gayathri (2022), Jat and Sharma (2017)). This underscores the importance of testing in the SDLC and its role in enhancing software development outcomes.

In safety-critical domains like medical device software development, Klespitz et al. (2015) , Dayo et al. (2022) emphasized the necessity of robust lifecycle management systems. Their case study highlighted the need for stringent documentation and objective measures to ensure effective and mature product development. This is particularly crucial for meeting safety standards and accelerating system implementation in medical software development.

Agile methodologies applied to IoT systems (IoTS) development were reviewed by Guerrero-Ulloa et al. (2023). Analyzing 60 documents from an initial 4,303, they found that 42.1% of IoTS developments used Scrum exclusively, with others combining it with methodologies like XP, Kanban, and Rapid Prototyping. The study highlighted the prominence of model-based approaches such as Model-Driven Engineering (MDE) and Model-Driven Development (MDD) in addressing IoTS technology heterogeneity Cabot (2012). However, gaps in requirement elicitation, maintenance, and withdrawal phases were noted, emphasizing the need for comprehensive methodologies.

Alfawair (2022) proposed the IoTSDLC model, which adapts the traditional SDLC to address IoT-specific challenges like heterogeneity and complex integration requirements. Despite its emphasis on detailed documentation and system specification, the model's linear approach lacks flexibility for agile development. Another example can be looked at Yang and Bi (2014). This rigidity and absence of agile integration guidelines limit its applicability in dynamic IoT environments where rapid prototyping and continuous iteration are essential.

These studies underscore the diverse challenges and methodologies in IoT systems development. They highlight the need for structured SDLC tailored to IoT projects, the critical role of early testing, robust management systems, and the application of agile methodologies. Integrating these insights into a unified SDLC model is imperative to meet the dynamic demands of current IoT environments, ensuring comprehensive support for developers and stakeholders. The proposed unified agile SDLC model integrates these insights to meet the IoT environments, ensuring comprehensive support for developers and stakeholders, and enhancing the overall effectiveness and adaptability of IoT system development.

## 3 Agile SDLC Model for IoT

Figure 1 illustrates the proposed Agile SDLC model tailored specifically for IoT projects. Unlike traditional Agile SDLC, which typically focuses on rapid iterative cycles for software delivery Yang and Bi (2014), this model is designed to address the extended lifecycle and complex integration challenges characteristic of IoT systems. This include device heterogeneity and system reliability Alberternst et al. (2021), Lakhan et al. (2022). Primary divergences from traditional Agile SDLC include:

1. **Long-term lifecycle management**: The model considers the extended lifecycle typical of IoT projects, which require careful planning not only for development but also for ongoing maintenance and adaptability. It incorporates continuous optimization and grooming to minimize future interventions, ensuring sustained performance over time.

2. **Grooming phase**: A dedicated phase called grooming is included to refine and optimize the system's architecture and components, ensuring robustness and scalability. This phase addresses the specific needs of IoT projects, where ongoing system refinement is critical.

3. **Maintenance and Support**: This phase introduces mechanisms for both small updates and major changes, providing structured support for the long-term maintenance of IoT systems:

   - **Small updates**: Incremental updates are applied to address minor bugs and optimize performance, ensuring compatibility with evolving technologies without disrupting stability.
   - **Major changes**: Significant upgrades or modifications are planned, tested, and deployed in a structured manner to achieve substantial improvements while minimizing disruption.

The model is structured into several distinct phases, each crucial for IoT project success. The **Requirement Phase** gathers specific requirements, considering

the diverse interactions within IoT systems. The **Design Phase** translates these requirements into a flexible design adaptable to iterative development. The **Rapid Prototyping Phase** involves continuous cycles of development and testing, particularly suitable for addressing hardware-related challenges. The **Grooming Phase** refines and optimizes the system, ensuring robustness before deployment. The **Deployment Phase** ensures seamless integration and operation of IoT components. Finally, the **Maintenance and Support Phase** manages both minor updates and major changes, ensuring the system remains adaptable and responsive to new requirements and technological advancements.

The proposed model serves as a flexible guideline rather than a rigid framework, accommodating the varied requirements of IoT projects. Iteration lengths and phase transitions are intentionally not fixed, as they depend on the project's scope, the involved stakeholders, and specific planning needs. In early phases, the process can either progress quickly or extend over several years. During the Rapid Prototyping phase, the development approach can vary—from Scrum to rapid prototyping or custom methodologies—allowing teams to adapt based on their preferences. Similarly, the Grooming and Maintenance phases are highly adaptable, recognizing that the timeline for product retirement is unpredictable Yang and Bi (2014).

The **Requirement phase** is essential for gathering detailed and specific requirements that account for the diverse components and interactions within IoT systems. This phase involves collecting functional and non-functional requirements, assessing project feasibility, identifying relevant stakeholders, and developing conceptual models International Organization for Standardization (2019). Leveraging Model-Driven Principles enhances the clarity and precision of requirements, ensuring that all system aspects are thoroughly considered. MDD can be particularly useful in IoT projects by enabling systematic transformations from high-level conceptual models to more detailed designs, facilitating ongoing refinement throughout the development process. This approach ensures that the system architecture remains adaptable to evolving requirements Cabot (2012).

In contrast, Model-Driven Architecture (MDA), a specific framework within MDD, is applied when there is a need to separate platform-independent models (PIMs) from platform-specific models (PSMs) Group) (2024). MDA is especially relevant in IoT environments where diverse hardware and software platforms must be supported. By using MDA, IoT systems can be designed in a way that allows for seamless deployment across various platforms, while ensuring the integrity and functionality of the system are preserved Kautz et al. (2018). Thus, while MDD provides the overall methodology for refining and evolving models, MDA should be employed in scenarios where cross-platform deployment and compatibility are major considerations.

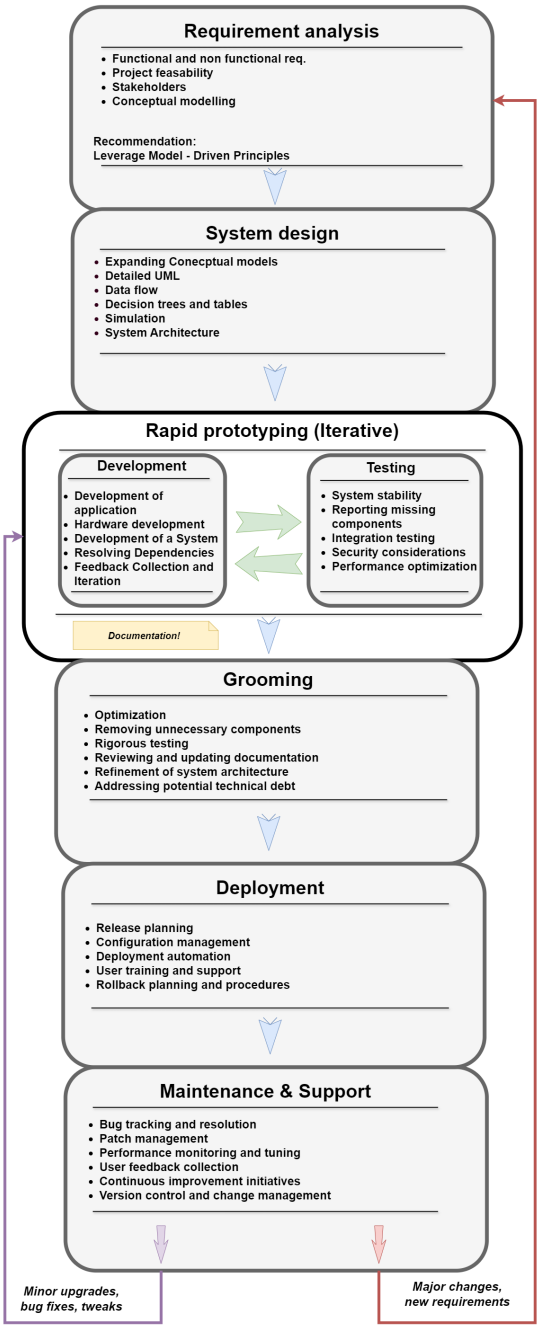



**Figure 1:** Proposed new model for agile IoT SDLC

In the **Design phase**, the focus shifts to translating the gathered requirements into a robust and flexible design. This phase includes expanding conceptual models into detailed UML diagrams, mapping data flows, constructing decision trees, and conducting simulations Brambilla et al. (2017). The objective is to develop a comprehensive system architecture that addresses all identified requirements while remaining adaptable to iterative enhancements.

The **Rapid prototyping phase** embodies the core principles of agile development, with continuous cy-

cles of development, testing, and coding. This iterative approach allows for rapid prototyping and immediate resolution of issues, enabling the system to evolve in alignment with user needs and technological advancements Guerrero-Ulloa et al. (2023), Zelfia et al. (2022).

The **Grooming phase** focuses on refining and optimizing the system, essential for IoT projects. This phase involves rigorous testing, optimizing performance, and updating documentation to reflect the latest system state. The Grooming Phase enhances overall system stability and performance, aligning it with the dynamic requirements of IoT environments.

The **Deployment phase** ensures smooth integration and deployment of IoT components, facilitating a seamless transition from development to operation. Deployment strategies are designed to minimize disruption and ensure the system's high performance and reliability Yang and Bi (2014).

Finally, the **Maintenance and support phase** provides mechanisms for both minor upgrades and major changes, ensuring the system remains adaptable and responsive to new requirements and technological advancements. Minor upgrades are handled through rapid development and testing, while significant changes trigger a comprehensive reassessment and planning process. This phase includes bug tracking, patch management, performance monitoring, and continuous improvement initiatives, ensuring the system remains robust throughout its lifecycle.

By integrating these phases, the proposed Agile SDLC model for IoT offers a structured yet flexible framework tailored to the unique challenges of IoT system development, emphasizing continuous iteration, comprehensive documentation and adaptive design.

# 4 Discussion

The proposed Agile SDLC model for IoT systems represents a framework developed from practical experience in IoT project environments, particularly in smart home and smart city domains where embedded systems play an important role. It requires thorough review and validation by peers and industry experts. One of the challenges in evaluating the effectiveness of this model lies in the difficulty of quantifying its benefits. While a comparative analysis using time as a metric—measuring the duration of similar projects with and without this model—could offer insights, it may not fully capture the model's advantages in adaptability and responsiveness.

Traditional SDLC models often struggle with the dynamic and iterative nature of IoT development. By integrating Agile methodologies, this model aims to enhance the ability to pivot, make changes, and adapt to evolving project requirements, which is a vital requirement in IoT-related projects. This approach seeks to offer a more flexible and responsive framework.

One significant potential benefit of this model is its focus on maintenance. It is designed to reduce the frequency and severity of interventions required post-deployment. Ideally, this model will lead to fewer maintenance incidents, and when issues do arise, they will be quicker and easier to address, particularly concerning embedded contexts. This could result in a more stable and reliable system over the long term.

Moreover, the proposed model integrates modern practices from IoT development, such as Model-Driven Architecture (MDA) and Agile methodologies, into a streamlined SDLC framework. This hybrid approach is intended to provide a balance between the structured processes of traditional SDLC and the flexibility needed for IoT projects. By doing so, it aims to be adaptable across various project types and sizes, offering a versatile tool for developers.

However, it is important to note that this model is likely to work most effectively with small development teams, as it aligns with the principles of Agile and Scrum practices. Agile methodologies are typically best suited for smaller teams due to their emphasis on close collaboration, rapid iteration, and flexibility. The model's effectiveness and scalability for larger teams remain uncertain, as it was not specifically designed for that context.

While the model shows promise, it is currently conceptual and has yet to be empirically validated through extensive real-world applications. The model's success hinges on its application in pilot projects, which will be crucial in identifying its strengths and weaknesses. These pilot implementations will provide valuable feedback, helping to refine the model and ensure it can address practical challenges and deliver tangible benefits.

Additionally, the model's effectiveness in different industry contexts needs to be assessed. IoT projects vary widely in scope and complexity, from small-scale consumer devices to large-scale industrial systems. The model's adaptability and scalability across these diverse contexts will be a main factor in its broader adoption. Future research should focus on empirical studies and case studies to evaluate how well the model performs in these diverse settings, thereby solidifying its practical applicability.

# 5 Conclusion

The proposed Agile SDLC model for IoT systems represents a significant advancement in addressing the complexities of IoT development. By integrating Agile methodologies, Rapid Prototyping, and Model-Driven Development, the model offers a flexible and iterative approach essential for managing diverse hardware integrations and complex software interactions inherent in IoT projects. Its emphasis on continuous iteration and collaboration enhances development efficiency, minimizes risks, and accelerates time-to-market for IoT so-

lutions. Moreover, the model's structured yet adaptable framework ensures responsiveness to changing requirements, technological advancements, and user feedback throughout the development lifecycle. By focusing on maintenance and support phases, it aims to reduce post-deployment issues and enhance system reliability, necessary for sustaining IoT solutions over the long term. The iterative development cycles facilitate rapid prototyping and validation, enabling early detection and resolution of issues while allowing for incremental enhancements based on real-world performance. Looking forward, further validation through empirical studies and case studies across diverse IoT applications will be essential to refine and scale the model for broader adoption in different industry contexts where embedded systems play a pivotal role. The proposed Agile SDLC model stands poised to drive innovation and efficiency in IoT development, contributing to the advancement of transformative IoT solutions in a rapidly evolving technological landscape.

# References


Alberternst, S., Anisimov, A., Antakli, A., Duppe, B., Hoffmann, H., Meiser, M., Muaz, M., and Spieldenner, D. (2021). Zinnikus, i. orchestrating heterogeneous devices and AI services as virtual sensors for secure Cloud-Based IoT applications. *Sensors*, 21.

Alfawair, M. (2022). Internet-of-things: A system development life cycle (sdlc). *J. Theoretical Appl. Inform. Technol.*, 100(5):1643–1653.

Brambilla, M., Umuhoza, E., and Acerbis, R. (2017). Model-driven development of user interfaces for IoT systems via domain-specific components and patterns. *Journal of Internet Services and Applications*, 8(1):14.

Cabot, J. (2012). Clarifying concepts: Mbe vs mde vs mdd vs mda. *Post at Modeling Languages, http://modeling-languages.com/clarifying-concepts-mbe-vs-mde-vs-mdd-vs-mda/*, 86. Accessed: 2024-8-12.

Dado, M., Janota, A., and Spalek, J. (2015). Challenges and unwanted features of the smarter cities development. In *Internet of Things. IoT Infrastructures*, Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 3–8. Springer International Publishing, Cham.

Dayo, Z. A., Aamir, M., Dayo, S. A., Khoso, I. A., Soothar, P., Sahito, F., Zheng, T., Hu, Z., and Guan, Y. (2022). A novel compact broadband and radiation efficient antenna design for medical IoT healthcare system. *Math. Biosci. Eng.*, 19(4):3909–3927.

Fahmideh, M., Ahmad, A., Behnaz, A., Grundy, J., and Susilo, W. (2022). Software engineering for internet of things: The practitioners' perspective. *IEEE Transactions on Software Engineering*, 48(8):2857–2878.

Fern, D. A. and Donaldson, S. E. (2003). Tri-Cycle: a prototype methodology for advanced software development. In *[1989] Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences. Volume II: Software Track*. IEEE Comput. Soc. Press.

Group), O. O. M. (2024). MDA specifications. `http://www.omg.org/mda/specs.htm`. Accessed: 2024-8-12.

Guerrero-Ulloa, G., Rodríguez-Domínguez, C., and Hornos, M. J. (2019). IoT-based system to help care for dependent elderly. In *Communications in Computer and Information Science*, Communications in computer and information science, pages 41–55. Springer International Publishing, Cham.

Guerrero-Ulloa, G., Rodríguez-Domínguez, C., and Hornos, M. J. (2023). Agile methodologies applied to the development of internet of things (iot)-based systems: A review. *Sensors*, 23(2).

Gupta, S. and Gayathri, N. (2022). Study of the software development life cycle and the function of testing. In *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*, pages 1270–1275.

International Organization for Standardization (2019). Iso/iec/ieee international standard – systems and software engineering - content of life-cycle information items (documentation). *ISO/IEC/IEEE 15289:2019(E)*, pages 1–86.

Jat, S. and Sharma, P. (2017). Analysis of different software testing techniques. *International Journal of Scientific Research in Computer Science and Engineering*, 5(2):77–80.

Kautz, O., Roth, A., and Rumpe, B. (2018). Achievements, failures, and the future of model-based software engineering.

Klespitz, J., Bíró, M., and Kovács, L. (2015). Aspects of improvement of software development lifecycle management. In *2015 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 323–327.

Lakhan, A., Mohammed, M. A., Abdulkareem, K. H., Jaber, M. M., Nedoma, J., Martinek, R., and Zmij, P. (2022). Delay optimal schemes for internet of things applications in heterogeneous edge cloud computing networks. *Sensors (Basel)*, 22(16):5937.

Lantz, K. E. (1986). *The prototyping methodology*. Prentice-Hall, Inc., USA.

Madakam, S., Ramaswamy, R., and Tripathi, S. (2015). Internet of things (IoT): A literature review. *J. Comput. Commun.*, 03(05):164–173.

Salo, O. and Abrahamsson, P. (2008). Agile methods in european embedded software development organisations: a survey on the actual use and usefulness of extreme programming and scrum. *IET Softw.*, 2(1):58.

Schwaber, K. and Sutherland, J. (2011). The scrum guide. *Acrum Allience*, 21.

Singh, D., Tripathi, G., and Jara, A. J. (2014). A survey of Internet-of-Things: Future vision, architecture, challenges and services. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*. IEEE.

Srivastava, A., Bhardwaj, S., and Saraswat, S. (2017). SCRUM model for agile methodology. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*. IEEE.

Yang, L. Q. and Bi, Y. Y. (2014). Internet of things technology implementation by applying SDLC model: The intelligent storage management system. *Appl. Mech. Mater.*, 556-562:5385–5390.

Zelfia, H., Simanungkalit, T., and Raharjo, T. (2022). Comparison of scrum maturity between internal and external software development: A case study at one of the state-owned banks in indonesia. In *2022 1st International Conference on Information System Information Technology (ICISIT)*.

# Digital Transformation and Corporate Culture

Maja Fatiga, Andrej Ignjatić
**Digitalization of Monitoring Side Effects Before, During and After Covid-19 Vaccines – Analysis of the Impact of the Pandemic on Reporting Side Effects in Republic of Croatia**

Robertina Zdjelar, Nikolina Žajdela Hrustek
**Public Policymakers' Attitudes on Planning and Implementing Measures for 54+ Population e-Inclusion**

Mirona Ana Maria Popescu, Andreea Barbu, Larisa Maria Fecioru, Elena-Teodora Miron
**An Analysis on Digital Public Services in Romania**

# Digitalisation of Monitoring Side Effects Before, During and After COVID-19 Vaccines – Analysis of the Impact of the Pandemic on Reporting Side Effects in Republic of Croatia

**Maja Fatiga, Andrej Ignjatić**

University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, Varaždin, Croatia
Agency for Medicinal Products and Medical Devices of Croatia
Ksaverska cesta 4, Zagreb, Croatia
{mfatiga, aignjatic}@halmed.hr

***Abstract**. This paper presents the medicinal product's side effect reporting information system (OPeN) used by the Croatian Agency for Medicinal Products and Medical Devices. The paper aims to show the trend the trend in reporting suspected side effects before, during, and after the COVID-19 pandemic, as well as before and after the digitalisation of side effects reporting with the introduction of the OPeN. In this research collected data of reported side effects from 2014 to 2023 is analysed. This paper intends to show the process of reporting suspicions of medicinal product side effect reactions linked to the digitalisation of public services.*

**Keywords.** Digitalisation, Information System, OPeN, Side effect, COVID-19

## 1 Introduction and Literature Review

Public administration is under strong pressure related to improving its service delivery methods (Vestues et al., 2021). Current trends in public administration development emphasise collaboration between the public administration and the users of its services, focusing on the service users themselves (Fossheim & Lund-Tønnesen, 2023; Greve et al., 2016). Users can participate in public administration through the process of digitalisation, which has notably impacted the public sector, leading to substantial changes in the workings of public administration and to enhancing access to its services. The digitalisation process has accelerated under special conditions, such as the COVID-19 pandemic, which sped up the digitalisation of certain segments of public administration, creating opportunities for further development of the public administration system (Đanić Čeko & Guštin, 2022).

The process of digitalisation substantially influenced the healthcare sector, especially during COVID-19, where different digital technologies, such as telemedicine, various mobile health applications, and electronic health records were integrated, which led to more direct and improved healthcare (Balenović et al., 2022; Eşiyok et al., 2023; Moulay, 2023). The digitalisation of healthcare has also impacted the reporting side effects of medicinal product (somewhere called "adverse drug reaction" or "adverse effect"). Adverse drug effects were not only reported to the competent adverse effect reporting bodies. Some studies have used social networks to detect adverse effects available in the market (Farooq et al., 2021). Additionally, websites can impact the knowledge of pharmacovigilance by independently collecting patients reports to uncover harmful consequences caused by medicinal products, which clinical trials and national pharmacovigilance systems may not have recognised (Lentacker, 2024). Furthermore, some studies have utilised machine learning to extract extensive data for identifying adverse drug effects (Dubey et al., 2021).

Pharmacovigilance describes the process of monitoring and evaluating adverse effects and is an essential component of medicinal product regulation and public health programs. The impact of reported adverse effects contributes to the early detection of new safety risks associated with medicinal product use. In the Republic of Croatia, according to the Medicinal Products Act and the Ordinance on Pharmacovigilance (HALMED, 2024b), healthcare professionals are required to report adverse drug reactions. An adverse drug reaction is defined as "an appreciably harmful or unpleasant reaction, resulting from an intervention

related to the use of a drug, which predicts hazard from future administration and warrants prevention or specific treatment, or alteration of the dosage regimen, or withdrawal of the product" (Edwards & Aronson, 2000). Sometimes this reaction may be characterised as a mild occurrence, and sometimes the consequences can be very dangerous for the patient's life (Kamble et al., 2023). Adverse drug reactions happen daily and affect patients' quality of life (Thomas, 2020). The adverse drug effects continue to be of interest to professional (pharmaceutical) and academic circles, considering how they begin in the medicinal product development phase (whether the drug is developed with a new or already known active substance), where possible adverse drug effects are initially explored (Elzagallaai, 2012). Furthermore, during the monitoring of drugs, new methods and ways of predicting possible adverse effects are defined, using the latest technologies, to reduce the occurrence of adverse effects or mitigate an adverse effect that may occur using a specific drug (Gawich & Alfonse, 2022; J. Wang et al., 2020; Mohsen et al., 2021).

The legal obligation to process reports of suspected adverse effects when using drugs in the Republic of Croatia is under the responsibility of the Agency for Medicinal Products and Medical Devices (HALMED, 2024b, 2024a). To fulfil its legal obligation more successfully, the Agency for Medicinal Products and Medical Devices began developing the OPeN information system in 2017, which is used for reporting adverse effects on medicinal products. The first significant upgrade to the OPeN information system occurred during the COVID-19 pandemic.

This paper aims to show the trend in reporting suspected adverse effects of medicinal products before, during, and after the COVID-19 pandemic, as well as before and after the digitalisation of adverse effect reporting. To our knowledge, this is the first paper that illustrates the process of reporting suspected adverse effects of medicinal products after significant changes in OPeN related to the digitalisation of public services. The paper is divided as follows. The first section presents the paper's goal and provides an overview of the area and literature. The second section describes the OPeN information system and the data collection method. The third section graphically presents the collected results. The paper concludes with a discussion, conclusion, and suggestions for upgrading the OPeN information system.

# 2 Information System for Reporting Side-Effects – OpeN

Until 2018, HALMED collected adverse drug reactions in various ways, most commonly using mail and email. In 2018, HALMED developed the OPeN information system for collecting adverse drug reactions to digitalise its pharmacovigilance-related services. The OPeN system is divided into two interconnected parts, as graphically illustrated in Fig. 1. The development of the OPeN information system was described by Rajh, Sudić, and Gvozdanović (Rajh, Gvozdanović, et al., 2018; Rajh, Sudić, et al., 2018).

The internal part of the system enables HALMED employees to streamline and facilitate business processes related to pharmacovigilance, which includes reporting suspected adverse reactions during drug use, recording and evaluating reported adverse reactions, and updating all additional information related to reported adverse reactions. In the internal part, employees are provided with the distribution of safety notices and the creation of various questionnaires within the educational module. This module represents HALMED's platform for educating healthcare professionals about activities undertaken to ensure the safe use of medicinal products and medical products, developed during the COVID-19 pandemic.

The external part of the OPeN system is intended for all external users of HALMED services, whether they are healthcare professionals (pharmacists or physicians), industry employees (i.e., marketing authorisation holders), or patients (or their legal representatives). External users can use the OPeN information system to report suspected adverse drug reactions and harmful events during the use of medical products. Additionally, healthcare professionals can use the functionalities of the OPeN information system to access the educational module, where they can access specific notices or respond to questions in the questionnaires. Healthcare professionals receive a predefined number of points for all activities within the OPeN information system, which serves to extend their licence at the respective chambers.

For the OPeN information system to successfully cover HALMED's business processes, the system is connected via web services and exchanges data with other information systems within and outside HALMED. Within HALMED, OPeN is connected to the information system for office operations and case management (Centrix), the document management system (DAIS), the National Medicinal Product Registry (NRL), and the OLIMP information system, in which the assessment of reports of harmful events on medical products is carried out. The OPeN information system is bidirectionally connected via web service to the VigiFlow information system, which manages reports of adverse drug effects and includes standardised medical terminology (Uppsala Monitoring Centre, 2024).

For the purposes of this research, secondary data were accessed and analysed, initially collected by HALMED within its legal authority and available in its business records. Data were collected from 2014 to 2023, such that from 2014 to 2018, open cases in the office operations system were summarised, and from 2019 to 2023, the number of open cases in the OPeN information system was summarised.

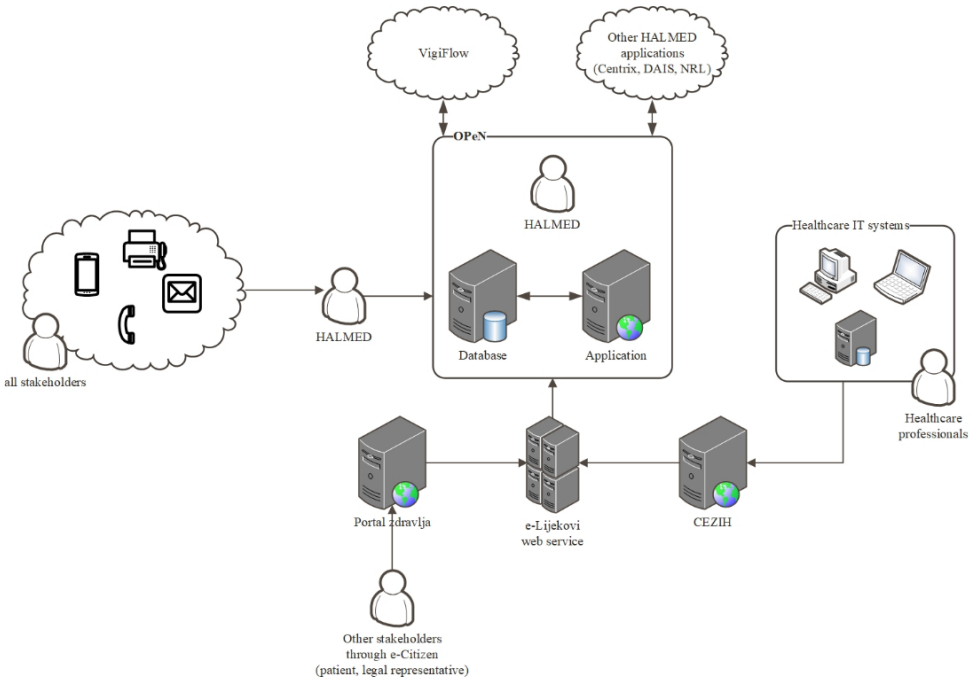



**Figure 1.** OPeN's structure scheme

# 3 Results

Results are graphically shown with Fig. 2 and Fig. 3. Fig. 2 graphically shows the number of reported adverse drug reactions from 2014 to 2023. From Fig. 2, it is evident that the highest number of adverse drug reactions was reported in 2021 (8491), while the fewest were reported in 2023.

Fig. 3 graphically represents the number monthly reported adverse reactions during the COVID-19 pandemic, between 2020 and 2023. In Fig. 3, the line represents reports of adverse reactions unrelated to COVID-19 (C-19), while the columns show reports of adverse reactions related to the COVID-19 pandemic. Observing this period, it is noted that the highest number of reports on suspected adverse reactions unrelated to COVID-19 was in March 2021, and the highest number of reports on suspected adverse reactions related to COVID-19 was in May 2021.

Fig. 4 shows the number of reported or analysed side effects through OPeN by reporter role between 2018 and 2023. The results in Fig. 4 shows that patients during COVID-19 reported far more side effects than any other reporter role.

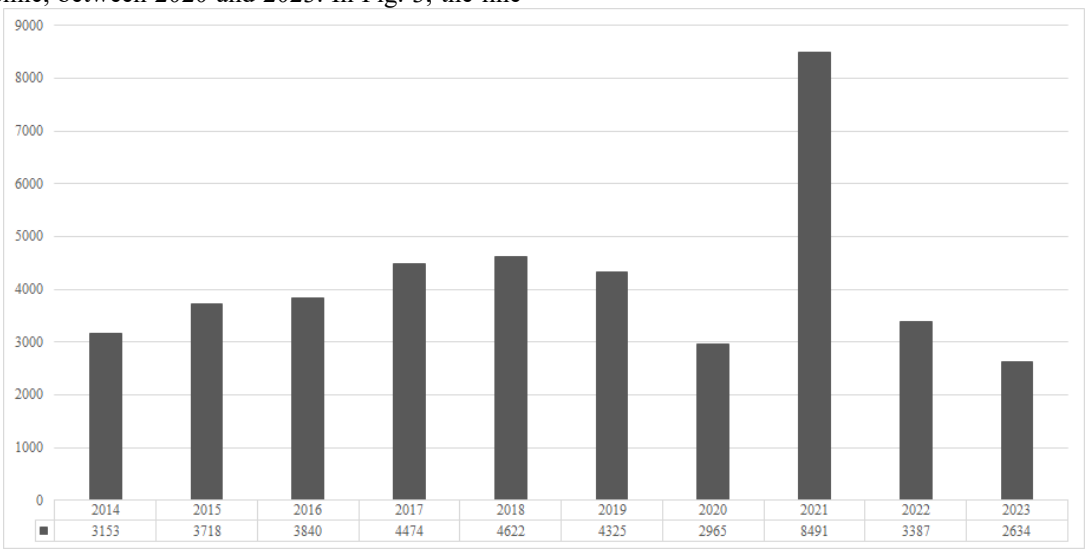



**Figure 2.** Number of reported adverse drug reactions

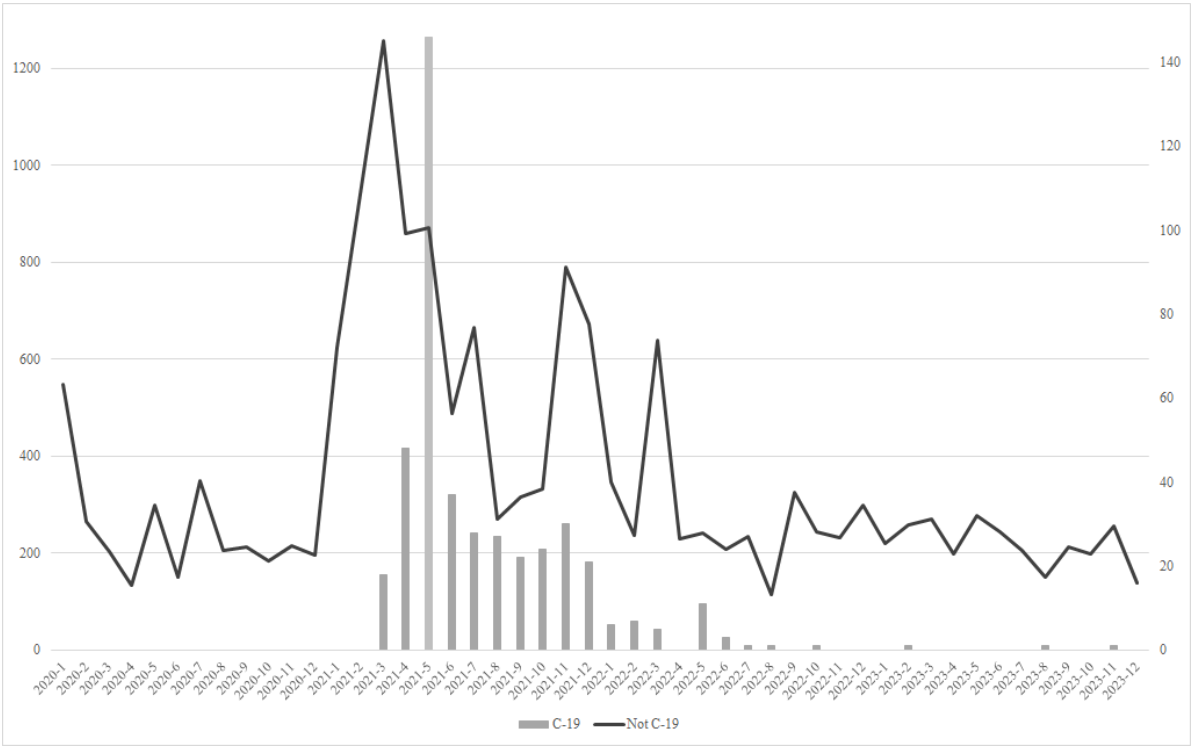



**Figure 3.** Number of reported adverse reactions by month during the COVID-19 (C-19) pandemic

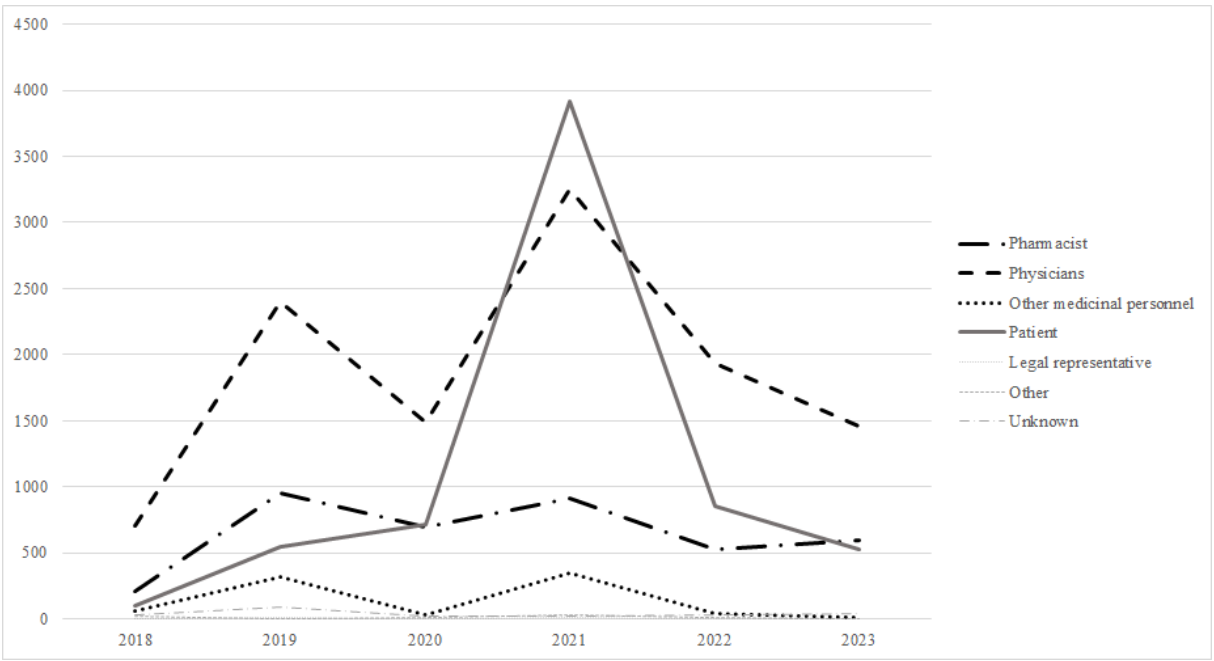



**Figure 4.** The trend of reporting adverse drug reactions by the reporter role

## 4 Discussion and Conclusion

Until 2018 in Croatia, adverse drug reactions could only be reported through traditional means—either by sending a paper form or an email to HALMED, the National Competent Authority (NCA). In 2018, HALMED launched the OPeN for electronic reporting of suspected adverse drug reactions. In addition to OPeN, it became possible to submit reports through a mobile application or by filling out a form on HALMED's website. This digitised the entire reporting

process, making access easier for all stakeholders in the pharmaceutical market. Additionally, the process of reporting adverse drug reactions has been promoted in various ways. Moreover, during the COVID-19 pandemic, HALMED significantly expanded the functionalities of the OPeN to adapt to pandemic conditions.

Discussion of Fig. 2 and Fig. 3. Fig. 2 shows that over the last decade, from 2014 to 2023, the number of reported adverse reactions was roughly consistent, with no significant fluctuations (with the exception of 2021). The highest number of reports was during 2021 due to the start of vaccination against the virus causing COVID-19 in Croatia. This is also evident from Fig. 3, which shows the peak of reports on COVID-19 vaccines in May 2021, when vaccination became available to all interested citizens, whereas previously only the most vulnerable groups were eligible. However, it is interesting to note that in 2023, the total number of reported adverse reactions decreased, despite an easy trend of reduction in reports since 2018 (except for 2021). While we cannot definitively state why there was a decline in reports in 2023, we have identified several factors that we believe contributed to this general reduction. First, we believe that the implementation of quality education and promotion of the reporting process led to fewer reports because stakeholders would not report expected and mild adverse reactions. Second, we think people have become more passive over time and less inclined to report adverse reactions, possibly due to the normalisation of certain side effects. Third, we believe there might be increased trust in medications, especially if they have proven effective and safe. Fourth, we think media coverage is responsible for reporting adverse reactions. Pharmacovigilance activities are not a common media topic (except during 2021 because of the COVID-19 pandemic), hence we believe this led to the peak in reports in 2021. If we consider the slight decline in reports after 2018, it could be attributed to "digital inaccessibility," i.e., the fact that some citizens in Croatia might not be able or willing to use digital tools.

Discussion of Fig. 4, which displays the trend of reported adverse drug reactions from 2018 to 2023, by role of reporter. Fig. 4 shows that all lines follow a similar trend – grow in 2019 and 2021, with drop in 2020, 2022 and 2023. Furthermore, from Fig. 4, it is apparent that the majority of adverse drug reaction reports are actually submitted by healthcare professionals (pharmacists and physicians), which is unsurprising given the legal obligation of these professionals to report (HALMED, 2024c), and also the fact that by reporting drug reactions, pharmacists accumulate points needed for issuing or renewing the Authorisation for Independent Practice (HLJK, 2024). In 2019 the grow on graph is because 2019 is first full year of using OPeN in HALMED. Furthermore, in 2021, there is a peak of reported adverse drug reactions from all stakeholders, especially patients which exceeds healthcare professionals, because patients reported the most adverse reactions during the vaccination campaign during the COVID-19 pandemic which started in 2021.

Moreover, using DATAtab (DATAtab, 2024), we checked if there is a correlation between number of reported adverse drug reactions from all stakeholders and Drug Utilisation Reports (we used DDD/1000/day value from each year) from 2019 until 2022 (HALMED, 2024d), because 2019 is the first full year of using OPeN, while HALMED did not publish Drug Utilisation Reports for 2023. The results of the Pearson correlation $(r(2) = 0.12, p = 0.882)$ showed that there was a low, positive correlation between the number of adverse drug reactions and the DDD/1000/day value of Drug Utilisation Reports ($r = 0.12$). However, the correlation was not statistically significant ($p = 0.882$), because $p$ value greater than the significance threshold (0.05) suggests that the observed correlation could very well be due to random chance.

Currently, HALMED is upgrading the OPeN information system to enable reporting through hospital information systems and general practice information systems. We believe this is beneficial as it will allow for quicker reporting of adverse drug reactions with "one click," thereby encouraging healthcare professionals to be more active in the reporting process. Additionally, it will enable the possibility of reporting through the e-Citizens portal.

All these systems share a joint component base called eLijekovi, which contains information about all the medicines available in the Republic of Croatia. The data in eLijekovi are recorded in the form of the International Organization for Standardization (ISO) for the identification of medicinal products (IDMP) standards (EMA, 2024) which the EMA has prescribed as the standard for recording and exchanging medicine data.

In this paper, we also proposed an extended model (Fig. 5) of the system for collecting side effects, which on the HALMED system side includes the use of artificial intelligence (AI) for receiving and processing data which could be used as a proof of concept (PoC) in the near future in accordance with HALMED's time and financial resources. Given the standardisation of medicinal product data at national and European levels, artificial intelligence could use coding lists with standardised terms to recognise and classify adverse event data while receiving an adverse reaction suspicion. Additionally, in the case when the received data are not in standardised form, AI could be trained to match the data to standardised form. This data can also be used also by an AI in the process of analysing, identifying, and classifying the adverse drug reaction as mild, moderate, severe or lethal, comparing the received reaction of the same drug and suggesting the correction measures if possible. For training the AI, all the data currently stored in OPeN could be used.

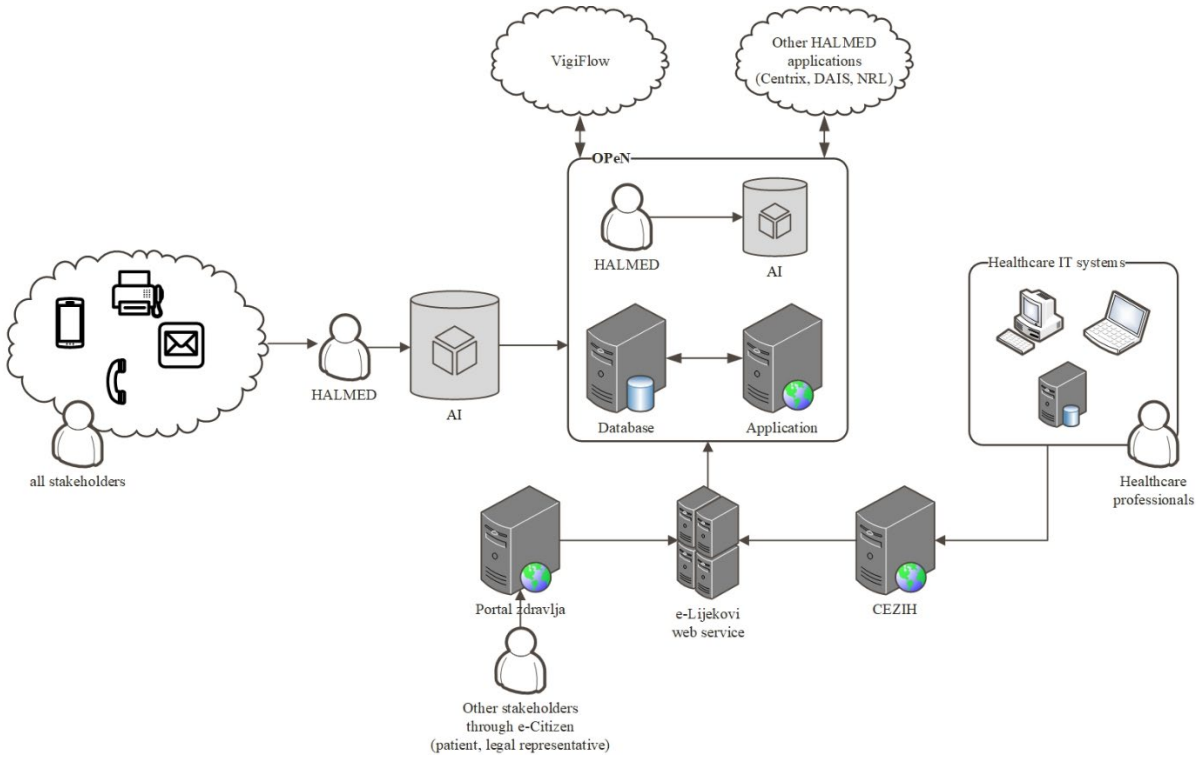



**Figure 5**: Proposed OPeN's structure scheme

This data were manually checked and all received adverse reaction suspicions were classified by HALMED's subject-matter expert.

The above mentioned process would speed up and simplify analysis and processing, as well as increase data accuracy. Suppose an adverse event report is received digitally, using artificial intelligence. In that case, the responsible HALMED employee only needs to verify the received data (without needing to overtype), expediting the business process of analysing received side effects.

## Acknowledgments

We want to acknowledge Prof. Siniša Tomić, PhD, head of the Agency for Medicinal Products and Medical Devices (HALMED) for granting us the right to use aggregated data.

## References


Balenović, A., Lazić, V., & Mesarić, J. (2022). Elektronički zdravstveni zapis—Prednosti i prepreke u uspostavljanju i korištenju. *Zbornik sveučilišta Libertas*, *7*(7), 19–30. https://doi.org/10.46672/zsl.7.7.2

DATAtab (2024). *DATAtab: Online Statistics Calculator*. DATAtab team. Retrieved from https://datatab.net

Dubey, S., Bhamre, P., Patil, A., & Kumar, R. (2021). Drug Safety Report Generator. *International Journal of Scientific Research in Science and Technology*, 38–49. https://doi.org/10.32628/CSEIT21743

Đanić Čeko, A., Guštin, M. (2022). Digitalizacija hrvatske javne uprave s posebnim osvrtom na sustav socijalne skrbi. *Zbornik radova Pravnog fakulteta u Splitu*, 793-821. https://doi.org/10.31141/zrpfs.2022.59.146.793

Edwards, I. R., & Aronson, J. K. (2000). Adverse drug reactions: Definitions, diagnosis, and management. *The Lancet*, *356*(9237), 1255–1259. https://doi.org/10.1016/S0140-6736(00)02799-9

Elzagallaai, A. (2012). Adverse Drug Reactions: An Overview of the Old and Continuous Challenge to Drug Therapy and Development. *The Libyan Journal of Pharmacy and Clinical Pharmacology*, *1*. https://doi.org/10.5542/LJPCP.v3i0.v1i0.571902

EMA (European Medicines Agency) (2024) *Data on medicines (ISO IDMP standards): Overview*. Retrieved from https://www.ema.europa.eu/en/human-regulatory-overview/research-development/data-medicines-iso-idmp-standards-overview

Eşiyok, A., Uslu Divanoğlu, S., & Çelik, R. (2023). Digitalisation in Healthcare—Mobile Health (M-Health) ApplicationsSağlık Hizmetlerinde Dijitalleşme- Mobil Sağlık (M-Sağlık) Uygulamaları. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *15*. https://doi.org/10.52791/aksarayiibd.1241287

Farooq, H., Niaz, J. S., Fakhar, S., & Naveed, H. (2021). Leveraging digital media data for pharmacovigilance. *AMIA Annual Symposium Proceedings*, *2020*, 442–451.

Fossheim, K., & Lund-Tønnesen, J. (2023). Digitalisation of public sector organizations over time: The applicability of quantitative text analysis. *International Review of Administrative Sciences*. https://doi.org/10.1177/00208523231183569

Gawich, M., & Alfonse, M. (2022). A Proposed Model for Drugs' Review Analysis and Adverse Drug Reaction Discovery. *2022 7th International Conference on Mathematics and Computers in Sciences and Industry (MCSI)*, 139–145. https://doi.org/10.1109/MCSI55933.2022.00028

Greve, C., Lægreid, P., & Rykkja, L. (2016). *Nordic Administrative Reform. Lessons for Public Management*. https://doi.org/10.1057/978-1-137-56363-7

HALMED. (2024a). *Activities | General Information and Documents | About HALMED: HALMED*. General Information and Documents | About HALMED. Retrieved from https://www.halmed.hr/en/O-HALMED-u/Osnovni-podaci-i-dokumenti/Djelatnosti/

HALMED. (2024b). *Legislation | About HALMED: HALMED*. Legislation | About HALMED. Retrieved from https://www.halmed.hr/en/O-HALMED-u/Zakoni-i-pravilnici/

HALMED. (2024c). *Pharmacovigilance: HALMED*. HALMED. Retrieved from https://www.halmed.hr/en/Farmakovigilancija/Kako-prijaviti-nuspojavu/Postupak/

HALMED. (2024d). *Drug Utilisation Reports*. HALMED. Retrieved from https://www.halmed.hr/en/Promet-proizvodnja-i-inspekcija/Promet/Potrosnja-lijekova/Izvjesca-o-prometu-lijekova

HLJK. (2024). *Pravilnici*. Retrieved from https://www.hljk.hr/pravilnici-a16

Kamble, S. U., Bais, S.K. & Hugar, S. R. (2023). Review on Adverse Drug Reactions. *International Journal of Advanced Research in Science, Communication and Technology*, 342–351. https://doi.org/10.48175/IJARSCT-7985

Lentacker, A. (2024). Epistemology of the side effect: Anecdote and evidence in the digital age. *BioSocieties*, *19*(1), 84–111. https://doi.org/10.1057/s41292-022-00293-5

Mohsen, A., Tripathi, L. P., & Mizuguchi, K. (2021). Deep Learning Prediction of Adverse Drug Reactions in Drug Discovery Using Open TG–GATEs and FAERS Databases. *Frontiers in Drug Discovery*, *1*. https://doi.org/10.3389/fddsv.2021.768792

Moulay, B. (2023). Digital Innovation in Healthcare. In A. C. Bouarar, K. Mouloudj, & D. Martínez Asanza (Eds.), *Integrating Digital Health Strategies for Effective Administration* (pp. 1–19). IGI Global. https://doi.org/10.4018/978-1-6684-8337-4.ch001

Rajh, A., Gvozdanović, K., & Sudić, D. (2018). "OPeN"—The National Adverse Reactions Database in the Croatian eHealth System. *International Journal on Advances in Life Sciences*, *10*, 147–155.

Rajh, A., Sudić, D., & Gvozdanović, K. (2018). *OPeN: Linking the National Adverse Reactions Database with Clinical IT Systems in Croatia*.

Thomas, R. (2020). An Over View on Adverse drug Reactions. *Journal of Pharmacovigilance*, *8*(5), 1–1. doi: 10.35248/2329-6887.20.8.291

Uppsala Monitoring Centre. (2024). *The backbone of medicines and vaccines safety monitoring*. Retrieved from https://who-umc.org/pv-products/vigiflow/

Wang, J., Deng, Y. Shu, L. & Deng, L. (2020). Machine learning-based methods and novel data models to predict adverse drug reaction. *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1226–1230. https://doi.org/10.1109/BIBM49941.2020.9313093

Vestues, K., Mikalsen, M., & Monteiro, E. (2021). *Using digital platforms to promote a service-oriented logic in public sector organizations: A case study*. https://doi.org/10.24251/HICSS.2021.269

# Public Policymakers' Attitudes on Planning and Implementing Measures for 54+ Population e-Inclusion

**Robertina Zdjelar**
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, Varaždin
rzdjelar@foi.hr

**Nikolina Žajdela Hrustek**
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, Varaždin
nikolina.zajdela@foi.unizg.hr

***Abstract**. Social inclusion is a crucial component of sustainable development, and digital technologies now play a facilitating role for those who have access and the skills to use them effectively. However, the challenge lies in the exclusion of groups without Internet access or digital skills, which often includes the 54+ population. Addressing this problem, known as the digital divide, requires a systematic and comprehensive approach to creating and implementing public policies. This paper aims to examine public policymakers' awareness of the 54+ population e-inclusion issues and to explore the connection between their views on the importance of recognizing the 54+ population e-inclusion in public policies and their personal levels of e-inclusion. Data were collected electronically using a validated measuring instrument across eight counties in the Republic of Croatia. Inferential statistical analysis of the data confirms a connection between public policymakers' views on the importance of e-inclusion for the 54+ population in public policies and their personal levels of e-inclusion.*

**Keywords.** E-inclusion, Public policies, vulnerable population groups

## 1 Introduction

Recognition of this problem on national, as well as international level, resulted in development of active aging measures (UN, 1990). The growing percentage of the elderly population in the total population has become a global problem today and is recognized in many countries. Predictions by the United Nations (UN) say that the number of people over the age of 60 could account for about 34% of total population in Europe by 2050 (United Nations, 2017, p. 1). Active aging measures are nowadays contained in a series of acts and documents of international organizations, such as Agenda 2020 on sustainable development, which aims to "provide the possibility of inclusion, the right to equality in education and the promotion of lifelong learning opportunities for all people" (European Commission, 2014). The 2030 Sustainable Development Agenda contains a plan of activities within the 17 set goals related to population, planet, prosperity, peace and partnership. Among other things, the sustainable development policy contains guidelines for solving today's social problems adopted for the period up to 2030 (United Nations, 2015a). One of the conclusions is the comprehensiveness of digital inclusion and the widest possible inclusion of citizens with the aim of achieving equal opportunities in sustainable development. Achieving equal conditions for all citizens cannot be attained without determining the main causes of digital divide, identifying homogeneous groups of users who are not e-Inclusive, as well as a systematic approach by the Public Sector in the creation and implementation of measures aimed at preventing the digital divide. Recent crises such as the Covid-19 pandemic have shown that at least part of the negative social and economic consequences of crisis situations can be mitigated (or even prevented) by e-inclusion of citizens (Robinson et al. 2020). The advantages of e-inclusion are particularly evident in the case of the population living in rural areas and/or affected by some of the more difficult life circumstances such as poverty, unemployment, health problems, disability, etc. Measures to restrict movement and daily migration caused by the pandemic led to isolation and inaccessibility or lower quality of public services, especially in the case of such vulnerable population groups (Guenther et al., 2020). Therefore, solving the problem of insufficient e-inclusion of certain social groups has significant implications on the development of economic and other activities in society, especially during emergency and crisis situations.

In the continuation of this Paper, the specifics of the 54+ population are explained, and the research problem is defined. This is followed by the conceptualization of the research question, research objectives and hypotheses, after which the research methodology is described. In the following chapter, the research results are presented and analysed. The last chapter highlights the most important conclusions and recommendations, as well as limitations of the conducted research and indications for future research.

# 2 54+ population e-inclusion

The growing percentage of the 54+ (population aged 55 to 74) in total population makes the e-inclusion issue an increasingly important social issue. At the same time, a prerequisite for a successful implementation of digital transformation is the public policymakers' and the general public awareness of the importance of quality aging. Through their joint efforts, United Nations member countries accepted the challenge by creating national, regional, and local strategies. In a series of acts, the European Union focused on digital technologies, accelerating access to knowledge, economic growth and job creation, as well as equality to contribute to the implementation of the 2030 Agenda for Sustainable Development (Vincent, 2019, United Nations, 2015a).

Bearing in mind the specifics of the 54+ population, public policy measures that should be emphasized and that can be considered relevant are: ensuring adequate digital infrastructure available to users equally in urban and rural areas, developing e-government, ensuring digital content (open data, English open data) and electronic public services (e-administration, English e-government), the availability of which, on the one hand, affects the quality of life of citizens in general, but also the empowerment to participate in the development of society (e-participation, English e - participation). On the one hand, a sufficient level of e-Involvement of public officials is needed (van Deursen, van Dijk, 2010), while on the other hand, user involvement in the development and provision of services is an important component of their accessibility. The reason for this is that the involvement of users in the development of practical application of digital solutions contributes to the prospects for acceptance and legitimacy, and thus to a higher level of their use (Simonofski et al. 2019, Tuan, 2022, Callens, 2023). Promoting the benefits resulting from the use of different types of electronic services - especially aimed at target groups of users - is another necessary measure. This contributes to the effective use of certain public digital services (e.g. e-learning, e-health, e-culture), as well as commercial services (e-commerce, e-banking), while more economic activity yields higher tax revenues for the budget in general, regardless of the level of public administration we are talking about. The analysis by the Center for Economics and Business Research (CEBR, 2023) in Great Britain which conducted a study of ten-year investment cycle of funds through numerous measures in e-inclusion showed that it resulted in multiple profits since the effects of e-inclusion of certain sensitive population groups reflect on: efficient use of time, increased income, employability, benefits in the supply sector, communication, data exchange with public administration, efficiency and savings due to digitalization of systems, corporate benefits and general benefits.

Furthermore, necessary measures include encouraging the activities of non-profit organizations dealing with issues of vulnerable population groups (citizens of the third age, people with disabilities) to organize activities aimed at raising digital literacy and encouraging e-inclusion of the target population. Encouraging and promoting the policy of active aging can be carried out precisely through the activities of non-profit organizations. Encouraging greater e-inclusion as designed in the Digital Agenda (European Commission, 2014) is expected to bring economic benefit (resource savings, availability of goods and services on the global market) and the achievement of high social goals of environmental protection and socially responsible business. Achieving these goals is possible provided that a large part of the population is active in the modern digital environment and if the e-inclusion is supported by public policies (Yu, Ndumu, Mon and Fan, 2018). Bearing in mind EUROSTAT data (Eurostat, 2020, e-society), this represents a challenge in the case of the 54+ population, since this particular group shows a significantly lower level of e-inclusion compared to the younger population, as evident from Table 1. In Table 1. data presents four categories of EU countries grouped by the percentage of 54+ population that never used Internet (in 2019). As it can be seen the in most of EU countries the percentage of digital divided 54+ between 31-40%. At that time the percentage of digital divided 54+ population is over 40% (41%), as well as it is in Portugal, Greece and Bulgaria.

**Table 1.** Percentage of respondents aged 55 to 74 who have never used the Internet in 2019 (European Union, 28 countries)

| % of resp. 54+ popul. | % countries | Comment |
|---|---|---|
| Up to 10% | 21,43 | Denmark,Netherlands, Sweden, Finland, Luxembourg, UK |
| 11 - 20% | 14,28 | Germany, Belgium, France, Estonia |
| 21-30% | 17,86 | Spain, Ireland, Czech Republic, Austria, Latvia |
| 31-40% | 28,57 | Slovakia, Slovenia, Italy, Lithuania, Hungary, Malta, Cyprus, Poland, Romania |
| 41%+ | 17,86 | Croatia, Portugal, Greece, Bulgaria |
| Total | 100 | |

Source: Zdjelar, according to EUROSTAT, e-society (2020).

# 3 Conceptual framework of the research

Considering that a systematic and horizontal approach in the creation of public policies is necessary for the successful integration of e-inclusion into public policies, and one of the assumptions being the willingness of public policymakers to adopt the issue of e-inclusion of the 54+ population, two basic research questions were formulated.

1. What are the attitudes of public policymakers in the Republic of Croatia towards the need to include 54+ population e-inclusion in public policies, and what is the personal level of e-inclusion of public policymakers?
2. Is there a connection between the level of readiness of public policies regarding 54+ population e-inclusion and the attitudes and personal level of e-inclusion of public policymakers?

In an effort to answer the formulated research questions, the research goal is to develop an instrument for measuring attitudes towards e-inclusion of the 54+ population in public policies and the personal level of e-inclusion of public policymakers, which would serve as a base for drawing a conclusion about the existence of a connection between attitudes about the importance of including 54+ population e-inclusion in public policies and the public policymakers' personal level of e-inclusion.

The research hypothesis is: H1: There is a connection between the views on the importance of including e-inclusion of the 54+ population in public policies and the personal level of e-inclusion of public policymakers.

In the continuation of the Paper, the research methodology is described, in order to be able to answer the research questions accordingly, achieve the research goals and accept or reject the set research hypothesis.

# 4 Paper Research methodology

In order to examine and determine the views of public policymakers on planning and implementation of measures for e-inclusion of the 54+ population through public policies and their personal level of e-inclusion, a validated measuring instrument was developed for which the steps of the methodology developed by DeVellis (2012) were used. The basis for determining the constructs (categories), attributes and items of the measuring instrument was an overview of the relevant literature in the field of e-inclusion. As the most suitable model for the production of a new measuring instrument, the general theoretical conceptual model of e-inclusion of the general population developed by Žajdela Hrustek (2015) was chosen together with a validated measuring instrument developed during the research by Žajdela Hrustek (2015). The newly constructed measuring instrument is formulated in such a way as to put emphasis on the impact of public policies and their focus on e-inclusion of the 54+ population.

In the measuring instrument, 31 claims are formulated, simple or with several sub-claims. The measuring instrument was divided into three parts, and included a total of 10 general variables on subjects and 79 variables on attitudes, of which six nominal variables and 73 rank variables. Respondents expressed their agreement or disagreement with the claims on an ordinal semantic scale and expressed their views on frequency. The claims in the measuring instrument were of a closed type, with the possibility of selecting exclusively one of the multiple answers offered.

After the measuring instrument was created, data was collected in the second step using the survey. The defined population of the research included a total of 128 respondents from the following areas: Zagreb County, Međimurje County, Krapina-Zagorje County, Varaždin County, Koprivnica-Križevci County, Bjelovar-Bilogora County, Virovitica-Podravina County and Istria County. The selection of counties was made on the basis of geographical similarity and readiness for cooperation of county services, because a significant part of the counties did not provide public e-mail addresses to members of the county assembly, and a significant part of them was unavailable due to the General Data Protection Regulation (GDPR).

The survey was conducted in the period from (October to December 2020). The respondents were members of county assemblies and county prefects, their deputies and heads. The measuring instrument was forwarded by random selection to a total of 386 respondents, of which a total of 128 (33%) responded voluntarily, who completed the questionnaire in full for the following three data sets: general data on respondents (county, gender, age, level of education and place of residence); a number of information and communication technology claims (ICT abbreviated) – claims 6 to 23; a series of claims about public policy preparedness for the e-inclusion of the population of people over the age of 54 – claims 24 to 31.

The measuring instrument was completed by respondents through an online questionnaire. The link leading to the online questionnaire was delivered to the respondents directly, i.e. to the professional service of a particular county that distributed a further sent request to the respondents by e-mail with a request for it to be filled in within the stipulated deadline. The introduction of the online questionnaire briefly describes the purpose of data collection, notes on confidentiality or the fact that the answers are anonymous in order to collect the highest quality and honest answers. For further processing and analysis, only questionnaires that were fully completed were

taken into account, while questionnaires that were not fully completed were excluded from further analysis. The following are the methods of statistical analysis that were applied after the data collection phase and the preparation of data for analysis:

- descriptive methods (percentages, mean values, dispersion, asymmetry and roundness measurements, and Spearman's rank correlation coefficient);
- inferential methods (Kolmogorov-Smirnov distribution normality test, chi-square test and Kruskal-Wallis test);
- multivariate methods (reliability analysis; Cronbach α coefficient).

Based on SPSS files, statistical analyses of IBM SPSS Statistics 25 were performed using a statistical package. Conclusions related to differences and correlation between variables were made at a level of significance of 0.05 or with a reliability of 95%.

# 5 Research results and discussion

The results of the analysis are presented and described in the following two chapters: descriptive statistical analysis and inferential statistical analysis.

## 5.1 Descriptive statistical analysis

The sample of respondents consisted of 71 males (55%) and 57 females (45%). The representation of age groups of respondents was predominantly between 35 and 54 years of age (58%), mostly university degrees (55%).

Respondents' claims regarding information and communication technology (ICT) indicating their personal e-inclusion show that they form a very homogeneous group with regard to experience in ICT and internet use; between 3 and 5 years for 91% of respondents.

The majority of respondents acquired digital skills at work independently or with the help of colleagues (44%), followed by those who acquired these skills independently at home (24%), as part of their education (19%), while the least was acquired through courses (13%), which may indicate the need for additional research on how much attention is being paid to the acquisition of digital skills in the employers' organization or with the encouragement of other profit and non-profit organizations or public authorities.

The majority of respondents declared that they have a desktop computer, a laptop and a mobile device (60.16%), a desktop computer and a mobile device (7.03%), a laptop and a mobile device (31.25%), only a mobile device (0 .78%), while 0.78% do not have any of the three mentioned devices.

In order to express the frequency of using the device with a single number, based on the values 1, 2, 3, 4, or 5, the arithmetic averages were calculated, which amount to: for desktop computers $\overline{X}$ = 3.4; for laptops $\overline{X}$ =4.1; for mobile devices $\overline{X}$ =4.7. Based on the above averages, it can be concluded that desktop computers are the least frequently used devices (average $\overline{X}$ = 3.4), while mobile devices are the most frequently used (average $\overline{X}$ = 4.7).

Accessing the Internet via a fixed telephone network or via a mobile broadband network is frequent or very frequent (93% of respondents). The averages are very high, both amounting to $\overline{X}$ =4.55 (between the terms "often" and "very often"). The most common place of access to ICT and the Internet is from home and from the workplace (the averages are $\overline{X}$ =4.54 and $\overline{X}$ =4.38, which is between the terms "often" and "very often"). Access in public places is very rarely used or not used at all, which leads to the conclusion that public places are not sufficiently placed in the function of serving citizens in terms of free Internet access.

ICT and Internet users are relatively satisfied with the speed or availability of the Internet network they use the most (the averages are $\overline{X}$ =3.80 and $\overline{X}$ =3.90, which is the closest to the term "agree"). 12% of the respondents are not satisfied with the speed of the Internet network, and 9% with the availability.

The intensity of using ICT and the Internet is usually between 1 and 5 hours per day on average, both for work and personal needs (averages are $\overline{X}$ =3.39 for work and $\overline{X}$ =2.83 for personal needs, while both medians are 3).

Respondents rate their ICT and internet skills as predominantly "very good". They rate the following four skills as the worst: using forums to exchange information, using software for creating charts in a spreadsheet (Excel and Lotus), downloading and installing software tools on a computer, and creating presentations (Power Point). For these skills, the averages are between $\overline{X}$ =3.4 and $\overline{X}$ =3.5 respectively. Respondents rate their skill in using software for writing and editing text (Word) and using internet search engines (Google, Yahoo) as the best, for which averages of $\overline{X}$ =4.2 and $\overline{X}$ =4.4 were calculated. On the whole, it can be said that the self-assessed skills are rated with averages between $\overline{X}$ =3.4 and $\overline{X}$ =4.4 and that these assessments are uniform (coefficients of variation range between 19% and 34%).

Most often, the respondents agree with the statements that examine attitudes towards security and privacy of the use of digital technologies. What is surprising is that the least accepted is the statement regarding personal data privacy concerns (average is $\overline{X}$ =3.59), which can be positive in the sense that it is not an obstacle to using ICT. In a negative sense,

however, it may indicate insufficient attention to an important aspect of (irresponsible) ICT use.

From the results related to "social support for users", it is evident that it is mostly "very common". Therefore, the arithmetic averages are slightly higher - above 3.7.

Regarding the statements related to the attitudes of the purpose of using ICT and from the results, an extremely positive attitude towards the use of ICT and the Internet is visible, especially for the purpose of employment, the use of public services via the Internet and for communication. Respondents see the use of ICT and the Internet useful for educational purposes, entertainment and hobbies, as well as for buying and selling products and services. They express their positive attitudes somewhat less regarding the use of ICT and the Internet for cultural content and for health purposes.

Statements related to the field of "e-participation" are most often accepted by respondents either rarely or often. Therefore, averages were also calculated for these statements between the terms "rarely" (3) and "often" (4). They are $\overline{X}$ =3.48 $\overline{X}$ =3.79 =3.57 and $\overline{X}$ =3.33 while the median values are also between 3 and 4. Public services via the Internet related to e-counselling are used the least, with around 60% of respondents declaring that they either "do not use" them at all or use them "very rarely" or "rarely". What ICT and the Internet are used for more often is finding "information I need related to the work of the Government and other public administration bodies", where more than 69% of respondents declared that they use ICT and the Internet for this activity "often" or "very often ".

Respondents mostly completely agree with the statements related to "e-democracy". Therefore, the arithmetic means are also high, between $\overline{X}$ =4.16 and $\overline{X}$ =4.55, while the coefficients of variation are moderate (between 15% and 24%). Great agreement is observed, for example, for the statement about availability of information and documents of public authorities via ICT and the Internet, more than 94% of respondents "agree" or "completely agree" with the statement. This is also valid in terms of finding information about "various social organizations/associations and their initiatives", which is very important nowadays for the normal functioning of modern society.

The respondents' engagement with social media is very diverse. According to the results, respondents do not prefer the use of ICT and the Internet in terms of activities such as "creating and editing a blog or website" either for themselves or for others. It is also evident that "forums for obtaining information and sharing opinions" are rarely used. Those activities that are slightly more prevalent in use than all the ones offered are "social networks", which nowadays boils down to the most commonly used social networks such as Facebook, Twitter and Instagram, as well as "communication via the portal and reading comments" and "adding independently created content". Given that nowadays social media plays a very important role related to social inclusion/exclusion, the role of e-inclusion of the 54+ population holds special significance in this context.

Respondents predominantly responded to statements regarding the importance of the topic of "e-inclusion of the 54+ population" in public policies with "I completely agree", which largely confirms the respondents' positive attitude regarding its importance. This is extremely important information since it is indicative of the awareness of public policy holders regarding the benefits of digitization and the inclusion of the 54+ population in the information society, as well as all the advantages and possibilities thereof. On the other hand, respondents stated that despite the fact that e-inclusion of the 54+ population is important, it is not incorporated into public policies in practice. This is reflected in their not so proactive action on the subject of "e-inclusion of the 54+ population". Less acceptance of that statement is reflected in the smallest average, which is $\overline{X}$ =3.6, while the other three averages are $\overline{X}$ =4.4. The respondents are relatively homogeneous in their attitudes since the coefficients of variation are smaller (between 17% and 32%).

The remaining four statements in the measurement instrument also refer to public policies: their influence, need to encourage, need to empower and their impact on improving the quality of life of the 54+ population. The largest number of respondents fully agree with the statement that public policies need to influence the greatest possible availability of ICT and the Internet. This can be seen from the highest frequencies as well as the arithmetic means, which for the mentioned three modes of influence amount to $\overline{X}$ =4.7, as well as low coefficients of variation (between 10% and 13%), i.e. very similar opinions of respondents. The respondents' positive attitude is particularly significant, as they are responsible for implementing measures into strategies, putting these strategies into action, and allocating financial resources. This is crucial, especially when considering budget creation and the management of public funds, which often present the greatest obstacles to "e-inclusion" for the 54+ population.

Respondents have very similar opinions regarding encouraging 54+ population to use ICT in different ways. This is confirmed by the highest frequencies for the rating "completely agree", arithmetic means that are high (between $\overline{X}$ =4.5 and $\overline{X}$ =4.6) and coefficients of variation that are low (between 12% and 15%). What the majority of respondents decided to "completely agree" on is in the matter of "skills development", which indicates the fact that policymakers are aware of the importance of their adoption and constant upgrading and improvement, given that the trends in the field of ICT and the Internet are subject to rapid change. They attach great importance to support, which is one of the key factors

when it comes to the inclusion of the 54+ population. They attach equal importance to strengthening motivation as well as changing attitudes, which should also be taken into account when it comes to the 54+ population.

When it comes to empowering the 54+ population for active e-Involvement and digitization, as well as the use of digital content, the respondents' opinions are also very similar and positive. This statement can be supported in several ways: through the highest frequencies for the rating "completely agree", through arithmetic means which are high ( $\overline{X}$ =4.20 and $\overline{X}$ =4.4) and through low coefficients of variation (between 17% and 19%). Thus, according to the results related to the claims related to the empowerment of the 54+ population using ICT and the Internet for active participation in "decision-making processes", "expanding and deepening political participation" and "networking and social inclusion" more than 80% of respondents "partially agrees" or "completely agrees" with the statements. A slightly smaller percentage also expresses agreement on the issue of empowerment related to "creating digital content".

The empowerment component is very important; as shown by the Covid-19 pandemic, as well as situations of natural disasters (earthquakes, floods, fires...), digitalization can greatly facilitate the availability and use of public and commercial e-services. In addition, some often undesirable and risky situations can be prevented if timely and accurate open data is available. It is possible to raise the quality of life of the 54+ population by developing and encouraging the use of services related to the domain of e-health, e-Learning, e-business, e-commerce, e-banking, e-entertainment, e-culture and e-government. This is the prevailing opinion expressed by the respondents in the last statements in the measuring instrument. This is shown by the highest frequencies for the rating "completely agree", high averages (between $\overline{X}$ =4.4 and $\overline{X}$ =4.7) and low coefficients of variation (between 13% and 20%). Respondents expressed the greatest agreement with the statements regarding the development and encouragement of the use of e-government, e-banking and e-health services. All this can both contribute to increasing the quality of life of the 54+ population, but also cause synergistic positive effects in terms of creating increased added value for the economy as well as the development of society as a whole. A large number of variables were defined as part of the measuring instrument, which is why it was necessary to summarize them for the purposes of inferential statistical analysis. For this purpose, two composite variables were formed, and one of them was further analysed.

Specifically, the following scales or subscales were formed: Scale A: personal level of e-inclusion of public policymakers (includes statements 9 – 26, without statement 13 and statement 16, which are negatively worded); Scale B: attitudes about the importance of including the topic of e-inclusion of the 54+ population in public policies (includes statements 27-31), which is broken down into five subscales: Subscale B0: includes statements 27a - 27d, related to the county development strategy; Subscale B1: includes statements 28a - 28c, which need to be influenced by public policies; Subscale B2: includes claims 29a - 29d, which needs to be encouraged by public policies; Subscale B3: includes statements 30a - 30d, which needs to be strengthened by public policies; Subscale B4: includes statements 31a - 31g, which aspects of the quality of life should be improved by public policies. Scales A and B and subscales B0, B1, B2, B3 and B4 were possible to form thanks to the coding of respondents' answers always in five degrees according to frequency and in five degrees according to acceptance of statements. At the same time, the lowest level 1 indicates the lowest frequency, and level 5 the highest frequency, i.e. the lowest and highest acceptance of statements that are always positively formulated. For each respondent, the variables were aggregated by adding up their responses and then dividing that total by the number of statements included. This calculation produces averages (arithmetic means with two decimal places), which can be used to classify data, compare differences, and determine the existence or absence of connections between scales or subscales, as well as between these scales/subscales and independent variables such as gender, age, education level, and place of residence. On scale A (personal level of involvement of public policymakers), the average value is = 3.69 ± 0.480, which gives a low coefficient of variation of 13%. The distribution of these values is slightly asymmetric to the left (Sk= -0.24) and is somewhat flatter (with a lower peak) than the normal distribution (K = -0.57).

On scale B (attitudes about the importance of including the 54+ population in public policies), the average value of high $\overline{X}$ = 4.46 ± 0.477, which gives a low coefficient of variation of 11%. The distribution of these values is also less asymmetric to the left (Sk = -0.86) and its peak is very similar to the normal distribution (K = 0.09). Subscales B0, B1, B2, B3 and B4 have high arithmetic means ( $\overline{X}$ = 4.21 to 4.70) and low dispersion (coefficients of variation are 10% to 17%). The distributions of these values are all left-sided asymmetric and most of them have a sharper peak than the normal distribution (especially distribution B4).

## 5.2 Multivariate statistical analysis

The reliability of the measuring scales (Reliability Analysis) of the mentioned seven scales and subscales was analysed with the Cronbach alpha coefficient for individual groups of variables. The Cronbach alpha coefficient is a measure of the internal consistency of a set of statements, and can take a value between 0 and 1; the closer it is to 1, the more reliable the measuring scale is, and Cronbach's alpha coefficient values

around 0.7 can be considered acceptable (Gliem, Gliem, 2003, Mejovšek 2008). The results of the analysis of the reliability of the measuring instrument based on the internal consistency of the Cronbach's alpha coefficient show that the answers to certain groups of statements are well aligned with each other according to individual scales or subscales, all Cronbach's alpha coefficients were satisfactory (Cronbach α >0.70; 0.76 and 0.94) which ensures reliability of the measuring instrument, i.e. that by using the same measuring instrument in repeated measurements, the same or similar measurement indicators can be expected.

## 5.3 Inferential statistical analysis

The empirical verification of the statistical significance and connection of the data from the survey was started in the first step by checking the normality of the distribution of seven composite variables (scale and subscale) using the Kolmogorov-Smirnov test. The distribution on scale A is the only one that follows a normal distribution ($p = 0.200$), while the distribution of other scales or subscales is not similar to the normal distribution ($p < 0.001$). This means that when applying inferential statistical analysis, it is possible to use parametric tests for scale A, while for all other scales (scale B, subscale B0, B1, B2, B3 and B4) it is possible to use only non-parametric tests. Values on scale A and values on scale B were classified into three categories based on quartile values for the purposes of testing with inferential statistical analysis tests. In this way, three categories of respondents were obtained with approximately a quarter of the respondents, half of the respondents and a quarter of the respondents (like a normal curve).

Therefore, the ratio variable was transformed into a ranking variable in the following way: a) scale A are not e-Inclusive at all (2.41-3.29) 32 respondents or 25% are partially e-Inclusive (3.30-4, 06) 64 respondents or 50% are fully e-Inclusive (4.07-4.56) 32 respondents or 25% b) scale B disagree (3.00-4.10) 30 respondents or 23% are undecided (4.11-4.82) 65 respondents i.e. 51% agree (4.83-5.00) 33 respondents i.e. 26% For scale B to each assigned expression ("do not agree", "are undecided", "agree") should add the phrase "that public policies can influence the key components of e-inclusion of the 54+ population".

### 5.3.1 Inferential method – Distribution normality

In the first part of the inferential statistical analysis, the connection between variables on scale A and values on scale B was examined in order to gain insight into whether there is a connection between the personal level of e-inclusion of public policymakers (scale A) and attitudes about the importance of e-inclusion of the 54 + population into public policies (scale B). The analysis of this connection was examined using the following three methods: - Spearman's rank correlation coefficient and scatter diagram, - the chi-square test, which tested the hypothesis of independence of variables in the contingency table, and - the non-parametric Kruskal-Wallis H test. Only non-parametric tests were used, because the agreement tests confirmed that the values on the B scale were not normally distributed. The rank correlation was obtained by Spearman's rank correlation coefficient, which is $rS = 0.39$. Thus, the values on the A scale are related to the values on the B scale. In addition, it can be stated that this connection is positive and statistically significant ($rS = 0.39$ $n = 128$ $p < 0.001$). This means that those respondents who are less e-Inclusive (low personal level, i.e. lower value on scale A) on average less accept the views that public policies can influence the key components of e-inclusion of the 54+ population (lower value on scale B). It can also be said that respondents who are more e-Inclusive (indicated by a higher value on the A scale) tend to more strongly believe that public policies can influence the key components of e-inclusion for the 54+ population (higher value on the B scale).

### 5.3.2 Inferential methods - chi-square test

Another method of statistical analysis is the chi-square test based on the observed frequencies in the contingency table. The results of the chi-square test ($\chi 2 = 12.028$ $df = 4$ $n = 128$ $p = 0.017$) show that the hypothesis of no association between the two variables (two scales) in the contingency table cannot be accepted because $p < 0.05$. Therefore, there is a statistically significant correlation between the three categories of scale A and the three categories of scale B. The nature of this connection can be discerned from the percentages calculated on the basis of the contingency table. That is why horizontal percentages were calculated for the two end categories "not at all" and "completely". Of the total number of respondents (32 or 100%) who were classified as "not at all" on the A scale, 41% disagree, and only 9% agree that public policies can influence the key components of e-inclusion of the 54+ population. On the other hand, out of the total number of respondents (32 or 100%) who are classified as "fully" e-Included on the A scale, only 9% of them disagree, and 38% agree that public policies can influence on the key components of e-inclusion of the 54+ population. Therefore, with a lower personal level of e-inclusion, there is a lower acceptance of the views that public policies can influence key components of e-inclusion of the 54+ population. On the other hand, with a higher personal level of e-inclusion, there is greater acceptance of the views that public policies can influence key components of e-inclusion of the 54+ population.

### 5.3.3 Inferential methods - Kruskal-Wallis test

The third method used is the Kruskal-Wallis H test, which compares three or more medians that are better means than arithmetic means in distributions that do not resemble normal distributions. Respondents who are classified into different categories according to

scale A statistically significantly differ with regard to the views that public policies can influence key components of e-inclusion of the 54+ population (H = 13,341 df = 2 n = 128 p = 0.001). With regard to scale A, respondents who are in the "not at all" group least accept the views that public policies can influence key components of e-inclusion of the 54+ population. Respondents who, with regard to scale A, are in the "fully e-Inclusive" group most accept the views that public policies can influence key components of e-inclusion of the 54+ population.

The research hypothesis reads "There is a connection between the views on the importance of including e-inclusion of the 54+ population in public policies and the personal level of e-inclusion of public policymakers". According to the previously presented analyses and the described results of the three methods used, it can be concluded that the research hypothesis can be accepted.

# 6 Conclusion

Considering the problems and consequences caused by the growing social and democratic digital divide, and the fact that public policies are crucial in reducing this gap, it became clear that it is important to explore public policymakers' attitudes towards the e-inclusion of the 54+ population within public policies, as well as how these attitudes are related by their own levels of e-inclusion. To determine the views of public policymakers on including e-inclusion for the 54+ population in public policies and their personal levels of e-inclusion, a measuring instrument was developed as part of this research. After conducting empirical research and statistical analysis of the collected data, a positive and statistically significant association between the personal levels of e-inclusion of public policymakers and their attitudes towards the importance of including e-inclusion for the 54+ population in public policies was determined. This confirms the research hypothesis: "There is a correlation between views on the importance of including e-inclusion for the 54+ population in public policies and the personal levels of e-inclusion of public policymakers."

As sustainable development becomes a key focus in the upcoming programming period until 2030, it is evident that significant changes in thinking, economic and social structures, as well as in consumption and production patterns are necessary. This shift will inevitably affect the 54+ population, who must become active stakeholders in the digital society due to their own interests and those of society. Further research should be conducted across a wider geographical area of the Republic of Croatia and compared with other EU Member States to gain a comparative insight and provide a larger research sample. Additionally, it is important to investigate whether strategic plans for the digitalization of the economy and society, for the new programming period until 2030, include substantial measures to promote e-inclusion of the 54+ population, both at the European and national levels, and to understand how these measures will be implemented.

# References

Callens, C. (2023). User involvement as a catalyst for collaborative public service innovation. *Journal of public administration research and theory*, 33(2), 329-341.

Centre for Economics and Business Research (CEBR) (2023) The economic impact of digital inclusion in the UK, available on https://www.goodthingsfoundation.org/insights/the-economic-impact-of-digital-inclusion-in-the-uk/, accessed on September 28th 2023

DeVellis, R.F. (2012) *Scale Development – Theory and Applications* (Third Edition). Sage.

European Commission (2014) Digital Agenda 2020, The Digital Skills and Jobs Coalition, available on https://ec.europa.eu/digital-single-market/en/digital-skillsjobs-coalition, accessed on December 1st 2016

European Commission (2014) Digital Agenda 2020, The Digital Skills and Jobs Coalition, available on https://ec.europa.eu/digital-single-market/en/digital-skillsjobs-coalition, accessed on December 1st 2016

EUROSTAT (2020) e-society, Dataset: Individuals - internet use [ISOC_CI_IFP_IU], Last updated: 15/04/2020 23:00, https://ec.europa.eu/eurostat/, accessed on October 11th 2020.

Gliem, J. A., Gliem, R. R. (2003) Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales, Midwest Res. to *Pract. Conf. Adult, Contin. Community Educ. Calc*., pp. 82– 88, 2003.

Guenther, J., Smede, B., & Young, M. (2020) Digital inclusion in central Australia: what is it and what makes it different?. *Rural Society*, 1- 17.

Mejovšek, M. (2008*) Metode znanstvenog istraživanja u društvenim i humanističkim znanostima*. Naklada Slap, Zagreb, Croatia.

Robinson, L., Schulz, J., Dodel, M., Correa, T., Villanueva-Mansilla, E., Leal, S., ... Khilnani, A. (2020) Digital inclusion across the Americas and Caribbean. *Social Inclusion*, 8(2), 244-259.

Simonofski, A., Melin, U., Lindgren, I., Vanderose, B., & Snoeck, M. (2019, August). Towards a decision support guide for user participation in public e-service development. *In 25th Americas Conference on Information Systems (AMCIS 2019): New Frontiers in Digital Convergence*

*(Vol. 2, pp. 912-921).* Association for Information Systems.

Trilar, J., Kos, A., Jazbinšek, S., Jensterle, L., & Stojmenova Duh, E. (2018). ICT to Promote Well-Being within Families. *Sensors*, 18(9), 2760.

Tuan, N. M. (2022). Customer readiness–customer participation link in e-services. *The Service Industries Journal*, 42(9-10), 738-769.

United Nations (1990) Resolution 45/106 Implementation of the International Plan of Action and on Ageing and related activities, 68th plenary meeting, December 14th 1990., available on https://undocs.org/en/A/RES/45/106, accessed September 8th 2019., accessed on 05.01.2018.

United Nations (2015) Transforming our world: The 2030 Agenda for Sustainable Development, available on A/RES770/1 310 https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf, accessed on 10.08.2020.

United Nations (2015a) Transforming our World: The 2030 Agenda for Sustainable Development. accessed August 1st 2018, from https://sustainabledevelopment.un.org/post2015/transformingourworld, accessed on 20.01.2019.

United Nations (2017) Population facts: Population ageing and sustainable development. Department of Economic and Social Affairs, Population Division, No. 2017/1, available on http://www.un.org/en/development/desa/population/publications/pdf/popfacts/PopFact s_2017-1.pdf , accessed on 20.01.2019.

Van Deursen A., van Dijk J. (2010) Civil Servants' Internet Skills: Are They Ready for E-Government?. In: Wimmer M.A., Chappelet JL., Janssen M., Scholl H.J. (eds) Electronic Government pp 132-143. EGOV 2010. Lecture Notes in Computer Science, vol 6228. Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-14799-9_12.

Vincent, J.A. (2019) Older People and Sustainable Development. Social and economic development, Vol. IV., Department of Sociology, University of Exeter, UK, available on https://www.eolss.net/Sample-Chapters/C13/E1-20-04-01.pdf, accessed on January 15th 2019

Yu, B., Ndumu, A., Mon, L.M., Fan, Z. (2018) "E-inclusion or digital divide: an integrated model of digital inequality", *Journal of Documentation*, Vol. 74 Issue: 3, pp.552-574

Zdjelar, R. (2022) Spremnost javnih politika za e-uključivost populacije 54+ (Doctoral thesis). Varaždin, Faculty of Organization and Informatics, https://repozitorij.foi.unizg.hr/islandora/object/foi%3A7055

Žajdela Hrustek, N. (2015) Multidimensional and multiperspective approach for monitoring e-inclusion (Doctoral thesis). Varaždin, Faculty of Organization and Informatics, https://urn.nsk.hr/urn:nbn:hr:211:404534

# An Analysis on Digital Public Services in Romania

**Mirona Ana Maria Popescu, Andreea Barbu**
National University of Science and Technology POLITEHNICA Bucharest
Academy of Romanian Scientists, Ilfov 3, 050044 Bucharest, Romania
{mirona.popescu, andreea.barbu2901}@upb.ro

**Larisa Maria Fecioru, Elena-Teodora Miron**
miDOR Digital services
Ecaterina Varga 86 C, Sibiu, Romania
{larisa.fecioru, elena.miron}@midordigital.com

***Abstract***. *The digitalization of public institutions brings with it an increased level of efficiency and transparency and should be one of the development pillars of any intelligent community. Through digitalization, the whole activity of public institutions becomes much more efficient on all three levels: internally, intra-institutionally and externally, regarding the relationship with citizens and the rest of the institutions.*

*As research methods, the authors have performed a literature review of the articles published on digital public services to provide an overview of the current situation and challenges in this field. Furthermore, in this article the authors carry out a comparative analysis on the main e-government platform in Romania, using data provided by the online tool Similarweb. As key findings, there are a lot of Romanian citizens that are still not using digital public services, which is also reflected in the DESI index. A series of measures and projects are in progress to increase the accessibility level and develop more platforms to strengthen the relation between citizen and government by providing reliable and secure digital public services.*

**Keywords.** digital public services, citizen engagement, Romanian citizens

## 1 Introduction

Since 2014, the European Commission has been monitoring the progress made by Member States in the digital field through the Digital Economy and Society Index (DESI) reports. Starting in 2023 and in line with the Digital Decade 2030 policy programme, DESI is now integrated into the Digital Decade Status Report and used to monitor progress towards the overarching digital goals.

Digitalization requires public authorities to offer their services at a better quality, via the Internet, without the need to go to a counter. Digitalization brings multiple benefits to the public sector. A digitalized institution is more transparent, fairer and is also able to manage and correlate information faster and more efficiently. The digitalization of local public administration involves the use of information and communication technologies by the public sector with the aim of improving the provision of information and services, encouraging citizen participation in the decision-making process and the accountability of the institutional partner in a transparent and efficient manner.

The lowest level of digitalization performance is found in Romania, which ranks last in the EU in terms of digitalization according to the Digital Economy and Society Index (DESI) of 2021, a document developed by the European Commission, which indicates the weak virtual connection of citizens and the private environment with the administration and public institutions. At EU level (EUR-Lex, 2024), the European Commission is taking steps to help public administrations in member states make the transition to digital, so that all citizens can enjoy the advantages of smart public services. It focuses on reducing barriers to public services and ensuring cross-border access to them. Actions that are already under way include: ensuring that EU platforms can work together and interact with each other; financing large-scale e-participation projects; encouraging public services, businesses, and citizens to exchange solutions through the joined platform, a common space that enables the exchange and reuse of IT solutions and best practices and facilitates communication and collaboration on IT projects across the EU; standardization of electronic health records.

An essential requirement of digital public services is the existence of a secure digital identity. This allows citizens and businesses to prove who they are – or what business they own – and thus access relevant services online.

According to the National Institute of Statistics, in 2020, 78.2% of households in Romania had access to the internet at home, an increase compared to 2019 by 2.5 percentage points, 60.9% of which were concentrated in urban area. In other words, more than three-quarters of the Romanian population today has

access to the Internet, regardless of whether we are talking about the rural or urban environment, a population that could have significant benefits once some of the services offered by the Romanian state and interactions go online. Unfortunately, the statistics show a low percentage of digital literacy of people in the administration who have are specialized in this field at work, a fact that could complicate any digitalization process. However, the challenges can be successfully overcome if they are approached methodically, coherently and without further delay. (Digi, 2021).

The aim of this paper is to highlight the research that has been made in the area of digital public services in Romania. By determining the existing challenges, solutions and strategies can be implemented to support digitalization of public institutions and provide online services that meet both citizens and government expectations.

# 2 Methodology

The objective of this study is to determine the state of digitalization of public services in Romania. Therefore, 3 directions were analysed:

1. The first perspective is represented by scientific research on digital public services in Romania. Thus, bibliographic research was carried out, considering articles from 2015 to 2024, which addressed this topic, and the results and conclusions obtained were presented.
2. The second perspective consisted in the analysis of scientific sources to determine statistics carried out by state organizations, along with ongoing projects or initiatives pursued by the government.
3. The third perspective consisted in using the Similar web tool to extract data related to e-government platforms in Romania and to perform a comparative analysis of them based on indicators. The indicators sought to see the engagement of citizens for digital public services.

To pursue the first perspective a literature review was made, considering scientific papers related to the development and utilization of digital public services in Romania from 2015 until 2024. There were used databases like SpringerLink, ScienceDirect and Google Academic, to identify articles by the following keywords: digital public services, Romanian e-government, digital transformation. A synthesis analysis was performed to highlight the results, challenges, and trends in the field of digital public services.

The second part of the study was made through the analysis of official sources and state organizations that publish relevant data. The collected data was structured based on specific indicators as digital public services utilization, e-government projects, and legislative initiatives.

For the last perspective, the comparative analysis between the main e-government portals was used the online tool Similarweb to extract the raw data for each. There were identified the main key indicators, and the results were presented in tables and figures to illustrate the Romanian citizens engagement.

# 3 Results

### 3.1. State of the art on digital public platforms in Romania

These days, the crucial process known as "digitalization" controls the growth of businesses, government agencies, and society as a whole. The rules of the game are completely altered by technological evolution, which also brings forth a new paradigm in which cutting-edge technologies like artificial intelligence and big data, as well as block chain technology and conventional computer-based systems, are used. From this angle, Romania's services industry, which is part of the country's expanding economy, needs to follow global trends (RĂBONȚU & Babucea, 2020).

With the digital transformation component, the Romanian Digitalization Authority is presently putting out a new vision for the public sector. It's an outlook on the corporate world. It is predicted that the advantages of digital transformation—such as simplicity, efficiency, and transparency—could result in significantly increased process productivity. The vision that ADR will create in its work is predicated on the realization of public institution leaders for the necessity of new technologies, the ongoing adaptation to citizen's needs, and the provision of high-quality, quick, safe, and online public services. To draw attention to the unique features of Romania, the digitalization processes of public administration in other nations were found to have some peculiarities. It was possible to conduct a comparative study of sustainable development indicators (Burlacu et al., 2022).

The creation of a cogent and integrated digital infrastructure at the level of Romanian public administration, which could offer top-notch digital services to both individuals and businesses, is a key goal of the National Recovery and Resilience Plan. The author's research aims to determine the variables influencing Romania's public service digitalization process and to establish any existing relationships between them (Virlanuta et al., 2024).

To ensure transparency in public administration, it is necessary to guarantee the fact that the citizens have broad access to information held by state authorities. Additionally, individuals must be included in the decision-making process by being informed first and consulted second. As a result, maintaining openness necessitates, particularly in the present, the establishment and ongoing maintenance of the public authority's website, which makes it simple for citizens

to obtain any necessary public information. However, there are some communities in Romania that do not post publicly available information online. To guarantee adherence to the ideal of openness in public administration, a digitalization plan that can be implemented at both the national and local levels is necessary, allowing information of public importance to be readily available to all (Slabu, 2021).

The nuances of the interaction between technology and public administration are analyzed by examining how digitalization affects accessibility, efficiency, and transparency in public services. The benefits and difficulties of integrating cutting-edge technology like blockchain, data analytics, and artificial intelligence are given consideration. The study aims to uncover best practices and lessons learnt by conducting a comparative analysis of digitalization projects across various administrative settings. This provides valuable insights into the elements that contribute to the success of digital transitions (Goldbach, 2024).

The primary goal of digitalization is to create value based on innovation, digitalization, and digital technologies, therefore enhancing performance and efficiency in the public sector, and contributing to the deep change of the economy, public administration, and society. The study intends to demonstrate and raise awareness of the following issues: the population's degree of digitalization skills; the low degree of digital technology integration in the commercial environment; the incompatibility of IT systems in public administration; etc. Romania is ranked lowest in most analysis indicators regarding the level of digitalization in the European Union, according to surveys and data. During the pandemic, small and medium-sized businesses with a strong technological foundation have the best opportunity of finding new business prospects. Public administrations must, therefore, adjust to the new reality and provide services that speed up work completion and simplify procedures. Therefore they must also priorities improving citizen interaction by offering digital solutions, so individuals may avoid needless excursions to governmental institutions with the intent to solve most of their concerns. From the time the epidemic first appeared until today, it can be said that Romania benefited from this erratic and utterly unusual period, particularly the faster digitalization of both private businesses and public administration (Panait & Rădoi, 2022).

Another study examines how Romanian authorities have used present technology to facilitate resident's access to public services, as well as the potential benefits of future technological advancements. In addition, a conclusion is drawn that although Romania has made some progress towards digitalization, its citizens still cannot access it, nor has it established an easily navigable cloud-based government system. Greater training on how to use digital public services and if the connection is secure should be provided by the public administration to people and public personnel. Also, investments in education must be made in advance since digital literacy must be incorporated into a citizen's life from an early age (Cojocaru et al., 2024).

Romania is notable for having one of the least developed online public service networks in the EU, despite having a vast telecommunications infrastructure that provides relatively affordable and quick network access. This is caused by the public's resistance, who was skeptical of the provided e-government solutions, as well as the government's policy, which up until 2019 did not demonstrate much interest in the nation's digitalization process or to propose a comprehensive program for its implementation. The COVID-19 epidemic, however, brought about a breakthrough: it compelled Romanian society to use public digital services and prompted the government to implement previously delayed legislation changes and new technology solutions (Calus, 2021).

Significant digital inequality is now present in Romania, and the COVID-19 pandemic epidemic has made it worse. The objective was to determine the obstacles that Romanian older individuals faced while attempting to utilize technology and to set up the groundwork for a more extensive study. The influence of the economic, social, cultural, and political difficulties caused by the pandemic on the well-being and quality of life of the elderly will also be covered in this article, along with public policy in the process of digitalization (Motorga, 2022).

It is also discussed how Romania's public administration is becoming more digitally innovative. For certain payments, including penalties, taxes, etc., there are several platforms/portals; nonetheless, paying these "obligations" in person at a counter is simpler than using platforms. It was analyzed the present state of innovation in Romanian public administration through digitalization. Although the study is theoretical, it will highlight several real-world elements of the problem of digitalization in Romania's public sector and place the country in various ranks for innovation brought about by digitalization. Results: Romania does not place well in the ranks of digitalization, mostly due to the lack of education and distrust of new technology among Romanian residents (Apostol, 2022).

Other researchers analyze Romania's digital technology trends in relation to other EU nations by considering various DESI index metrics for the years 2017–2018. After analyzing twenty-eight European nations, four primary components were developed: digital mass market, advanced digital features, ultramodern broadband access, and pricing or communication consideration. Additionally, clustering techniques identified two primary patterns with respect to advanced digital features: those that score significantly higher on this component and those that do not (highest number of mobile broadband services, highest penetration of digital public services, high usage of Internet content, and high penetration of

advanced digital skills). The younger demographic is more impacted and has the potential to have their behavior changed quickly because they are the true early adopters of new technology. Thus, to make up for Romania's lack of digital capabilities, particular attention should be paid to how digital technologies affect the youth there (Burlacioiu et al., 2018).

### 3.2. Perspective on digitalization of public services in Romania

The special Eurobarometer "Digital Decade 2024" shows that 45% of Romanian respondents believe that the EU protects their digital rights. Confidence in digital privacy is 48%, slightly below the EU average. Concerns include children's online safety (48%) and control over personal data (41%), both showing significant growth. These findings underscore the need to strengthen digital rights at the national level. Positive trends include the importance given by Romanians to digital technologies for accessing public services (64%) and connecting with friends and family (71%). Romania must improve its performance in terms of the targets and objectives of the digital decade, promote competitiveness, resilience, sovereignty and promote European values and climate actions (European Commission, 2024).

| Digital public services | Romania | | UE |
|---|---|---|---|
| | Place | Score | Score |
| DESI 2022 | 27 | 21,0 | 67,3 |

**Figure 1.** Traffic source overview February – April 2024 (Worldwide)

| Indicator | Romania | | | UE | |
|---|---|---|---|---|---|
| Current Usage: Internet and Digital Applications | Q1 2023 | Q1 2022 | Change | Q1 2023 | Q1 2022 |
| % of internet users | 15.5% | 16.4% | -0.9% | 17% | 17.5% |
| Data Consumption | | | | | |
| Average data consumption per user (GB/user/month) | 29 GB | 23 GB | 26% | 61 GB | 55 GB |
| Online Purchases | | | | | |
| Percentage of people who shop online (%) | 43% | 36% | 7% | 63% | 60% |
| Digital Skills | | | | | |
| % of people with basic or above digital skills | 42% | 39% | 3% | 65% | 63% |
| Internet Penetration | | | | | |
| % of households with internet access | 75% | 73% | 2% | 91% | 89% |

**Figure 2.** Traffic source overview February – April 2024 (Worldwide)

The provision of digital public services in Romania remains difficult (Fig. 1). All metrics (Fig. 2), including the provision of digital public services for individuals (44 vs the EU average of 75) and enterprises (42 versus the EU average of 82), show that the nation performs much worse than the EU average. Because just 17% of internet users use e-government services, there is also a limited level of digital contact between public authorities and the general population. The potential to enhance these outcomes is provided by the significant portion of digital investments and reforms allocated to this aspect in Romania's national resilience and recovery strategy. When these steps are implemented in a timely manner, they will help realize the goal of the Digital Decade, which is to provide 100% of critical governmental services online for European citizens and businesses by 2030.

Romania does not yet have an electronic identity system in place. For Romanians to connect with public, private, and sectoral entities more easily, digital signatures and electronic identity cards must be implemented. A total of EUR 200 million has been set aside for the issuance of 8.5 million electronic identity cards as part of the PNRR. Two digital certificates that enable (i) online public service authentication and (ii) qualified electronic signatures will be stored on the electronic identification card.

The establishment of a single architecture for the government cloud and the online delivery of necessary public services are the other two goals of investments funded by PNRR. It consists of the following: (i) setting up government cloud infrastructure (675 million EUR); (ii) developing and migrating to the cloud (€187 million); (iii) developing the e-health and telemedicine system (400 million EUR); (iv) digitalizing the judicial system (€162 million); (v) implementing electronic forms in the public procurement domain (0.85 million EUR); (vi) creating an electronic identity card and digital signature (EUR 200 million); (vii) digitalizing the NGO sector (€10 million) and civil service management (€10 million).

By utilizing cutting-edge technologies, these investments seek to modernize public administration while keeping people and companies' requirements front and center. Alongside this, the prerequisites for data-driven policymaking should be guaranteed, and the interoperability of current digital technologies should be strengthened. The construction of an integrated architecture for digital public services is also supported by the changes. The Government Cloud Act and the Information Systems Interoperability Act are the two main pieces of legislation that make up the PNRR's reform of government cloud services.

The Authority for the Digitalization of Romania (ADR) released the Catalog of Public Services, which will be updated annually, to track the advancements made in the digital transformation of Romania's public administration and to assist the government's strategic choices. This will be the programming tool used to design interventions, coordinate the delivery of digital public services, and digitize public services.

In its attempts to digitalize the public services sector, Romania has highlighted one of the main structural problems as being lack of experience in combining IT and administrative procedures inside institutions. Consequently, to fully profit from the significant measures currently in place, it is imperative to consistently enhance the administrative ability to conduct digitalization projects (DESI, 2022).

### 3.3. Perspective on digital public platforms in Romania

An analysis was realized on the internet to find and establish the digital public services that are available for Romanian citizens through web platforms. Even though government institutions have developed web platform to encourage citizens to move some of their public related activity in the online environment, the numbers are quite low (Table 1). The e-government web platforms that have the highest number of visits are Portal.just.ro, followed by anaf.ro and then e-licitatie.ro. If a visit is unique (1 visit per citizen), it means that only 44% of the Romanian population uses portal.just.ro, which is according to the number of visits in the last 3 months the most frequented one. Therefore, Romanian citizens are using e-government portals in a small percentage, meaning the traditional procedure remains the preferred one, according to the numbers.

**Table 1.** Total visits February 2024 – April 2024 (Worldwide)

| Nb. | Platform | Total visits |
|---|---|---|
| 1 | Ghiseul.ro | 26,308 |
| 2 | Regista.ro | 844,834 |
| 3 | Edirect.e-guvernare.ro | 218,392 |
| 4 | E-guvernare.ro | 1,594,000 |
| 5 | Anaf.ro | 8,291,000 |
| 6 | e-licitatie.ro | 2,127,000 |
| 7 | Pasapoarte.mai.gov.ro | 703,298 |
| 8 | Portal.just.ro | 8,718,000 |
| 9 | Cnas.ro | 1,569,000 |
| 10 | Econsulat.ro | 1,625,000 |

From the list of e-government web platforms analyzed, even if are developed to be accessed both by desktop and mobile, the Romanian citizens preferences are split equally: for Regista.ro, Edirect.e-guvernare.ro, E-guvernare.ro, Anaf.ro and e-licitatie.ro most of the traffic comes from desktop devices and for the rest of the 5 platforms the traffic comes from mobile devices. The differences between the desktop traffic and mobile web traffic for almost all the 10 e-government platforms are high, being more than 50% (Table 2).

**Table 2.** Device distribution February 2024 – April 2024 (Worldwide)

| Nb. | Platform | Device distribution (%) | |
|---|---|---|---|
| | | Desktop | Mobile Web |
| 1 | Ghiseul.ro | 43,32 | 56,68 |
| 2 | Regista.ro | 88,6 | 11,4 |
| 3 | Edirect.e-guvernare.ro | 53,49 | 46,51 |
| 4 | E-guvernare.ro | 86,67 | 13,33 |
| 5 | Anaf.ro | 74,4 | 25,6 |
| 6 | e-licitatie.ro | 93,55 | 6,45 |
| 7 | Pasapoarte.mai.gov.ro | 14,27 | 85,73 |
| 8 | Portal.just.ro | 32,66 | 67,34 |
| 9 | Cnas.ro | 33,3 | 66,7 |
| 10 | Econsulat.ro | 38,35 | 61,65 |

There is no website rank present for 2 of the 10 e-governments web platforms, neither global nor local (Table 3). The best rank for both categories is held currently by anaf.ro. This rank is given after analyzing all the web platform that are used at the time being by Romanian citizens, including all web platforms that have content and are available worldwide. ANAF has the mission to ensure the resources for the public expenses of the society by collecting effectively and efficiently administering the taxes, fees, contributions, and other amounts due to the general consolidated budget, as well as to improve the relational state of taxpayer – institution through which the new requirements of the citizen are met in this dual role of user and beneficiary.

**Table 3.** Country rank 2024 (Worldwide)

| Nb. | Platform | Rank | |
|---|---|---|---|
| | | Country | Global |
| 1 | Ghiseul.ro | 317 | 59,131 |
| 2 | Regista.ro | 867 | 162,932 |
| 3 | Edirect.e-guvernare.ro | - | - |
| 4 | E-guvernare.ro | 676 | 125,775 |
| 5 | Anaf.ro | 75 | 15,423 |
| 6 | e-licitatie.ro | 296 | 59,149 |
| 7 | Pasapoarte.mai.gov.ro | - | - |
| 8 | Portal.just.ro | - | - |
| 9 | Cnas.ro | 700 | 126,084 |
| 10 | Econsulat.ro | 1,217 | 74,335 |

Engagement of citizens can be measured by monthly visits, visit duration, pages per visit and bounce rate. Each of the 10 platforms analyzed along with the number associated for each factor that is representing the engagement is presented in Table 4. Anaf (2.763M), Portal.just.ro, (2.906M) and Ghiseul.ro (1.158M), have the most visits per month. This can indicate that these are e-government platforms for services that frequently used by Romanian citizens to acquire different services and resolve state duties. If the duration of visit as an average is considered, on Regista.ro and E-licitatie.ro the citizens spent more time than on the other digital platforms for public services. On Anaf (8.57) and E-licitatie (9.08), Romanian citizens visit more pages on each visit to

accomplish their purpose. The bounce rate, which indicates the proportion of users who arrive at the website, exit (or "bounce"), and do not browse additional pages on the same website, is higher on Cnas (64.66%) and Pasapoarte.mai.gov.ro (57.86%).

**Table 4.** Engagement overview February – April 2024 (Worldwide)

| Nb. | Platform | Engagement overview | | | |
|---|---|---|---|---|---|
| | | Monthly visits | Visit duration | Pages / Visit | Bounce rate |
| 1 | Ghiseul.ro | 1,158M | 00:05:04 | 5.11 | 32.72% |
| 2 | Regista.ro | 281,611 | 00:12:01 | 7.81 | 17.78% |
| 3 | Edirect.e-guvernare.ro | 72,797 | 00:02:41 | 3.09 | 43.43% |
| 4 | E-guvernare.ro | 516,384 | 00:05:15 | 2.27 | 45.53% |
| 5 | Anaf.ro | 2.763M | 00:08:39 | 8.57 | 20.91% |
| 6 | e-licitatie.ro | 709,250 | 00:11:42 | 9.08 | 29.08% |
| 7 | Pasapoarte.mai.gov.ro | 234,433 | 00:01:38 | 2.04 | 57.86% |
| 8 | Portal.just.ro | 2.906M | 00:05:28 | 6.43 | 32.63% |
| 9 | Cnas.ro | 523,167 | 00:02:27 | 2.37 | 64.66% |
| 10 | Econsulat.ro | 541,938 | 00:08:49 | 7.38 | 37.33% |

Traffic of Romanian citizens can be direct (visitors who access the website without using a search engine, links from other sites or social networks. They reach the site either by directly entering the URL in the browser or by bookmarking), email (visitors that enter the site via email marketing campaigns), referrals (visitors that enter the domain who did not find the domain from a Google search and arrived from other websites. Google monitoring systems identify a visitor as a referral when they click on a link from a social network or website and land on another website), social (traffic coming to the website, mobile site or mobile app from social networks and social media platforms), organic search (a search engine's method of finding and ranking website content naturally, without the influence of paid advertising), paid (any customer that visits a website after seeing paid advertising promotions. Many platforms offer paid traffic services including search engines, social media platforms, and media networks. One of the most common forms of paid traffic is Pay Per Click (PPC) advertising which utilizes Google Ads), display ads (images, videos, or gifs shown to users on websites or apps). As seen in Fig. 3, the most traffic on the e-government platforms come as direct traffic and organic search. Also regista.ro, e-licitatie.ro and econsulat.ro receive the highest values from direct traffic. The organic search contributes to platforms as cnas.ro, pasapoarte.mai.gov.ro and edirect.e-guvernare.ro. Therefore, the traffic for Romanian digital public services does not come from paid advertising or digital marketing.

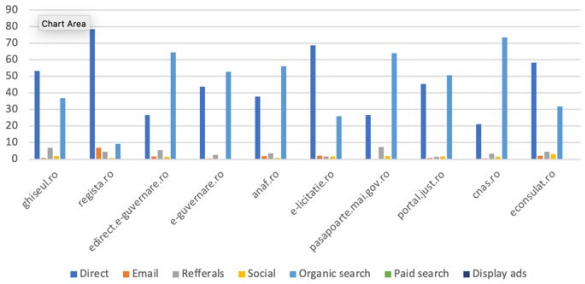



**Figure 3.** Traffic source overview February – April 2024 (Worldwide)

# 3 Conclusion

The conclusion taken based on the literature review shows that there are still challenges when it comes to digital public services, mainly due to the apprehension towards change. Even though the implementation of digital technologies in public administration has benefits such as process optimization and accessibility for citizens, there are rural areas where resources are limited. Therefore, digital divisions and data security concerns can be observed.

From the statistical data performed by specialized institutions from Romania, the efficacy of digital public services varies considerably. Data indicates a continuous necessity of investment in digital infrastructure and in the education of citizens to improve their digital skills and assure equal access to these services.

As a result of the platform analysis, the action to attract new users and maintain them due to potential issues such as accessibility, functionality, or not enough promotion is difficult. To improve upon this, the government should prioritize user experience and increase visibility of digital platforms. Also, Romania must improve its performance in terms of the targets and objectives of the Digital Decade, to stimulate competitiveness, resilience, sovereignty and to promote European values and climate actions.

In the last seven years, the DESI index for Romania increased by only 3.7 points, compared to 13.4 – the average increase of EU countries. Romania ranks last in the EU ranking, with a score of 30.6 points, compared to an EU average of 52.3 points. Finland, Denmark, the Netherlands, and Sweden continue to be in the top spots in the EU.

Romania has relatively good results in terms of connectivity, this being the dimension for which it gets the best score. The proportion of use of services with fixed broadband coverage of at least 100 Mbps (57 %) and via very high-capacity fixed networks (87 %) exceeds the EU average. This is also important given the Digital Decade's goal of 100% coverage of all households by gigabit networks by 2030.

Regarding digital public services, the current state is due to several reasons, one of which is the lack of implementation of the national digital identity.

However, there is PSCID (Centralized Software Platform for Digital Identification) in implementation. Specialists point out that a significant change in Romania's pace of training in terms of digital skills is essential for the EU to achieve the Digital Decade objective of basic digital skills and ICT specialists. Even though Romania has a strong IT sector, recognized at an international level for innovation and technological competences, this potential was not used to develop public services. A stronger collaboration between government and IT companies may accelerate the modernization of digital infrastructure, improve cybersecurity, and increase efficiency and accessibility of services provided to citizens. Therefore, Romania can benefit from a robust and competitive e-government system, adapted to current needs of population.

This study has its limitations, the first one being represented by the limited access to detailed data and updated information towards government projects and public data that can be incomplete or optimistic. Also, the generalization of results may not reflect the reality in all areas and among all citizens, since are differences in the regional parts of Romania. The lack of a close collaboration with these institutions and IT sector has limited a deep understanding of challenges and opportunities in the field of digital public services.

# References


Apostol, M.G. (2022). Innovation through Digitalization of the Romanian Public Administration. In Proceedings of the 28th International RAIS Conference on Social Sciences and Humanities (pp. 89-93). Scientia Moralitas Research Institute.

Burlacu, S., Pargaru, I., Iacob, O.C. and Gombos, S.P. (2022). Digital Public Administration and the Perspectives of Sustainable Development in Romania. European Journal of Sustainable Development, 11(4), pp.230-230 (Goldbach et al., 2024).

Całus, Kamil. (2021). "Digitisation in Romania accelerates during the pandemic."

COJOCARU, L., PICA, A.Ș. and STOICA, A.A. (2024). THE IMPACT OF DIGITIZATIONON PUBLIC ADMINISTRATION MANAGEMENT IN ROMANIA. Romanian Review of Political Sciences & International Relations, 21(2).

Comisia Europeana. (2024). Retrieved from https://digital-strategy.ec.europa.eu/ro/factpages/romania-2024-digital-decade-country-report

DESI. (2022). Retrieved from https://digital-strategy.ec.europa.eu/ro/policies/desi-romania

Digi24. (2021). Retrieved from https://www.digi24.ro/opinii/digitalizarea-serviciilor-publice-exemple-de-bune-practici-1765815

EUR-Lex. (2024). Digital public services. Retrieved from https://eur-lex.europa.eu/RO/legal-content/glossary/digital-public-services.html#:~:text=Acestea%20sunt%20servicii%20publice%20inteligente,nivel%20local%2C%20regional%20și%20național

Goldbach, D., Popa, M. and Paraschiv, O.V. (2024). Navigating the Digital Frontier: A Comprehensive Examination of Romanian Public Administration Digitization and Its Perceptual Implications. In Proceedings of the International Conference on Business Excellence (Vol. 18, No. 1, pp. 3452-3464).

Irimie, R.C., 2015. Digital agenda for Romania, progress towards 2020 targets. Studia Universitatis Babes-Bolyai-Studia Europaea, 60(4), pp.35-66.

Motorga, M.E. (2022). Public Policy and Digital Government Public Framework in Romania during the pandemic. The case of aging people. In Political Studies Forum (Vol. 3, No. 1, pp. 93-124). Editura Universității de Vest.

Panait, N.G. and Rădoi, M.A. (2022). Accelerating the Digitization Process in the Public Sector. Global Economic Observer, 10(1), pp.112-118.

RĂBONȚU, C.I. and Babucea, A.G. (2020). DIGITAL SERVICES IN ROMANIA-PRESENT AND PERSPECTIVE. Annals of'Constantin Brancusi'University of Targu-Jiu. Economy Series/Analele Universităţii'Constantin Brâncuşi'din Târgu-Jiu Seria Economie, (1).

Slabu, E. (2021). Transparency and Digitalization in the Public Administration of Romania. Public Administration & Regional Studies, 14(1), pp.49-56.

Virlanuta, F.O., Barbuta-Misu, N., Bacalum, S., David, S. and Moisescu, F. (2024). DIGITALISATION OF PUBLIC SERVICES IN ROMANIA. A CORRELATIONAL STUDY. Transformations in Business & Economics, 23(2).

Worldometer. (2024). Retrieved from https://www.worldometers.info/world-population/romania-population/).

# Innovation, Entrepreneurship and Economic Growth

Marijana Bubanić, Dina Korent
**Profitability of IT companies in the Republic of Croatia in (non)crisis periods: Skillings-Mack test**

Leon Šaško, Ivana Fojs, Kristina Detelj
**The rise of the gaming entrepreneurship – overview of the gaming industry ecosystem, case of Croatia**

Zrinka Lacković Vincek, Suzana Keglević Kozjak, Vladimir Kovšca
**More than twenty years of the eXtensible Business Reporting Language**

Tamara Šmaguc, Marina Klačmer Čalopa, Mirjana Grčić Fabić
**Phenomenological Inquiry in Entrepreneurship and Small Business with Insights into the Digital Economy Research Context**

Iva Gregurec, Magdalena Kuštelega, Goran Vlašić
**Exploring the trend of scientific research of eye-tracking for digital advertising insight – A Bibliometric Analysis**

# Profitability of IT companies in the Republic of Croatia in (non)crisis periods: Skillings-Mack test

**Marijana Bubanić, Dina Korent**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{mabubanic,dkorent}@foi.unizg.hr

***Abstract***. *Exogenous shocks jolt companies, which is often manifested in realized profitability. The aim of the paper is to determine whether there is a statistically significant difference in the profitability of the Croatian companies operating in the IT sector, before, during and after crisis periods. The paper covers two global crises: financial crisis and COVID-19 pandemic. Based on secondary data from the company's financial statements for the period from 2008 to 2023 and by applying the Skillings-Mack test, it has been proven that there are (in)equalities in the profitability of companies from the analysed industry, and that there is a statistically significant difference in profitability measured by the gross return on assets at a significance level of 1 % in the observed periods.*

**Keywords.** IT activity, Skillings-Mack test, panel data, crisis

## 1 Introduction

Without delving into a discussion about the ultimate purpose of the company's operations, Tamminen (2017) states that profitability is the final outcome of the overall management of pricing strategies of all products in relation to the overall cost structure with regard to the competitive environment. However, the question of what determines the profitability of a company is the focus of many researches[1].

Given Marx's theory of economic cycles, Fichtenbaum (1988) states that the participants in the discussion over this theory agree on one point, namely that crises are connected with a drop in the rate of profit. However, they disagree on the factors that caused this drop. This is supported by the fact that, according to Binz (2020) profitability is the result of the interplay between consumption, output and investment, which determine income and expenses, as well as the fact that during times of economic uncertainty, demand falls and income falls, resulting in a decline in profitability.

According to Jermann and Quadrini (2009) the crisis that began in the United States of America (USA) due to subprime loans in the summer of 2007 suggests that the financial sector plays an important role as a source of business cycle fluctuations. The financial shock, or disruption that originated directly from the financial sector, which consequently affected the *global village*, spilled over into the real sector as well. Prasad et al. (2015) in their research find evidence that the financial crisis reduced the profitability of companies from Indonesia, Korea and Thailand included in the sample.

After the recovery, the world was shocked by a new economic shock – the COVID-19 pandemic. The COVID-19 pandemic, as one of the external shocks, accelerated the digitization of various aspects of the company's operations by up to four years in just a few of months. It also increased the level of digital products and services in business portfolios that would otherwise have taken seven years to reach. (McKinsey & Company, 2020) Under pandemic conditions, digital technology has become a pillar for the development and creation of new business models that enable an improvement in relationships with customers and stakeholders as well as an overall improvement of the company. The above implies that the COVID-19 crisis affected the growth of demand for IT industry products and services. Based on the findings of their study of the IT sector in the Republic of Croatia, Zubak et al. (2022) conclude that the IT industry proved to be less sensitive to the negative circumstances of the crisis in 2020 and 2021 than most other industries, especially services.

This provokes the question: is the force of economic destruction damaging the profitability of companies in all sectors? For this reason, this paper aims to determine whether there is a statistically significant difference in the profitability of the Croatian IT industry, specifically companies operating in the IT sector, before, during and after the crisis. In

[1] Goddard et al. (2009) provide a qualitative overview of studies into the factors of firm profitability.d

this context, the goal is to determine (in)equalities in the profitability of business operations of the mentioned companies', not only in the pre-crisis and post-crisis period, but also in the crisis period. Profitability is therefore regarded as one of the indicators of the IT sector's resilience to external shocks.

The time horizon of consideration is the sixteen-year period from 2008 - 2023, whereby the period from 2008 - 2014 defines as the period of financial crisis, the period from 2015 - 2019 as the post-crisis period of the financial crisis, i.e., the pre-crisis period of the COVID-19 crisis, the period 2020 - 2021 as the period of the COVID-19 crisis and 2022 and 2023 as the post-crisis period of the COVID-19 crisis.

# 2 Literature Review

The aim of many researches is to test the existence of differences in determinants that affect the profitability of companies before and/or during and/or after the economic crisis (for example, Adlina, 2015; Argyrou et al., 2016; Cheong & Hoang, 2021 etc.), and among them there are especially those that investigate the aforementioned on samples of the banking sector (for example, Adelopo et al., 2018; Demirhan, 2013; Dietrich & Wanzenried, 2011; Rossi et al., 2018 etc.). However, there are few studies that empirically investigate whether there is a statistically significant difference between the profitability of companies before, during and/or after the crisis period.

Research that includes companies from the non-financial sector primarily differs in terms of the scope of the company's activities, and the time period or crisis that they examine. For instance, Prasad et al. (2015) investigate the mentioned issue on the example of industrial companies, Spitsin et al. (2020) on manufacturing companies, and Mazliza & Kusumaningdyah Adiati (2023) on listed manufacturing companies, Mulyono (2023) on listed companies but from the technology and telecommunications industry, Migliaccio (2018) and Pavone et al. (2023) on hotels with revenues over 800 thousand euros, and Migliaccio & De Palma (2023) on real estate companies. Apart from the aforementioned, the studies also include companies from various industries, which are listed on the stock exchange (Denčić-Mihajlov, 2014; Maziarczyk, 2020; Paulus Tahu & Yuesti, 2021; Rizaldi et al., 2022; Sulistyorini et al., 2023).

Furthermore, previously listed studies studies are primarily focused on the time period of a single crisis. For example, Maziarczyk (2020), Migliaccio (2018) and Pavone et al. (2023) investigate the existence of statistically significant differences in the profitability of companies due to the global crisis of 2008, which affected most countries of the world. The main findings of the aforementioned studies show that there is a statistically significant difference in profitability indicators in the non-crisis period versus the period of the global economic crisis. In particular, the results of this investigations reveal a decrease in the profitability of companies due to the global crisis.

The most recent external shock that jolted the world related to the COVID-19 pandemic is also considered by scientists as a crisis that affected the profitability of companies. Therefore, Mazliza & Kusumaningdyah Adiati (2023), Migliaccio & De Palma (2023), Mulyono (2023) and Paulus Tahu & Yuesti (2021) investigate the difference in company profitability due to the COVID-19 pandemic. The results of the mentioned research are not unambiguous. First, the main results of some of these studies show that there is a difference in the profitability indicator before the pandemic and during the COVID-19 pandemic. This is confirmed by the research results conducted by Mazliza & Kusumaningdyah Adiati (2023) and Mulyono (2023). Thay agree that the pandemic has reduced the profitability of companies. Second, some of these studies only partially confirm that there is a difference in profitability in pandemic versus non-pandemic periods, such as research conducted by Migliaccio & De Palma (2023). The main findings of that research can only to a limited extent confirm the difference in profitability before and during the pandemic, since the profitability of companies fell slightly and only in certain geographical areas. Finally, the main results of some studies shows that there isn`t statistically significant difference in the profitability of companies in the non-pandemic period versus the pandemic period. This is confirmed by the study results conducted by Paulus Tahu & Yuesti (2021). They didn`t find empirical evidence, i.e., there is no significant difference in profitability ratios before and after the announcement of the first case of Covid-19.

According to the author's knowledge, no research on IT activities has been conducted at this point, and unlike the aforementioned studies, which considered only one crisis, the subject research includes two exogenous shocks: the 2008 global crisis and the COVID-19 pandemic.

# 3 Methodological Framework

## 3.1 Sample

The research sample consists of 80,490 company-year observations, that is, quantitative secondary data on companies registered in the Republic of Croatia in the period from 2008 to 2023. The sample is based on the approach used by the Croatian Chamber of Economy (CCE) in its research of the Croatian IT sector (Zubak et al., 2022), and it includes companies registered in the classes of activities that comprise the IT industry according to the National Classification of Activities 2007 (Vlada Republike Hrvatske, 2007). First, the analysis includes companies/trades that are in the corporate income tax system and have submitted

annual financial reports to the Financial Agency (FINA); second, companies are classified according to how they registered their core activity; and third, the IT industry is defined as the following classes of the National Classification of Activities i.e., Statistical classification of economic activities in the European Community (NACE Rev. 2) (Eurostat, 2008; Vlada Republike Hrvatske, 2007):

- C 26.11 Manufacture of electronic components
- C 26.12 Manufacture of loaded electronic boards
- C 26.20 Manufacture of computers and peripheral equipment
- C 26.30 Manufacture of communication equipment
- G 46.51 Wholesale of computers, computer peripheral equipment and software
- G 46.52 Wholesale of electronic and telecommunications equipment and parts
- G 47.41 Retail sale of computers, peripheral units and software in specialised stores
- G 47.42 Retail sale of telecommunications equipment in specialised stores
- J 58.21 Publishing of computer games
- J 58.29 Other software publishing
- J 62.01 Computer programming activities
- J 62.02 Computer consultancy activities
- J 62.03 Computer facilities management activities
- J 62.09 Other information technology and computer service activities
- J 63.11 Data processing, hosting and related activities
- S 95.11 Repair of computers and peripheral equipment
- S 95.12 Repair of communication equipment.

The time period of the analysis covers the sixteen-year period from 2008 to 2023, with the period from 2008 to 2014 defined as the period of the financial crisis, the period from 2015 to 2019. as the post-crisis period of the financial crisis, i.e., the pre-crisis period of the COVID-19 crisis, the period 2020 - 2021 as the period of the COVID-19 crisis and the period of 2022 - 2023 as the post-crisis period of the COVID-19 crisis. The selection of the mentioned period is primarily the result of the defined problem and research objective.

Table 2., attached to Appendix 1, demonstrates the distribution of the number of companies in the defined time period by the covered classes of IT sector activities. Companies in activity class 62.01 Computer programming hold the largest share of the Croatian IT industry. In 2008, the number of companies in class 62.01 covered 34.81 % of the total IT industry, however by 2023, the share has climbed to 60.74 %. While the total number of companies in the IT industry expanded by 167.57 % between 2023 and 2008, the number of companies in class 62.01 increased for 3,816 companies, or 366.92 %.

## 3.2 Data and variables

The data used for research purposes are secondary in nature and includes data on the companies that make up the defined sample. The data about the companies in question includes both basic and quantitative (financial and non-financial) data. Primarily, quantitative data are mostly used to calculate indicators of variable at the company level, which are organized into a panel data structure. Basic data on companies, as well as data required to construct variable indicator at the company level, were obtained from the Financial Agency's databases (FINA, 2023). Based on the data obtained using the MS Excel software package, a profitability indicator was generated. The profitability indicator is chosen as a primary interest variable based on its theoretical basis, significance in current studies, and availability of the data needed for calculation. As an indicator of profitability, the gross profitability of assets (ROA) was chosen, which relates profit / loss before taxation and total assets, and is expressed as a percentage, which is often used in current studies (Maziarczyk, 2020; Mazliza & Kusumaningdyah Adiati, 2023; Migliaccio, 2018).

In the case of companies who in the reference year are obliged to submit both non-consolidated and consolidated financial statements, data from consolidated financial statements were used[2]. As the Republic of Croatia replaced the Croatian kuna as the official currency on January 1, 2023, and financial results in 2023 were expressed in euros, data from previous years' financial statements were converted into euros using the official exchange rate.

Considering business demographics, births, deaths, entry and exit of companies from a particular class of activity, the data was formed into an unbalanced panel structure. Regardless of the sample's imbalance, companies that survive from period to period imply sample dependency.

Appendix 2 provides a Table 3. of descriptive data for the variables used (total assets, gross profit / loss, and asset profitability) for the analysed period. The results confirm the imbalanced panel structure that is common in economic research while also showing the asymmetry of the distribution of the variable of primary interest. Therefore, the median is used as a representative measure of the mean value. Fig. 1. illustrates an estimate of the median gross return on assets for the IT industry in the Republic of Croatia from 2008 to 2023. The statistics results indicate a fall

[2] With the exception of 2023, where at the time of data availability, only data from non-consolidated financial statements were available.

in asset profitability during the global financial crisis beginning in 2008, with the first hints of recovery (% ROA higher than 2008) showing in 2014.

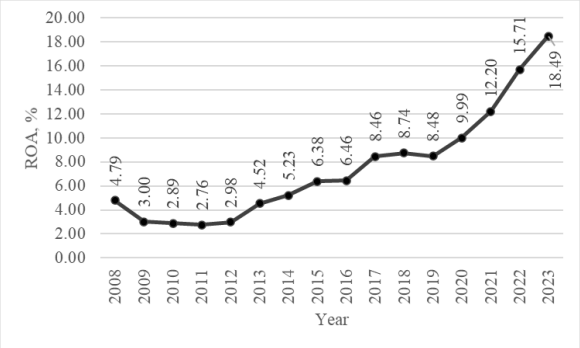



**Figure 1.** Median ROA of IT companies in the period from 2008 to 2023 [Author's work based on Stata output]

## 3.3 Methods

In addition to the standard techniques of traditional financial analysis for calculating indicators of the variable of primary interest and descriptive statistical methods, the paper used inferential statistics. Descriptive statistics were used to describe the sample in the previous chapter. Among the methods of inferential statistics, the Skillings-Mack test is a non-parametric pairwise multiple comparison of dependent groups that estimate the existence of statistically significant inequalities between indicators of the profitability variable (ROA) in specific years between 2008 and 2023. In previous studies, the ANOVA test with repeated measurements (Maziarczyk, 2020), or the paired sample t-test (Sulistyorini et al., 2023) were often used, where the aforementioned was attainable due to the use of balanced samples. In this research, the Skillings-Mack test is an appropriate method because it is a generalization of the Friedman test that can be used to an unbalanced panel data structure (Chatfleld & Mander, 2009). It is suitable for asymmetrical data distributions and dependent samples.[3]. With the aim of investigating (in)equality in the profitability of IT companies during and after the financial crisis, as well as before, during and after the crisis caused by the COVID-19 pandemic, in the first step, the Skillings-Mack test is used to test the existence of general statistically significant differences in ranks of the investigated indicator between certain periods, that is, between the years of interest. Then, in the spirit of post-hoc analysis and an effort to reveal differences more precisely, the Skillings-Mack test is used to assess the existence of specific statistically significant differences between all 120 pairs of years.

[3]For more on the Skillings-Mack test, see Chatfleld & Mander (2009).

# 4 Results and discussion

Table 1 displays the findings of the study on the existence of general statistically significant differences in the ranks of asset profitability or the years of interest. Based on the results, it is evident that in the observed period from 2008 to 2023 in the IT industry in the Republic of Croatia, there is a statistically significant difference in profitability measured by the gross return on assets at a significance level of 1 %.

**Table 1.** Results of the Skillings-Mack test for the period 2008 - 2023

| Year | N | Wsum CRank | SE | Wsum / SE |
|---|---|---|---|---|
| 2008 | 2,832 | 2070.49 | 173.69 | 11.92 |
| 2009 | 3,249 | 732.38 | 184.75 | 3.96 |
| 2010 | 3,455 | -91.78 | 192.73 | -0.48 |
| 2011 | 3,648 | -247.74 | 199.64 | -1.24 |
| 2012 | 3,766 | -728.74 | 205.03 | -3.55 |
| 2013 | 4,080 | 510.21 | 212.85 | 2.4 |
| 2014 | 4,338 | 0.27 | 218.86 | 0 |
| 2015 | 4,591 | 43.67 | 223.56 | 0.2 |
| 2016 | 4,868 | -2.01 | 227.44 | -0.01 |
| 2017 | 5,171 | 769.58 | 230.97 | 3.33 |
| 2018 | 5,549 | 84.02 | 234.28 | 0.36 |
| 2019 | 5,854 | -776.86 | 235.59 | -3.3 |
| 2020 | 6,161 | -918.48 | 236.39 | -3.89 |
| 2021 | 6,707 | -805.44 | 236.39 | -3.41 |
| 2022 | 7,490 | -508.03 | 234.01 | -2.17 |
| 2023 | 7,161 | -131.53 | 229.7 | -0.57 |
| Skillings Mack = 230.938 *** | | | | |
| P-value (No ties) = 0.0000 | | | | |
| Empirical P-value (Ties) ~ 0.0000 | | | | |

* 10 %; ** 5 %; *** 1 %
[Author's work based on Stata output]

Following the spirit of post-hoc analysis, Appendix 3 in Table 4 provides estimates of specific statistically significant differences between all 120 pairs of years of the analysed period. Due to the results' comprehensiveness, the most noteworthy ones are interpreted. The results indicate that there are statistically significant differences in the gross profitability of assets in all observed years compared to 2008, with the exception of 2023. When the global financial crisis hit in 2008, it affected almost every nation on Earth, and numerous companies suffered a sharp decline in their profitability. In the conducted

analysis, this is evident on the graph of the median ROA (Fig.1), which demonstrates a notable decline in the profitability of Croatian IT companies during the period of the global financial crisis. The above highlighted agrees with the research of Migliaccio (2018) who likewise use ROA as an indicator of profitability but on a sample of large Italian hotels. If we look more closely at the results of the Skillings-Mack test in 2023 versus 2008, where no statistically significant difference was found in profitability measured by gross return on assets, the data show that only 1.4781 companies survived in 2023 compared to 2008[4]. The reported surviving companies' median return on assets (ROA) was 8.78 % in 2008 and 8.83 % in 2023, according to available data.

It is interesting to observe the differences in profitability between the years of the period analysed (e.g. 2009 versus 2008, 2010 versus 2009, 2011 versus 2010, etc.). A statistically significant difference in the ranks of asset profitability can be seen in 2009 compared to 2008 and between 2010 and 2009. A look at the weighted sums of the centered ranks of the compared years, as well as the medians of both the entire sample and the medians of companies that persist from year to year, indicates a decrease in the profitability of companies in the IT industry during the global economic crisis. As stated by Ergović & Gambiraža (2023), during the global crisis there was a decline in Croatian purchasing power, which was partly a consequence of high indebtedness, which consequently reduced consumption and imports, and led to a global decline in domestic demand. The general decline in stimulus and consumption may have also affected the profitability of IT activities. The test results confirm that there is no statistically significant difference in ROA in 2011 versus 2010. The aforementioned is also evident based on the median profitability, both of the entire cause of companies in the mentioned years, as well as of only persistent companies. The gross profitability of assets was low in the observed years, and in 2011 compared to 2010 it suffered a slight decline. In 2012 versus 2011, the results of weighted sums of the centered ranks show that there was a decrease in profitability, at the level of statistical significance of 5%. If we look at the median ROA of companies that persisted in 2011 and 2012, it decreased by 0.34 percentage points (2011: 3.38 %, 2012: 3.04 %). From the foregoing, it can be concluded that in the aforementioned years of the financial crisis, the profitability of the Croatian IT industry continued to decline. Although the results of both the weighted sums of the centered ranks and the median show that there was a slight increase in the company's profitability in 2013 compared to 2012, it was still not statistically significant. In the next couple of years, 2014 ─ 2013, 2015 ─ 2014, and 2016 – 2015, there are statistically significant differences in the profitability of the analysed activity, as well as the appearance of an interesting phenomenon. Namely, despite profitability generally records growth, the results of the Skillings-Mack test, as well as the medians of persistent companies between pairs of years, show its decrease. For example, in 2014 compared to 2013, based on the Fig. 1, the growth of the median ROA is evident. Nonetheless, when looking at the weighted sums of the centered ranks of the years mentioned, as well as the medians of persistent companies in those years (2013: 5.43 %, 2014: 5.12 %), a decline in profitability is obvious. The aforementioned provokes the thesis that new companies entering the market compensate for the decline in profitability of companies that persist in their activities and may react to the crisis with more conservative business management. Even no empirically significant difference in profitability was found in 2017 compared to 2016, the results of the Skillings-Mack test as well as the median ROA confirm its growth. Pairs of years of the pre-pandemic period 2018 – 2017, 2019 – 2018, and 2020 – 2019 (pandemic versus pre-pandemic period) indicate that there are empirically significant differences in the gross profitability of assets. A similar phenomenon appears again as in the previously described case, in the periods 2018 – 2017 and 2019 – 2018, profitability is increasing globally, however, among companies that persist in pairs of periods, profitability is declining. In 2020 ─ 2019, the pandemic and pre-pandemic year, the results are not unambiguous. While the weighted sums of the centered ranks indicate a slight decline, the median ROA of all companies in the observed years, as well as the companies that persisted in those years, indicate an increase in the profitability of IT companies[5]. The results of the Skillings-Mack test indicate that there is no statistically significant difference between the pairs 2021 ─ 2020, although the growth of the median gross ROA is again noticeable. Among the persistent companies of the aforementioned years, there was an increase in profitability, but the number of companies that achieved a higher gross ROA in 2021 compared to 2020 is smaller than those whose gross ROA in 2020 was higher than in 2021. These findings during the pandemic period are consistent with those of Mulyono (2023), who also found that the profitability of more companies in the technology and telecommunications sectors listed on the Indonesia Stock Exchange decreased rather than increased. In 2022 versus 2021, the situation is similar

[4] Specifically, 1,479 survived, however, one company had no recorded assets, which made the ROA invalid and recorded as missing. As a result, the Skillings-Mack test did not consider this company. Note N= 8,017 not included as only had one observation.

[5] Companies that persisted in 2019 and 2020, and in both cases had a valid gross ROA indicator, indicate that 2,669 companies had higher ROA rates in 2020 compared to 2019, and 2,887 companies had lower ROA rates. For this reason, the Skillings-Mack test shows a decline, although the companies generally achieved an increase in the gross return on assets measured by the median of this indicator.

to that of a couple of years before, only that it is statistically significant. In the final few years of the observed period, using the weighted sums of the centered ranks results, a slight decrease in profitability was detected, although not statistically significant, accompanied by a decrease in the median ROA of persistent companies in the mentioned years, in same time a global increase in profitability when all companies are included of the observed period is evident. Ergović & Gambiraža (2023) state that the COVI-19 pandemic has certainly changed consumer habits. The ban on leaving the place of residence, the closing of borders and social distancing have led many consumers to shop from the comfort of their homes, i.e. many have started shopping online. Changing consumer preferences required business entities to adapt to online shopping. The COVID-19 crisis has introduced new global trends in business. (Ergović & Gambiraža, 2023) The aforementioned has certainly stimulated the increase in demand for IT services. On the one hand, many business entities had to offer an online way of shopping and thus introduce some form of IT business solution, while on the other hand, the population in order to buy goods and services had to have access to the Internet, they began to use more online forms of communication, as well as other IT solutions to adapt your life to numerous limitations. The aforementioned may have been reflected in the global increase in the median profitability indicators of Croatian IT companies.

# 5 Conclusion, contribution and limitation

By conducting empirical research, the main goal of the study was achieved, i.e. (in)equality was determined in the profitability of IT companies, not only in the pre-crisis and post-crisis period, but also in the crisis period. The research does not cover the period of one, but two major global crises: the 2008 financial crisis and the COVID-19 crisis. The main findings of the conducted research show that in the observed period from 2008 to 2023 in the IT industry in the Republic of Croatia, there is a statistically significant difference in profitability measured by the gross return on assets at a significance level of 1%. Also, the findings show that there is (in)equality in the profitability of these companies between pairs of years. These results are important both for policymakers and business leaders. Policymakers have to make decisions at the time of a certain crisis, which are primarily reflected in changes in fiscal and monetary policy. Therefore, it is important that they know how the measures they adopted in a particular period affected certain activities, in this case the IT activity, which is becoming increasingly important in the Croatian economy. On the other hand, it is important for business leaders to know whether the profitability of their company moves in accordance with the activity in which they operate, and how exogenous shocks affect its profitability globally.

The scientific contribution of the paper is manifested in the time span of a sixteen-year period, i.e. in a long term which, compared to many other studies that span the period of one crisis, spans the time period of two global crises. In contrast to existing studies mainly oriented towards non-financial companies that are listed on the stock exchange, as well as studies that often includes the banking sector, the research in question covers IT activity and not only large companies but companies of all sizes. Unlike many studies that test differences in the determinants of profitability before and/or during and/or after the crisis, the subject study examines the existence of statistically significant differences in profitability using a statistical test that is not common in existing studies. In addition to the above, the contribution is also evident in the number of tests of pairs of years, which estimated a total of 120 pairs of years, with the aim of drawing more detailed conclusions about the differences in the profitability of the companies from the sample between the years of interest.

The limitation of the study can be seen in the selection of only one profitability indicator compared to some studies that consider several profitability indicators. Therefore, other profitability indicators should also be considered in future studies. Although the topic of research covers various classes from different fields of activity, it focuses on research in the IT sector. As the instrument mentioned was not used in the previous studies and other activities were examined, comparability with these studies is limited. In this context, future research should also focus on other activities in order to compare the results and come to more comprehensive conclusions about the profitability of the Croatian economy in times of crisis and non-crisis. Furthermore, it would be a valuable extension of research to conduct comparative studies with companies from different countries. Finally, the observed phenomenon of global growth in the profitability of the companies observed after the financial crisis, while at the same time the profitability of the companies that persisted in that period became more modest, leaves open the question of whether the global crisis conditioned their more cautious, i.e. more conservative business management.

# Acknowledgments

This paper was supported in part by the Croatian Science Foundation under the project number [HRZZ-IP-2022-10-6703].

# References


Adelopo, I., Lloydking, R., & Tauringana, V. (2018). Determinants of bank profitability before, during, and after the financial crisis. *International Journal of Managerial Finance*, *14*(4), 378–398. https://doi.org/10.1108/IJMF-07-2017-0148

Adlina, U. N. (2015). *The Determinants of the Profitability of Malaysian Public Listed Companies* [Master of Science, Universiti Utara Malaysia]. https://etd.uum.edu.my/4662/1/s813062.pdf

Argyrou, M.-D., Bassiakos, Y., & Tsipouri, L. (2016). Determinants of the Profitability of Greek Enterprises before and after the Crisis. *MIBES Transactions*, *10*(2), 122. http://mtol.teilar.gr/vol10_issue2_2016/Tsipouri,%20Bassiakos%20&%20Argyrou.pdf

Binz, O. (2020). *Managerial Response to Macroeconomic Uncertainty: Implications for Firm Profitability* [Doctoral disertation, Graduate School of Duke University]. https://dukespace.lib.duke.edu/server/api/core/bitstreams/affaa452-50a2-416f-b380-0bf20bd45730/content

Chatfleld, M., & Mander, A. (2009). The skillings-mack test (friedman test when there are missing data). *Stata Journal*, *9*(2), 299–305. https://doi.org/10.1177/1536867x0900900208

Cheong, C., & Hoang, H. V. (2021). Macroeconomic factors or firm-specific factors? An examination of the impact on corporate profitability before, during and after the global financial crisis. *Cogent Economics and Finance*, *9*(1). https://doi.org/10.1080/23322039.2021.1959703

Demirhan, D. (2013). Effects of the recent financial crisis on the determinants of bank profitability: Case of Turkish banking industry. *Journal of Yasar University*, *8*(31), 5203–5228. https://dergipark.org.tr/en/download/article-file/179405

Denčić-Mihajlov, K. (2014). Profitability during the financial crisis: evidence from the regulated capital market in Serbia. *South-Eastern Europe Journal of Economics*, *12*(1), 7–33.

Dietrich, A., & Wanzenried, G. (2011). Determinants of bank profitability before and during the crisis: Evidence from Switzerland. *Journal of International Financial Markets, Institutions and Money*, *21*(3), 307–327. https://doi.org/10.1016/J.INTFIN.2010.11.002

Ergović, A., & Gambiraža, I. (2023). Uzroci i posljedice gospodarskih kriza s naglaskom na Republiku Hrvatsku. *Oeconomica Jadertina*, *13*(1), 50–64. https://doi.org/10.15291/oec.3986

Eurostat. (2008). NACE Rev. 2 – Statistical classification of economic activities in the European Community. In *eurostat Methodologies and Working papers*. https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF

Fichtenbaum, R. (1988). "Business Cycles", Turnover and the Rate of Profit: An Empirical Test of Marxian Crisis. *Eastern Economic Journal*, *14*(3), 221–228. https://www.jstor.org/stable/40325197

FINA. (2023). Baza podataka FINA-e o poduzetnicima. In *Kupljeni podaci o poduzećima iz odabranih djelatnosti*.

Goddard, J., Tavakoli, M., & Wilson, J. O. S. (2009). Sources of variation in firm profitability and growth. *Journal of Business Research*, *62*(4), 495–508. https://doi.org/10.1016/J.JBUSRES.2007.10.007

Jermann, U., & Quadrini, V. (2009). *Macroeconomic Effects of Financial Shocks* (Working Paper 15338; NBER WORKING PAPER SERIES). http://www.nber.org/papers/w15338

Maziarczyk, A. (2020). Global financial crises, profitability and outsourcing in industrial companies in Poland. *Organization & Management: Scientific Quartely*, *1*(49). https://doi.org/10.29119/1899-6116.2020.49.6

Mazliza, M., & Kusumaningdyah Adiati, A. (2023). Analysis of the Financial Performance of Manufacturing Companies Before and During the Covid 19 Pandemic. *Journal of Social Research* , *2*(3). https://doi.org/https://doi.org/10.55324/josr.v2i3.742

McKinsey & Company. (2020). *How COVID-19 has pushed companies over the technology tipping point—and transformed business forever*. https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever#/

Migliaccio, G. (2018). The profitability of Italian hotels during and after the 2008 Economic Crisis. *African Journal of Hospitality, Tourism and Leisure*, *7*(6). http//:www.ajhtl.com

Migliaccio, G., & De Palma, A. (2023). Profitability and financial performance of Italian real estate companies: quantitative profiles. *International Journal of Productivity and Performance Management*, *73*(11), 122–160. https://doi.org/10.1108/IJPPM-02-2023-0075

Mulyono, M. (2023). Pandemic Covid-19 Impact on Technology and Telecommunication Industry: Evidence from Indonesia Listed Company. *E3S Web of Conferences 388; The 4th International Conference of Biospheric Harmony Advanced Research (ICOBAR 2022)* , *388*. https://doi.org/10.1051/e3sconf/202338803034

Paulus Tahu, G., & Yuesti, A. (2021). Analyzing the effects of COVID-19 pandemic on the financial performance of Indonesian listed companies. *Academy of Accounting and Financial Studies Journal*, *25*(3).

https://eprints.unmas.ac.id/id/eprint/542/1/17%20ANALYZING%20THE%20EFFECTS.pdf

Pavone, P., Migliaccio, G., & Simonetti, B. (2023). Investigating financial statements in hospitality: a quantitative approach. *Quality and Quantity*, *57*(Suppl 3), 383–407. https://doi.org/10.1007/s11135-020-01086-3

Prasad, D., Puri, Y. R., & Jain, R. (2015). Return to Profitability after Financial Crisis. *The International Journal of Business and Finance Research* ♦, *9*(1), 89–101. https://www.theibfr.com/download/2746/ijbfr-v9n1-2015/12892/ijbfr-v9n1-2015.pdf#page=91

Rizaldi, R., Nur Huda, A., Alfiyan, A., & Kholilah, K. (2022). Comparative Analysis of Profitability, Firm Size and Market Share Before and During the Covid-19 Pandemic in Indonesia. *Proceedings of the International Conference of Islamic Economics and Business (ICONIES)*, 89–98. http://conferences.uin-malang.ac.id/index.php/iconies/article/view/1779

Rossi, S., Borroni, M., Lippi, A., & Piva, M. (2018). Determinants of Bank Profitability in the Euro Area: What Has Changed During the Recent Financial Crisis? *International Business Research*, *11*(5), 18. https://doi.org/10.5539/ibr.v11n5p18

Spitsin, V., Ryzhkova, M., Vukovic, D., & Anokhin, S. (2020). Companies profitability under economic instability: evidence from the manufacturing industry in Russia. *Journal of Economic Structures*, *9*(1). https://doi.org/10.1186/s40008-020-0184-9

Sulistyorini, E., Metyopandi, V., Ahirya, M., Rishma, M. A., & Patalo, R. G. (2023). Differences in Financial Performance of LQ45 Companies Listed on the Indonesian Stock Exchange during the Covid-19 Pandemic. *International Journal of Current Science Research and Review*, *6*(4), 2522–2529. https://doi.org/10.47191/ijcsrr/V6-i4-34

Tamminen, S. (2017). Regional effects or none? Firms' profitability during the Great Recession in Finland. *Papers in Regional Science*, *96*(1), 33–60. https://doi.org/10.1111/PIRS.12222

Vlada Republike Hrvatske. (2007). *Odluka o Nacionalnoj klasifikaciji djelatnosti 2007. - NKD 2007*. Narodne novine. https://narodne-novine.nn.hr/clanci/sluzbeni/2007_06_58_1870.html

Zubak, D., Mirošević, H., & Šaravanja, G. (2022). *Pregled stanja hrvatske IT industrije u 2021.* https://www.hgk.hr/documents/pregled-stanja-hrvatske-it-industrije-u-2021638a05d188c2f.pdf

# Appendix

## Appendix 1

**Table 2.** Distribution of the number of companies with regard to classes of IT activities in the period from 2008 to 2023 in the Republic of Croatia

| NACE Rev. 2 Classes | Year | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
| 26.11 | 46 | 47 | 54 | 52 | 53 | 54 | 54 | 54 |
| 26.12 | 1 | 5 | 7 | 6 | 7 | 6 | 7 | 10 |
| 26.20 | 241 | 268 | 253 | 245 | 237 | 235 | 232 | 226 |
| 26.30 | 57 | 61 | 55 | 63 | 60 | 62 | 55 | 54 |
| 46.51 | 231 | 221 | 231 | 238 | 228 | 234 | 225 | 225 |
| 46.52 | 64 | 49 | 70 | 82 | 83 | 92 | 92 | 96 |
| 47.41 | 133 | 147 | 145 | 154 | 152 | 136 | 126 | 123 |
| 47.42 | 15 | 9 | 18 | 19 | 23 | 28 | 35 | 42 |
| 58.21 | 1 | | | 1 | 2 | 3 | 6 | 8 |
| 58.29 | 147 | 150 | 138 | 141 | 128 | 124 | 125 | 119 |
| 62.01 | 1,040 | 1,320 | 1,312 | 1,421 | 1,539 | 1,737 | 1,959 | 2,128 |
| 62.02 | 246 | 261 | 311 | 315 | 326 | 346 | 351 | 378 |
| 62.03 | 41 | 34 | 68 | 72 | 89 | 107 | 129 | 139 |
| 62.09 | 320 | 286 | 373 | 390 | 390 | 431 | 447 | 456 |
| 63.11 | 180 | 202 | 194 | 208 | 211 | 226 | 238 | 253 |
| 95.11 | 218 | 262 | 247 | 257 | 256 | 259 | 269 | 275 |
| 95.12 | 7 | 6 | 14 | 16 | 18 | 23 | 33 | 38 |
| NACE Rev. 2 Classes | Year | | | | | | | |
| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
| 26.11 | 56 | 59 | 66 | 66 | 68 | 69 | 72 | 72 |
| 26.12 | 14 | 17 | 17 | 17 | 20 | 21 | 22 | 22 |
| 26.20 | 221 | 215 | 216 | 207 | 210 | 204 | 208 | 195 |
| 26.30 | 54 | 55 | 56 | 57 | 56 | 51 | 50 | 50 |
| 46.51 | 222 | 223 | 216 | 222 | 222 | 214 | 209 | 199 |
| 46.52 | 96 | 98 | 108 | 113 | 109 | 118 | 122 | 122 |
| 47.41 | 123 | 118 | 118 | 126 | 121 | 125 | 119 | 112 |
| 47.42 | 48 | 48 | 44 | 42 | 40 | 38 | 40 | 39 |
| 58.21 | 9 | 10 | 11 | 15 | 18 | 26 | 29 | 26 |
| 58.29 | 125 | 121 | 119 | 124 | 126 | 129 | 140 | 141 |
| 62.01 | 2,382 | 2,632 | 2,931 | 3,166 | 3,403 | 3,820 | 4,503 | 4,856 |
| 62.02 | 387 | 399 | 428 | 449 | 474 | 501 | 531 | 573 |
| 62.03 | 148 | 162 | 171 | 179 | 192 | 216 | 232 | 256 |
| 62.09 | 463 | 468 | 491 | 486 | 518 | 580 | 631 | 695 |
| 63.11 | 259 | 260 | 269 | 282 | 289 | 302 | 317 | 327 |
| 95.11 | 265 | 273 | 274 | 274 | 270 | 275 | 263 | 249 |
| 95.12 | 42 | 49 | 56 | 62 | 61 | 61 | 63 | 61 |

[Author's work based on Stata output based on data from the Financial Agency (FINA, 2023)]

## Appendix 2

**Table 3.** Descriptive statistics of research variables

| Year | | Gross profit / loss | Total assets | ROA |
|---|---|---|---|---|
| 2008 | M | 44729.30 | 477781.80 | 6.48 |
| | Me | 1998.61 | 42761.76 | 4.79 |
| | SD | 578171.40 | 4842489.00 | 51.90 |
| | N | 2988.00 | 2988.00 | 2988.00 |
| | Skw | 39.85 | 33.06 | -7.21 |
| | Krt | 1886.75 | 1247.04 | 195.07 |
| 2009 | M | 25527.08 | 443414.00 | -65.22 |
| | Me | 1215.94 | 40508.79 | 3.00 |
| | SD | 479935.90 | 4599329.00 | 3929.06 |
| | N | 3328.00 | 3328.00 | 3328.00 |
| | Skw | 3.14 | 34.65 | -57.62 |
| | Krt | 797.43 | 1402.21 | 3322.48 |
| 2010 | M | 25282.52 | 402458.10 | -19823.31 |
| | Me | 1095.23 | 34969.87 | 2.89 |
| | SD | 231918.10 | 4340174.00 | 999574.90 |
| | N | 3490.00 | 3490.00 | 3490.00 |
| | Skw | 7.30 | 34.36 | -57.04 |
| | Krt | 301.64 | 1344.40 | 3314.16 |
| 2011 | M | 26127.56 | 378314.50 | -3908.71 |
| | Me | 1117.26 | 31907.69 | 2.76 |
| | SD | 187822.40 | 3824713.00 | 178116.60 |
| | N | 3680.00 | 3680.00 | 3679.00 |
| | Skw | 2.76 | 34.04 | -49.57 |
| | Krt | 176.53 | 1310.05 | 2584.17 |
| 2012 | M | 27982.02 | 385269.60 | -3214.68 |
| | Me | 1150.51 | 29489.22 | 2.98 |
| | SD | 387583.60 | 3746154.00 | 193212.20 |
| | N | 3802.00 | 3802.00 | 3802.00 |
| | Skw | 29.09 | 33.09 | -61.63 |
| | Krt | 1486.38 | 1265.12 | 3799.22 |
| 2013 | M | 32207.25 | 368590.50 | -316.96 |
| | Me | 1405.27 | 26122.37 | 4.52 |
| | SD | 357527.10 | 3502833.00 | 745770.90 |
| | N | 4103.00 | 4103.00 | 4101.00 |
| | Skw | 37.81 | 32.61 | 26.60 |
| | Krt | 2014.17 | 1246.71 | 2188.98 |

*Mean (M), median (Me), standard deviation (sd), count (N), skewness (Skw), kurtosis (Krt)

[Author's work based on Stata output]

| Year | | Gross profit / loss | Total assets | ROA |
|---|---|---|---|---|
| 2014 | M | 34635.03 | 363149.60 | -7409.35 |
| | Me | 1420.93 | 24691.62 | 5.23 |
| | SD | 273182.90 | 3120205.00 | 309587.80 |
| | N | 4383.00 | 4383.00 | 4381.00 |
| | Skw | 22.66 | 32.39 | -58.93 |
| | Krt | 839.83 | 1338.20 | 3688.60 |
| 2015 | M | 35571.39 | 377223.80 | -11610.70 |
| | Me | 1628.04 | 26767.01 | 6.38 |
| | SD | 337657.60 | 3261276.00 | 646307.90 |
| | N | 4624.00 | 4624.00 | 4623.00 |
| | Skw | 11.83 | 32.04 | -66.96 |
| | Krt | 567.99 | 1320.08 | 4526.63 |
| 2016 | M | 38607.36 | 381280.80 | -2601.06 |
| | Me | 1951.95 | 27154.22 | 6.46 |
| | SD | 356464.30 | 3198270.00 | 427634.60 |
| | N | 4914.00 | 4914.00 | 4914.00 |
| | Skw | 23.75 | 28.25 | -2.75 |
| | Krt | 930.42 | 1015.26 | 2347.16 |
| 2017 | M | 44735.63 | 388727.00 | -5143.32 |
| | Me | 2577.88 | 31509.99 | 8.46 |
| | SD | 377696.30 | 3092753.00 | 205640.40 |
| | N | 5207.00 | 5207.00 | 5206.00 |
| | Skw | 17.12 | 26.18 | -45.32 |
| | Krt | 528.38 | 892.38 | 2138.49 |
| 2018 | M | 49740.83 | 390544.60 | -1352.78 |
| | Me | 3016.39 | 33960.98 | 8.74 |
| | SD | 413779.50 | 3083861.00 | 49397.71 |
| | N | 5591.00 | 5591.00 | 5589.00 |
| | Skw | 28.09 | 27.29 | -50.16 |
| | Krt | 1065.81 | 971.00 | 2632.01 |
| 2019 | M | 50855.08 | 431452.50 | -58668.08 |
| | Me | 3146.73 | 38920.96 | 8.48 |
| | SD | 494817.30 | 3381768.00 | 4769072.00 |
| | N | 5887.00 | 5887.00 | 5885.00 |
| | Skw | 14.71 | 25.66 | -74.90 |
| | Krt | 719.72 | 861.37 | 5712.19 |

*Mean (M), median (Me), standard deviation (sd), count (N), skewness (Skw), kurtosis (Krt)

[Author's work based on Stata output]

Continued **Table 3.** Descriptive statistics of research variables

| Year | | Gross profit / loss | Total assets | ROA |
|---|---|---|---|---|
| 2020 | M | 56632.06 | 446206.90 | -12697.39 |
| | Me | 4109.63 | 42551.99 | 9.99 |
| | SD | 415889.10 | 3342378.00 | 602966.50 |
| | N | 6197.00 | 6197.00 | 6191.00 |
| | Skw | 9.96 | 24.34 | -61.86 |
| | Krt | 502.24 | 776.52 | 4294.69 |
| 2021 | M | 73922.94 | 485838.40 | -2940.47 |
| | Me | 6007.10 | 46638.26 | 12.20 |
| | SD | 497512.80 | 3812588.00 | 277596.90 |
| | N | 6750.00 | 6750.00 | 6741.00 |
| | Skw | 27.05 | 26.41 | -33.14 |
| | Krt | 1163.30 | 908.81 | 2231.55 |
| 2022 | M | 80122.94 | 510412.50 | -3892.42 |
| | Me | 8935.56 | 52501.82 | 15.71 |
| | SD | 543534.90 | 4143179.00 | 317887.00 |
| | N | 7551.00 | 7551.00 | 7545.00 |
| | Skw | 15.74 | 26.78 | -49.28 |
| | Krt | 509.65 | 911.33 | 3540.20 |
| 2023 | M | 83721.20 | 551214.10 | 2635.68 |
| | Me | 12427.25 | 58359.63 | 18.49 |
| | SD | 560701.90 | 6341141.00 | 2205466.00 |
| | N | 7995.00 | 7995.00 | 7985.00 |
| | Skw | 24.15 | 49.76 | 55.57 |
| | Krt | 867.37 | 3143.93 | 5206.36 |
| Total | M | 50717.36 | 434091.20 | -8484.41 |
| | Me | 2838.28 | 38050.30 | 7.69 |
| | SD | 436205.40 | 4010588.00 | 1519403.00 |
| | N | 80490.00 | 80490.00 | 80448.00 |
| | Skw | 23.09 | 40.70 | -153.84 |
| | Krt | 1115.63 | 2712.86 | 42914.03 |

*Mean (M), median (Me), standard deviation (sd), count (N), skewness (Skw), kurtosis (Krt)

[Author's work based on Stata output]

## Appendix 3

**Table 4.** Skillings-Mack Test for years pairs

| | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|
| 2009 | **50.548*** ** | | | | |
| | **0.0000 (0.0000)** | | | | |
| 2010 | **110.513*** ** | **32.834*** ** | | | |
| | **0.0000 (0.0000)** | **0.0000 (0.0000)** | | | |
| 2011 | **102.081*** ** | **19.868*** ** | 0.679 | | |
| | **0.0000 (0.0000)** | **0.0000 (0.0000)** | 0.4099 (0.4550) | | |
| 2012 | **104.054*** ** | **37.883*** ** | **20.425*** ** | **6.014** ** | |
| | **0.0000 (0.0000)** | **0.0000 (0.0000)** | **0.0000 (0.0000)** | **0.0142 (0.0170)** | |
| 2013 | **41.253*** ** | **5.252** ** | **6.992*** ** | 0.015 | 1.738 |
| | **0.0000 (0.0000)** | **0.0219 (0.0240)** | **0.0082 (0.0030)** | 0.9025 (0.9050) | 0.1874 (0.1630) |
| 2014 | **52.14*** ** | **14.672*** ** | **3.631*** | 1.544 | 0.088 |
| | **0.0000 (0.0000)** | **0.0001 (0.0000)** | **0.0567 (0.0570)** | 0.214 (0.2190) | 0.7667 (0.7670) |
| 2015 | **39.206*** ** | **10.464*** ** | **4.632** ** | 3.514 | 1.710 |
| | **0.0000 (0.0000)** | **0.0012 (0.0000)** | **0.0314 (0.0370)** | 0.0609 (0.0580) | 0.1910 (0.1950) |
| 2016 | **25.98*** ** | **4.151*** | 2.267 | 1.209 | 0.136 |
| | **0.0000 (0.0000)** | **0.0416 (0.0530)** | 0.1322 (0.1160) | 0.2715 (0.2740) | 0.7123 (0.6910) |
| 2017 | **6.514** ** | 0.031 | 0.248 | 0.000 | 0.569 |
| | **0.0107 (0.0030)** | 0.8602 (0.8310) | 0.6185 (0.6290) | 1.0000 (1.0000) | 0.4507 (0.4260) |
| 2018 | **15.694*** ** | **4.892** ** | 0.374 | 0.232 | 0.288 |
| | **0.0001 (0.0000)** | **0.027 (0.0330)** | 0.5408 (0.5700) | 0.6300 (0.6260) | 0.5915 (0.6010) |
| 2019 | **31.081*** ** | **6.167** ** | 1.564 | 0.380 | 0.087 |
| | **0.0000 (0.0000)** | **0.0130 (0.0110)** | 0.2111 (0.2190) | 0.5376 (0.5580) | 0.7680 (0.7760) |
| 2020 | **27.88*** ** | **6.441** ** | 1.608 | 0.000 | 0.380 |
| | **0.0000 (0.0000)** | **0.0112 (0.0110)** | 0.2048 (0.2180) | 1.0000 (1.0000) | 0.5376 (0.5600) |
| 2021 | **17.629*** ** | 2.433 | 0.542 | 1.085 | 0.799 |
| | **0.0000 (0.0000)** | 0.1188 (0.1100) | 0.4616 (0.4530) | 0.2976 (0.2410) | 0.3714 (0.3720) |
| 2022 | **8.516*** ** | 0.02 | 0.354 | 1.600 | 3.285 |
| | **0.0035 (0.0020)** | 0.8875 (0.9020) | 0.5519 (0.5480) | 0.2059 (0.1920) | 0.0699 (0.0750) |
| 2023 | 2.122 | 0.021 | **5.843** ** | **5.524** ** | **11.215*** ** |
| | 0.1452 | 0.8848 | **0.0156 (0.0170)** | **0.0188** | **0.0008** |

| Year | Skillings Mack |
|---|---|
| | P-value (No ties) (Empirical P-value (Ties)) |

* 10 %; ** 5 %; *** 1 %

[Author's work based on Stata output]

Continued **Table 4.** Skillings-Mack Test for years pairs

| | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|
| 2009 | | | | | |
| 2010 | | | | | |
| 2011 | | | | | |
| 2012 | | | | | |
| 2013 | | | | | |
| 2014 | **12.201*****<br>**0.0005 (0.0000)** | | | | |
| 2015 | **18.700*****<br>**0.0000 (0.0000)** | **5.508****<br>**0.0189 (0.0150)** | | | |
| 2016 | **14.626*****<br>**0.0001 (0.0000)** | **9.59*****<br>**0.0020 (0.0020)** | **5.591****<br>**0.0181 (0.0130)** | | |
| 2017 | **6.618****<br>**0.0101 (0.0130)** | 1.308<br>0.2528 (0.2550) | 1.998<br>0.1575 (0.1570) | 0.333<br>0.5639 (0.5640) | |
| 2018 | **11.376*****<br>**0.0007 (0.0010)** | 0.391<br>0.5318 (0.5010) | **10.998*****<br>**0.0009 (0.0000)** | **18.04*****<br>**0.0000 (0.0000)** | **20.549*****<br>**0.0000 (0.0000)** |
| 2019 | **10.935*****<br>**0.0009 (0.0000)** | **7.38*****<br>**0.0066 (0.0090)** | **10.035*****<br>**0.0015 (0.0020)** | **16.067*****<br>**0.0001 (0.0000)** | **40.533*****<br>**0.0000 (0.0000)** |
| 2020 | **8.383*****<br>**0.0038 (0.0040)** | **5.157****<br>**0.0232 (0.0220)** | **13.759*****<br>**0.0002 (0.0000)** | **15.279*****<br>**0.0001 (0.0000)** | **33.58*****<br>**0.0000 (0.0000)** |
| 2021 | **12.53*****<br>**0.0004 (0.0000)** | 3.675<br>0.0552 (0.0630) | **11.924*****<br>**0.0006 (0.0000)** | **4.130***<br>**0.0421 (0.0540)** | **24.972*****<br>**0.0000 (0.0000)** |
| 2022 | **4.779****<br>**0.0288 (0.0230)** | 0.540<br>0.4624 (0.4710) | 0.078<br>0.7800 (0.7880) | 0.215<br>0.6429 (0.6550) | **15.883*****<br>**0.0001 (0.0000)** |
| 2023 | 0.542<br>0.4616 | 0.382<br>0.5365 (0.5120) | 0.167<br>0.6828 (0.6840) | 0.179<br>0.6722 (0.6620) | 2.04<br>0.1532 (0.1340) |

| Year | Skillings Mack |
|---|---|
| | P-value (No ties) (Empirical P-value (Ties)) |

* 10 %; ** 5 %; *** 1 %

[Author's work based on Stata output]

Continued **Table 4.** Skillings-Mack Test for years pairs

| | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|
| 2009 | | | | | |
| 2010 | | | | | |
| 2011 | | | | | |
| 2012 | | | | | |
| 2013 | | | | | |
| 2014 | | | | | |
| 2015 | | | | | |
| 2016 | | | | | |
| 2017 | | | | | |
| 2018 | | | | | |
| 2019 | **26.971*****<br>**0.0000 (0.0000)** | | | | |
| 2020 | **29.889*****<br>**0.0000 (0.0000)** | **5.263****<br>**0.0218 (0.0240)** | | | |
| 2021 | **22.408*****<br>**0.0000 (0.0000)** | **11.12*****<br>**0.0009 ( 0.0000)** | 2.750<br>0.0973 (0.1090) | | |
| 2022 | **7.564*****<br>**0.0060 (0.0070)** | 3.154<br>0.0757 (0.0700) | 2.337<br>0.1263 (0.1370) | **3.837***<br>**0.0501 (0.0480)** | |
| 2023 | 3.119<br>0.0774 (0.0670) | 0.000<br>1.0000 (1.0000) | 0.019<br>0.8904 (0.9040) | 1.113<br>0.2914 (0.2990) | 0.948<br>0.3302 (0.3060) |

| Year | Skillings Mack |
|---|---|
| | P-value (No ties) (Empirical P-value (Ties)) |

* 10 %; ** 5 %; *** 1 %

[Author's work based on Stata output]

# The rise of the gaming entrepreneurship – overview of the gaming industry ecosystem, case of Croatia

**Leon Šaško**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
lsasko20@student.foi.hr

**Ivana Fojs, Kristina Detelj**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{ivfojs, kristina.detelj}@foi.unizg.hr

***Abstract****. The paper presents the video game industry, which is among the youngest and currently the fastest growing creative industries in the world. The global development of the industry in recent decades, and its impact on the Croatian gaming market has been investigated. Statistical research of secondary data on gaming in the Republic of Croatia and gaming in the world was conducted. Through an overview of collected data, an analysis of the current state of the gaming entrepreneurship ecosystem in Croatia was conducted. The current state of the gaming industry ecosystem was evaluated by examining the regulatory framework, incentives, education system and existing networks. The tendency of further development of gaming entrepreneurship, and its necessary factors, has been estimated.*

**Keywords.** creative and cultural industry, Croatia, ecosystem, gaming entrepreneurship, metaverse, video game industry

## 1 Introduction

Although the development of video games began inthe middle of the last century, this industry has seen significant growth in this millennium, partly due to the rapid advancement of technology, partly due to the global use of smartphones and the availability of the Internet. Consumer habits have also changed, the majority of the population spends a significant part of the day in front of a screen, and generations of new consumers are connected to the virtual world from an earlier age. The rise of online gaming and the growing demand for mobile games have created new opportunities for entrepreneurs. This statement is substantiated by the fact that in 2022 it was estimated that the industry will generate 184.4 billion dollars (Newzoo, 2023). Gaming entrepreneurship is a promising trend that encourages innovation, especially through the implementation of virtual reality, augmented reality and the metaverse. For the reasons mentioned, this research was carried out, i.e. the aim of the work is to present the current state of the gaming industry in Croatia, from the entrepreneurial aspect, but also to identify the potential in the future.

## 2 Development of the gaming industry

The first computer games appeared in the 1950s, and due to the limitations of the technology at the time,video games were very simple (ITU, 2011). With the appearance of the first functional games („Nim", „Oxo", „Tennis for two" and „Spacewar", and with the invention of the microprocessor and the beginning of the production of personal computers in 1970 and1971, the first arcade games were developed („Computer Space", 1971, „Galaxy game", 1971 and „Pong",1972). The industry implies the development, distribution and sales, so despite the existence and production of video games, it was only with the sale of the first Magnavox Odyssey player console in 1972 that the gaming industry began in its full sense (Langlotz et al., 2008). In the eighties of the last century, one of the first video games made in color „Pac-man" (1980), together with „Tetris" (1984) and „Super Mario Bros" (1985) became a global phenomenon that made video games more popular, and the industry began to attract more attention. During the 1990s, the video game industry has seen significant progress in technology, with the advent of 3D graphics and CDs as a gaming medium, and the rise of game consoles (Super Nintendo, Sega Genesis, and Sony PlayStation). The availability of mobile devices has contributed to the development of the industry in the past decade, attracting millions of players worldwide to games such as „Angry Birds" (2009) and „Candy Crush Saga" (2012) (Altamirano, 2012). With the emergence of a gaming streaming platform in 2011, Twitch has significantly influenced the development of computer games, primarily due to the possibility of

interaction between players and viewers. This created a community of players and spectators, which further popularized e-sports, and gaming became recognized and accepted as a serious competitor to traditional sports (Yaden, 2022). The current big shift in the industry is made possible by the development of Virtual Reality (VR) and Augmented Reality (AR).

# 3 Gaming entrepreneurship – definition and existing literature

Activities and business opportunities related to the video game industry encompasses gaming entrepreneurship. Gaming entrepreneurship includes the development of video games, which includes programming, character and world design, art, music, and game testing. The types of gaming entrepreneurs differ depending on the activities they perform; gaming developers manufacture and sell gaming hardware, including gaming consoles, peripheral devices (mice, keyboards, controllers), gaming headsets and other accessories that improve the gaming experience, publishers are placing video games to the market. They also provide financial support, marketing strategies, distribution and support for development studios or independent developers to help their games reach a wider audience (Petrić & Kajić, 2020). Some other directions of possible entrepreneurship in the gaming industry are the provision of services such as platforms for streaming games (Sutevski, 2023), organization of e-sport events, creation of streaming content related to video games and gaming education.

Considering the rapid increase in the number of gamers, games, companies and other participants in the gaming industry, it is important to assess its effects on the economy and to identify obstacles, as well as undiscovered potential.

Seen from the entrepreneurial aspect, previous research can be categorized into five research areas:

- **The impact of gaming on entrepreneurial habits** (La Guardia et al., 2014; Newbery et al., 2016);

Transversal entrepreneurial skills are encouraged by applying the concept of game-based learning. Gaming, more precisely, serious games and simulations, when included in entrepreneurial education, can give a realistic insight of the journey of starting a business.

- **The impact of gaming on entrepreneurial education** (Thavikulwat,1995; La Guardia et al., 2014; Newbery et al., 2016; Bajraktari et al., 2021). La Guardia et al. (2014) use the concept of experiential learning, which includes gaming with the aim of encouraging entrepreneurial intentions. Bajraktari et al. (2021) proved the positive impact of games on entrepreneurial education. In addition to entrepreneurial traits and intentions, the positive influence of gaming has been proven on cognitive and affective abilities of adolescents (Zioga et al., 2024);
- **Tools and technology in game development**: forms of technical creation and development of games (Dubbels, 2016), or technology usage for improvement of user experience (Palmquist et al., 2024);
- **Enterprise success factors and the potential of the gaming industry** (Santasärkkä, 2017; Zhou et al., 2022; Primasari, 2022; Palma-Ruiz et al., 2022; Huang and Huang, 2023; Białek-Jaworska et al., 2023; Cao, 2023; Gumasing et al., 2023; Peri, 2024; Rogov, 2024). Some of the identified success factors of the gaming industry is its resilience in unpredictable times such as the pandemic, as well as constant hardware innovation and the inclusivity of the gaming content (Peri, 2024). Zhou et al. (2022) also examined the impact of Covid-19 on the gaming industry in Macao. Due to results, authors emphasized the need for portfolio diversification, expansion into new markets and cooperation with surrounding areas. Huang and Huang (2023) identified the potential for gaming industry development through the usage of ChatGPT, whose application could enhance the gaming experience and transform the gaming ecosystem. In research regarding new forms of gaming industry directions, such as of cloud gaming, Gumasing et al. (2023) point to its potential, due to the positive attitude of Generation Z towards cloud gaming, but also the general economic well-being such as sustainability, efficiency and optimization of energy and other resources. The results of a study conducted by Białek-Jaworska et al. (2023) indicated increased lagged innovation and efficiency of companies in the gaming industry, because of intellectual property protection policy. According to Rogov (2024), gaming is a specific social institution, with a still unrecognized potential;
- **Entrepreneurial ecosystem in the context of video games** (Huang, 2024; Mcintyre et al., 2023; Wang, 2024; Białek-Jaworska et al., 2023; Cao, 2023; Rogov, 2024; Hagverdiyev, 2024; Peri, 2024). Huang (2024) investigates how local cultural policies affect the establishment of businesses in the gaming industry. Tsymlyakov (2024) also examined the impact of local policies on the development of gaming industry, but concluded that governmental restrictions can have a negative effect on the gaming industry. While evaluating the effectiveness of the new gaming public policy in Macao, Wang (2024) identified positive policy changes which are visible through the focus on economic diversification, rather than on the growth of the industry. Białek-Jaworska et al. (2023) investigate the effects of government policies, specifically the intellectual property and taxation, on the efficiency of companies in the gaming industry. Additionally, some existing research focuses on the gaming industry ecosystem, but specifically on the legal perspective and security. Cao (2023) examined the potential of the metaverse development, with gaming as one of the main factors. The author suggests investments in new platforms, but also governmental policies for security

and protection, for both, companies and consumers. The stated challenge coincides with the highlighted challenges of the European Parliament (2023) in the report Protecting gamers and encouraging growth in the video games sector. Another threat for gaming entrepreneurs in existing ecosystems is the lack of policies to protect them from the influence of large corporations (Rogov, 2024). Peri (2024) emphasized security as one of the main challenges of the gaming industry, as well as Hagverdiyev (2024) who underlined the importance of governmental security and privacy policies, especially in the context of Big Data Analytics.

Given the existing literature and the observed limiting effect of policies on the development of the gaming industry, the goal of this paper is the analysis of the ecosystem in Croatia, that is, the identification of existing support instruments for entrepreneurs, as well as potential additional areas of support. Since a quality workforce is a prerequisite for gaming industry development in the future, special emphasis during the analysis will be on the education system and the popularization of gaming through competitions and tournaments.

# 4 The gaming industry ecosystem in Croatia

## 4.1 The development and current state of the gaming industry in Croatia

The industry started in the 1980's with the first commercial game („Vruće ljetovanje", 1985), and had its greater development in the 1990's with the establishment of the first computer game development studio („Croteam", today „Abest", 1993). 2008 and 2010. were milestones for Croatian gaming industry development, due to the foundation of the highly successful company Nanobit, whose revenues make up to 60% of the total revenues of the industry (Pavelić, 2022), and Gamepires, with its release of the popular survival action-adventure game "SCUM".

Currently, the number of active E-sports players in Croatia is 265 (Rajput, 2023). In addition to the general increase in the number of players, the video game industry records a significant increase in female players, so for comparison in 2006, the percentage of female players was 38%, while currently the share of female and male players is almost equal, i.e. 49% refers to female players (Qazi, 2023).

A decade ago, while still developing, the total revenues of the Croatian computer game industry from 2009 to 2013 grew at an average annual rate of 50.4 percent (Rašić Bakarić et al., 2015) as shown in Figure 1.

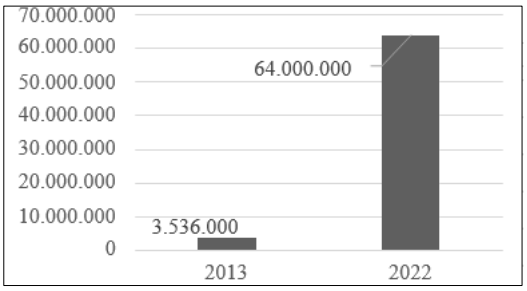



**Figure 1:** Croatian gaming industry revenue (in euro) over the last decade, source: Rašić Bakarić et al., 2015 and CGDA, 2022

Croatia gaming industry revenue in 2022 was 64 million euro (Knezović, 2023), as shown in Table 1, with comparison to Europe's 34.4 billion in revenue.

**Table 1:** Croatian gaming industry revenue compared to other European countries

| Country | | Revenue (euro) |
|---|---|---|
| 1 | France | 38.800.000,00 |
| 2 | Germany | 37.400.000,00 |
| 3 | Finland | 32.000.000,00 |
| 4 | Sweden | 27.130.000,00 |
| 5 | Spain | 12.810.000,00 |
| 6 | Poland | 11.910.000,00 |
| 7 | Netherlands | 430.000,00 |
| 8 | Romania | 293.000,00 |
| 9 | Czech Republic | 290.000,00 |
| 10 | Lithuania | 154.000,00 |
| 11 | Belgium | 83.000,00 |
| 12 | Slovakia | 69.000,00 |
| 13 | **Croatia** | **64.000,00** |
| 14 | Slovenia | 32.000,00 |
| 15 | Latvia | 24.000,00 |
| 16 | Portugal | 22.000,00 |

Source: Knezović (2023).

According to Pavelić, there were about 45 legal entities in Croatia in 2022, that were engaged in the development of video games, while a decade ago, there were only 10 legal entities, according to the research conducted by The Institute of Economics, Zagreb. The number of employees in Croatian gaming industry is also growing significantly (Rašić Bakarić et al., 2015).

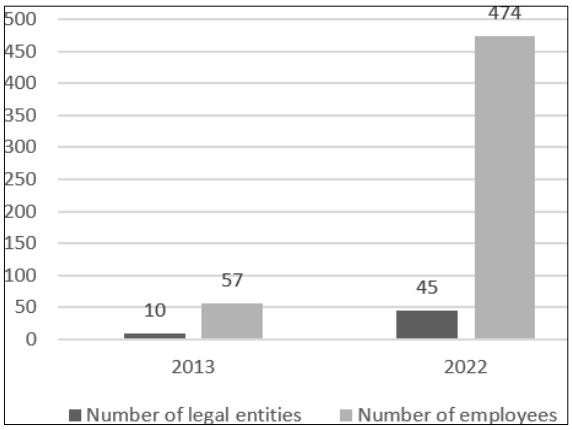



**Figure 2:** Rise of the number of legal entities and employees over the last decade, source: Rašić Bakarić et al., 2015 and CGDA, 2022.

**Table 2:** Top 10 enterprises by revenue in Croatia

| Name of enterprise | Revenue in mil $ |
|---|---|
| **Nanobit d.o.o.** | **36,08** |
| Pine studio d.o.o. | 7,83 |
| Gamepires d.o.o. | 4,42 |
| Abest d.o.o. | 3,07 |
| Orqa d.o.o. | 2,84 |
| DEV d.o.o. | 1,66 |
| Diversitas IT sustavi d.o.o. | 1,66 |
| Fury studios d.o.o. | 1,49 |
| Protopixel d.o.o. | 0,96 |
| Poster d.o.o. | 0,95 |

Source: CGDA (2022).

Table 2 shows the revenues of top ten Croatian gaming companies. Total number of employees in the industry in 2021 was 474 (CGDA, 2022). Share of employed women in the gaming industry in Croatia is 30%, which is above average comparing to countries in region (Slovenia 41%, Serbia 30%, Romania 30%, Germany 25%, Czech Republic 19%, Slovakia 17,4%).

## 4.2. Gaming ecosystem in Croatia

The gaming ecosystem in Croatia encompasses different elements that create the rules within the industry and dictate the directions of its development. The entities that are active in this ecosystem work according to the set of se regulatory framework, incentives, educations system and networks. For the clarity of its description, we have organized the ecosystem elements into the Table 3, and the relevant sources are cited within the table.

**Table 3:** Croatian gaming ecosystem

**Regulation, policies, documents**

- Competent institution: Ministry of Entrepreneurship and Crafts;
- The Law on Audiovisual Activities („Narodne novine“ No. 61/18, NN 114/22), which entered into force in 2018 (on the initiative of the Ministry of Culture and Media), includes the production of video games by audiovisual activities, which are additionally defined in Article 8 and Article 25;
- National Classification of Activities (NKD): Sub-sectors of creative and cultural industries according to NKD 2007, 58.21 Publishing of computer games;
- Rulebook on the procedure, criteria and deadlines for the implementation of the National Program for the Promotion of Audiovisual Creativity („Narodne novine“ No. 95/23), on entrepreneurship and crafts;
- Mapping creative and cultural industries in the Republic of Croatia (2015), The institute of economics, Zagreb, Croatian cluster of competitiveness of creative and cultural industries (Rašić Bakarić et al., 2015);

**Public calls for grants**

- The public call to encourage the development and production of video games: "Transformation and strengthening the competitiveness of cultural and creative industries" under the National Recovery and Resilience Program (Ministry of Culture and Media of the Republic of Croatia, 2024);
- The Croatian Audiovisual Center (HAVC) published first call for projects in 2021 to support the development of the video game industry in Croatia, with a budget of approximately 100,000 euros;
- Creative Europe MEDIA strand

**Financing**

- Creative Europe MEDIA strand
- Crowdfunding
- Venture capital

**Educational programs**

**1) High school**

Major: video game technician

- Technical School Sisak
- Secondary School in Novska

**2) Higher education (university)**

- Undergraduate study - Development of computer games, College of Algebra
- Undergraduate study - Information and Business Systems (IPS), a compulsory subject "Development of computer games", Faculty of Organization and Informatics, University of Zagreb

**3) Vocational**

- Two-year education: Game artist, United POP, Academy of Music, Media and the Arts, Zagreb

| |
|---|
| **Support infrastructure for entrepreneurship (institutions)**<br>Croatian entrepreneurial gaming incubator „PISMO“, City of Novska, with a campus extension planned (studio, student dormitory, business incubator for E-sport arena with 4000 seats and Gaming Industry Accelerator) |
| **Federations, associations, networks**<br>• Croatian eSPORTS Federation, member of the International Esports Federation (IESF).<br>• CGDA, Croatian game development alliance |
| **Competitions and tournaments, conferences, fairs**<br>• E-sport competition A1 Adria League, competition that brings together players from Croatia, Serbia, Macedonia, Slovenia, Bosnia and Herzegovina, Montenegro and Albania<br>• OGA Dota Pit tournament<br>• Odyssey Masters<br>• University (faculty) league<br>• Student Esports Tournament, since 2021 – 16 Croatian faculties are taking part in the competition (Faculty of Electrical Engineering and Computing in Zagreb, Faculty of Electrical Engineering, Mechanical Engineering and Shipbuilding in Split, Technical Polytechnic in Zagreb, Faculty of Organization and Informatics, University Algebra, Faculty of Electrical Engineering, Computing and Information Technologies Osijek, Polytechnic in Bjelovar, Jurja Dobrila University in Pula, Faculty of Mechanical Engineering and Shipbuilding, University of the North, Department for Mathematics, J. J. Strossmayer University in Osijek, Faculty of Science and Mathematics in Split, Faculty of Engineering, University of Rijeka, Polytechnic Velika Gorica, Faculty of Transport Sciences at the University of Zagreb and the Faculty of Philosophy in Osijek.<br>• Zagreb GameDev Meetup<br>• Gaming center „Hall of Game”<br>• The largest regional video game fair Reboot (InfoGamer fair, since 2011) |

Source: authors' work

# 5 Analysis and discussion

Based on the presented data, a significant growth of the gaming industry in Croatia is visible. Along with the increase in income on average by 50% per year, the development is visible in the number of employees and the number of companies, especially if we look at the last decade. In addition to the above, it is good to mention the presence of the female population among players, as well as employees, given that the observed industry is still considered predominantly male. The most successful Croatian companies in the gaming industry are known on the international scene, so there is no shortage of investment funds. There are formal and informal educational programs aimed at creating qualified personnel. However, opportunities for improvement were identified.

The share of exports in the total revenues of the industry is higher than 80%, which is a significant success considering the competition in the case of this truly globalized market, but it results in a double taxation, the problem that entrepreneurs cite as one of the most significant obstacles (Rašić Bakarić et al., 2015). Besides the most common threat in the video game industry, which is consumer protection and transparency (Borghese, 2023), developer protection is another risk that needs to be handled. Video game producers are generally not sufficiently familiar with the issues of intellectual property protection, and often do not have the resources to implement procedures for the protection of rights from infringement (WIPO, 2022). Globalization, with its opportunities, increased both competition and potential saturation. It is difficult for novice entrepreneurs to stand out on the global market, which is why additional actions of local authorities in the form of incentives, monetary as well as non-monetary, such as education, promotion, counselling and networking assistance, are necessary.

After considered obstacles and challenges, recommendations for a more effective gaming ecosystem are recognized:

- double taxation regulation
- governmental policies for protection of consumers and developers
- governmental funding (monetary incentives and state subsidies)
- education and cooperation with industry
- marketing – small Croatian enterprises
- inclusivity, but also heterogeneity – knowing your audience
- usage of AI
- promotion of healthy gaming
- use of talented employees
- strong public support for the industry.

Identified recommendations are comparable to recommendations identified in a study by Goldani & Goldani (2018), based on Finland's good practice gaming ecosystem, but adjusted for Iranian gaming industry. Governmental policies should not limit, for example through tax policies, but encourage the industry development. Policies should also focus on security and protection, for both companies and consumers. General public perceptions about the positive effects of gaming on mental health and education needs to be changed, since adequate games can contribute to creativity, resourcefulness, problem solving, and inclusion through group play. The same can be achieved with local and national projects promoting healthy gaming. Networking and visibility are important for entrepreneurs, therefore the establishment of more centralized places such as the aforementioned incubator in Novska, or clusters where new entrepreneurs can turn to and find support, are recommended.

One of the ways to build a sustainable gaming industry is the additional analysis of existing successful companies and

projects and their promotion as an example of good business model practice, user experience, marketing and sustainability. Recommendations can be applied on local, aswell as national and regional governmental level.

# 6 Conclusion

The positive impact of the gaming industry on economic growth is seen not only through revenue, employment and international business, but alsothrough innovation. Santasärkkä (2017) emphasizes how gaming industry players can be catalysts for innovation and knowledge-based growth, where knowledge spillover is an important factor of economic development. Further development is inevitable and also desirable, but with additional regulations to protect both users and developers, as well as entrepreneurs who are just starting out. The cooperation of all actors of the ecosystem (companies, government, entrepreneurial support institutions, educational system, society, networks) can contribute to the efficiency of the industry and better inclusivity.

# References


Altamirano, A. (2012). History of Video Games. https://www.scribd.com/document/109372526/His

Bajraktari, F. & Xiong, F. & Drieschner, C. & Möslein, V. & Ströbl, B. & Utesch, M. & Krcmar, H. (2021). A Serious Gaming Approach for Teaching Environmental Entrepreneurship. 1310-1317. 10.1109/EDUCON46332.2021.9454148.

Białek-Jaworska, A. & Teterycz, A. & Sichel, R. & Woźniak, M. (2023). IP box effects in the gaming industry. Central European Management Journal. 31. 10.1108/CEMJ-12-2021-0143.

Borgese, M. L. (2023). European Strategy for Regulating Online Video Games and New Digital Markets: A Comparison of Emerging Opportunities and the Current Italian Regulatory Framework. Athens Journal of Law - Volume 9, Issue 4, October 2023 Pages 667-676 https://doi.org/10.30958/ajl.9-4-10 doi=10.30958/ajl.9-4-10

Cao, Y. (2023). Microsoft's Layout in the Gaming Industry and Metaverse: Starting from Blizzard. Advances in Economics, Management and Political Sciences. 36. 99-103. 10.54254/2754-1169/36/20231791.

CGDA. (2022). Analiza industrije videoigara za 2022. godinu - CGDA (2022). Croatian game development alliance: https://cgda.eu/data-item/analiza-industrije-2022/

Croatian eSPORTS Federation, https://www.hess.hr/en/

Dubbels, B. (2016). Transforming Gaming and Computer Simulation Technologies across Industries. 10.4018/978-1-5225-1817-4.

European Parliament (2023). Protecting gamers and encouraging growth in the video games sector, available on https://www.europarl.europa.eu/news/en/press-room/20230113IPR66646/protecting-gamers-and-encouraging-growth-in-the-video-games-sector

Goldani, M. H., Goldani, A. (2018). A review study on effective factors of developing the Finnish gaming industry and some suggestions for Iran's game industry. 2ND NATIONAL AND 1ST INTERNATIONAL DIGITAL GAMES RESEARCH CONFERENCE: TRENDS, TECHNOLOGIES, AND APPLICATIONS (DGRC). Page123-127

Gumasing, M. J. & Alonzo, R. & Nazareno, J. & Guinto, K. (2023). Antecedents of Cloud Gaming Acceptance among Gen Z: Achieving Sustainability in the Digital Gaming Industry. Sustainability. 15. 9189. 10.3390/su15129189.

Hagverdiyev, F. (2024). DATA GOVERNANCE IN GAMING INDUSTRY. Problems of Information Technology. 15. 44-51. 10.25045/jpit.v15.i1.06.

HAVC. (2024). Razvoj i proizvodnja videoigara u 2023. https://havc.hr/o-nama/javni-pozivi/razvoj-i-proizvodnja-videoigara/razvoj-i-proizvodnja-videoigara-u-2023

Huang, G. (2024). Exploring Cultural Policy and Gaming Entrepreneurship in Shanghai: An Entrepreneurial Ecosystem Perspective. 10.1007/978-3-031-41504-3_5.

Huang, J. & Huang, K. (2023). ChatGPT in Gaming Industry. 10.1007/978-3-031-45282-6_9.

ITU. (2011). Trends in video games and gaming. https://www.itu.int/dms_pub/itu-t/oth/0B/15/T0B1500000F3301PDFE.pdf

Knezović, G. (2023). Snažan rast gejming industrije u Hrvatskoj. https://mreza.bug.hr/snazan-rast-gajming-industrije-u-hrvatskoj/

La Guardia, D. & Gentile, M. & Dal Grande, V. & Ottaviano, S. & Allegra, M. (2014). A Game based Learning Model for Entrepreneurship Education. Procedia - Social and Behavioral Sciences. 141. 195–199. 10.1016/j.sbspro.2014.05.034.

Langlotz, A. C. H.; Rhode, M.; Whaley, C. (2008) Video Games Industry Overview: An Analysis of the Current Market and Future Growth Trends. International Business Project.

Mcintyre, P. & Fulton, J. & Kerrigan, S. & Meany, M. (2023). Entrepreneurship in Gaming and IT. 10.1007/978-3-031-19455-9_13.

Ministry of Culture and Media of the Republic of Croatia, 2024, https://min-kulture.gov.hr/en

Newbery, R. & Lean, J. & Moizer, J. (2016). Evaluating the impact of serious games: the effect of gaming on entrepreneurial intent. Information Technology & People. 29. 733-749. 10.1108/ITP-05-2015-0111.

Newzoo (2023). The Games Market in 2022: The Year in Numbers. https://newzoo.com/resources/blog/the-games-market-in-2022-the-year-in-numbers

Palma-Ruiz, J., M., Torres-Toukoumidis, A., Gonzalez-Moreno, S., E., German Valles-Baca, H.(2022.) An overview of the gaming industry across nations: using analytics with power BI to forecast and identify key influencers

Palmquist, A., Jedel, I. and Goethe, O. (2024). Assistive Technologies for Attainable Gaming Experiences. 10.1007/978-3-031-30595-5_4.

Pavelić, M. (2022). Prva velika analiza CGDA: Hrvatske gaming tvrtke u 3 godine uprihodile 1,3 miljarde kuna". https://www.netokracija.com/analiza-cgda-hrvatska-gaming-industrija-igre-203789

Peri, J. (2024). Video gaming industry in the US. International Journal of Science and Research Archive. 11. 1257-1265. 10.30574/ijsra.2024.11.1.0194.

Primasari, C. H. (2022). Strategy for achieving IT-business alignment in gaming industry in Indonesia, Procedia Computer Science, Volume 197, Pages 469-476, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.12.163.

Qazi, A. (2023). HOW MANY GAMERS ARE THERE IN THE WORLD IN 2023? (LATEST GAMER STATISTICS). https://pcguide101.com/security/gamers-statistics/

Rajput, A. (2023). Metaverse in gaming – How it is Powering the gaming industry. https://300mind.studio/blog/metaverse-gaming/

Rašić Bakarić, I., Bačić, K., Božić, LJ. (2015). Mapiranje kreativnih i kulturnih industrija u Republici Hrvatskoj. Ekonomski institut, Zagreb

Rogov, I. (2024). Gaming: Between Science and Media. Humanities and Social Sciences. Bulletin of the Financial University. 13. 54-61. 10.26794/2226-7867-2023-13-6-54-61.

Santasärkkä, S. (2017.) The Digital Games Industry and its Direct and Indirect Impact on the Economy. Case study: Supercell and Finland. Helsinki Metropolia University of Applied Sciences, Metropolia Business School. https://www.theseus.fi/handle/10024/128453

Sutevski, D. (2023). Tech & Gaming Entrepreneurs: How to make money while playing. iz https://www.entrepreneurshipinabox.com/20780/tech-gaming-entrepreneurs-how-to-make-money-while-playing/

Thavikulwat, P. (1995). Computer-Assisted Gaming for Entrepreneurship Education. Simulation & Gaming. 26. 328-345. 10.1177/1046878195263005.

Tsymlyakov, A. (2024). Analysis of the Impact of the System Restrictions Imposed by the Chinese Regulator on the Key Performance Indicators of the Chinese Gaming Industry. Теоретическая и прикладная экономика. 39-52. 10.25136/2409-8647.2024.1.70010.

Wang, C. (2024). New developments in Macao's gaming public policy: implications and effectiveness. Asian Education and Development Studies. 10.1108/AEDS-08-2023-0107.

WIPO (2022). Copyright Infringement in the Video Game Industry, https://www.wipo.int/edocs/mdocs/enforcement/en/wipo_ace_15/wipo_ace_15_4.pdf

Yaden, J. (2022). What is Twitch? Everything you need to know about the top live-streaming platform? https://www.digitaltrends.com/gaming/what-is-twitch/

Zhou, F. & Schrier, T. & Zheng, T. (2022). Covid-19 And Its After-Effects On Macau's Gaming Industry. International Journal of Business & Management Studies. 03. 44-52. 10.56734/ijbms.v3n12a3.

Zioga, T. & Nega, C. & Roussos, P. & Kourtesis, P. (2024). Validation of the Gaming Skills Questionnaire in Adolescence: Effects of Gaming Skills on Cognitive and Affective Functioning. European Journal of Investigation in Health Psychology and Education. 14. 722-752. 10.3390/ejihpe14030048.

# More than twenty years of the eXtensible Business Reporting Language

**Zrinka Lacković Vincek, Suzana Keglević Kozjak, Vladimir Kovšca**

University of Zagreb Faculty of Organization and Informatics

Department of Economics

Pavlinska 2, 42 000 Varaždin, Croatia

{zlackovi, skozjak, vkovsca}@foi.unizg.hr

***Abstract.*** *eXtensible Business Reporting Language (XBRL) is being considered as the future language for business and financial reporting, aiming to provide precise, dependable, and timely financial and non-financial information to both internal and external stakeholders. Therefore, the aim of this paper is to study the scientific foundations of a research field that has been developing for about 20 years. Through bibliometric analysis methods, we will examine the conceptual and intellectual foundations underpinning the existing knowledge in the scientific literature on XBRL, identifying the main research trends and future directions for new research, which will delve deeper into various aspects of XBRL adoption.*

**Keywords.** XBRL, extensible business reporting language, accounting information systems, Internet financial reporting, bibliometric analysis

## 1 Introduction

With the growth in the availability of scientific literature, there is an increasing need for summarizing and critically analyzing specific research topics, as well as for consolidating, organizing, and synthesizing accumulated knowledge (Kunisch et al., 2018). Literature reviews are fundamental mechanisms for providing insights into the scientific foundations of a particular field of study, initiating new debates, advancing scientific discourse, and guiding future research directions. Linnenluecke et al. (2020) caution that the exponential growth of scientific literature results in an increase in low-quality research papers, particularly with the emergence of so-called predatory journals in open access, which are easily accessible. Therefore, finding relevant literature is more challenging than ever in terms of selecting sources and evidence on which to build and advance the research field. A highly useful form of literature review is bibliometric analysis, which summarizes the bibliometric and intellectual structure of a research field and analyzes relationships among various research components. Therefore, this paper will clarify the theoretical framework of bibliometrics as a statistical analysis of metadata from scientific papers aimed at assessing and predicting trends in the development of a specific scientific field, and present a systematic, transparent, and reproducible overview of the literature. In addition to the above, this paper will introduce advanced techniques that go beyond bibliometrics and complement scientific or bibliometric mapping methods through visualization, i.e., graphical presentation of analysis results. These techniques are based on graph theory, social network analysis, clustering, etc. The subject of the bibliometric analysis will be the concept of Extensible Business Reporting Language (XBRL), an essential topic in contemporary research and development of accounting information communication technologies. The aim of this paper is to explore the concept of XBRL from both theoretical and applied perspectives, synthesizing various research viewpoints on the subject. Furthermore, the goal was not to limit the range of literature over time in order to track the development of scientific thought from its initial appearance in the academic literature.

The implementation of new accounting information technologies, such as XBRL is a complex and context-specific process. Companies whose securities are traded on a regulated market are required to prepare financial statements in accordance with International Financial Reporting Standards.

Organizations may choose to implement XBRL to reduce the costs associated with regulatory compliance and enhance their competitive standing with potential investors. XBRL, which is rooted in XML, simplifies the automated production of

financial data. This collaborative initiative spans the globe, involving professionals from the accounting field, regulators, and businesses. Using information for wright purpose seems to be an important aspect evolving the technologies with application in accounting, whether exploring ERP systems creating an efficient data flow, or standard software to reduce the complexity of communication.

Examples of successful XBRL implementation in prominent countries like the USA, UK (Abdullah et al., 2009), Spain, Italy, and Brazil (Backus & Leiser,2021) underscore the advantages of XBRL, particularly in cases where regulators and governments collaborate to share data. Research findings also suggest that voluntary adoption rates fluctuate unless there is a mandate for adoption imposed by a reporting agent or organization.

Despite XBRL's slow adoption rate, it is steadily gaining worldwide recognition. Consequently, it's understandable that researchers, industry professionals, and governments continue to prioritize research on XBRL adoption. Moreover, there is an expectation of more studies in this field (Borgi, 2022).

The paper is structured into the following sections: After the introduction, Section 2 provides an overview of the theoretical knowledge on the given topic. Section 3 precisely describes the research methodology, while Section 4 presents the results and findings of the conducted research. Finally, Section 5 provide a discussion, concludes the paper and proposes avenues for future research.

# 2 Literature review

Bibliometric analysis is a quantitative, systematic review of scientific literature aimed at assessing the scientific progress in a particular research area (Ruhanen et al., 2015).

Over two decades of XBRL literature can reveal trends, advancements, and unresolved issues (Roohani et al., 2010). For example, a study by Locke & Lowe (2007) focuses on the development of XBRL using concepts drawn from a model of successful open source projects. The other studies examine the impact of XBRL adoption in different regions or countries, for example, Steenkamp & Nel (2012) in South Africa; Lakovic et al. (2019) in Montenegro; Liu et al. (2014) in PR China; Liu, Wang, & Yao (2014) in the USA; Bai, Sakaue & Takeda (2014) in Japan. In their research, Shan & Troshani (2014) confirmed the thesis that the application of XBRL can facilitate the auditing process of financial statements, as evidenced by lower auditing costs. Some studies like Gatea, Alnawas & Gali (2021), Ib, Jide & Zik-Rullahi (2016) and Chen (2012) highlighted XBRL's impact on financial reporting and its role in enhancing the transparency of information in financial statements. The authors found that XBRL adoption leads to a reduction in information asymmetry between managers and investors, thereby enhancing the overall transparency of financial statements. XBRL adoption improves the quality of financial disclosures and allows easier comparison of financial data across different jurisdictions.

Given the absence of bibliometric analysis on XBRL publications, this paper aims to explore the concept of XBRL from both theoretical and applied perspectives, synthesizing various research viewpoints on the subject. A study by Aidi, & Rosli (2019), for example, analyze and report the 370 documents related to XBRL, all of which were sourced from the Scopus database gathered from 2001 to 2019. A study by Erkus & Chiu (2014) presents an analysis of citation counts, co-citations, and publications based on 49 articles sourced from four databases: EBSCO Host, ISI Knowledge Web, Science Direct, and the Social Science Research Network. A study by El Ansary & Oubrich (2016) report the 113 articles related to XBRL based on the ProQuest ABI/Inform, Science Direct and Springer Link database gathered from 2000 to 2014. Study by Bartolacci et al. (2020) covered Scopus database and they selected only 142 articles published in peer-reviewed journals in English between 2001 and 2019, focusing primarily on the consequences of XBRL implementation on financial reporting. Uyob, Saad & Ahmi (2019) analyzed 86 articles published only in English with the specific focus on the impact of XBRL. The data were gathered from Google Scholar and Scopus database starting from the year 2003 until 2018. As a result, this study will expand upon the findings by utilizing the current sources and tools available.

# 3 Methodology

This analysis of this study covers papers from the Web of Science core collection database, as it is regarded as the most selective database, ensuring that the records in WoS are presumed to be of the highest quality research papers (Lacković Vincek et al., 2023), with the Visualization of Similarities (VOSviewer) program used as the analytical tool.

The search terms used to recognize the most relevant documents included “XBRL” or “extensible business reporting language” as keywords in the title or abstract. The retrieval date was February 1st, 2023. The database analysis was conducted using the free VOSviewer software program, which is specifically designed for bibliometric analysis. VOSviewer generates a map where the relatedness of variables is represented by the distance between them. The shorter the distance between the observed variables, the stronger the connection between them,

which is based on the "visualization of similarities" (VOS) approach (Goyal & Kumar, 2021).

Bibliometric analysis is a scientifically computer-assisted method that identifies core research areas or authors and their interrelationships, correlating all publications related to a specific research topic (de Bellis, 2009). The application of bibliometric analysis enables researchers to explore elements of research, trends in specific fields, intellectual structures within research domains, journal and article performance, and collaboration patterns (de Bruyn et al., 2023). Bibliometric analysis is applied as a specialized method for analyzing large volumes of scientific data across various research fields, including accounting.

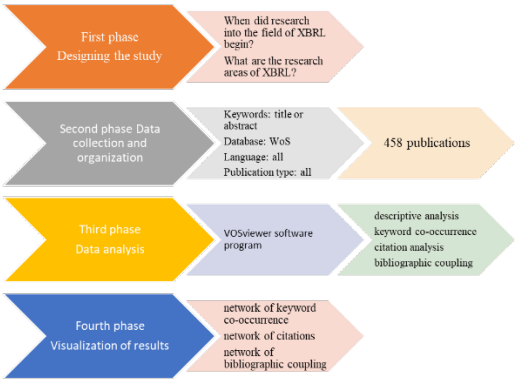



**Figure 1.** Stages of conducting bibliometric analysis on XBRL

Source: Author.

According to the specified criteria, a total of 458 publications were obtained across the entire WoS database in all categories, and they are considered for further analysis. The analysis of filtered documents was conducted using bibliometric mapping in the VOSviewer software. Descriptive analysis, keyword co-occurrence analysis, citation analysis, and bibliographic coupling analysis were applied (Fig. 1).

# 4 Results and findings

In the Web of Science core collection database, a total of 458 publications mentioning XBRL in the title or abstract were found from 2002 to 2023, which included 320 (69.87%) original research article, 110 (23.71%) proceeding papers, 10 (2.16%) book chapters and editorial materials, and 8 (1.72%) review articles (Fig. 2).

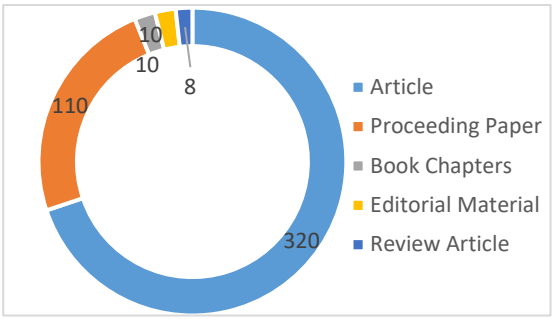



**Figure 2.** Publication types

Source: Own elaboration based on WoS 2023 data.

The dataset's published documents were also analyzed according to the annual publication frequency. The first publications on XBRL were published by Elliott (March 2002), as well as Nutz & Strauss (October 2002). According to the data presented in Fig. 3, it is evident that no publications on the topic of XBRL were published in 2003, followed by a noticeable slow growth until 2009. The number of published papers increased each year, with the highest number published in 2015, 2017, and 2018 (more than 40 papers per year). After 2021, there was a sudden decline again in the number of published papers on the topic of XBRL, which is surprising considering the fact that the European regulatory framework requires sole proprietorships with securities traded on a regulated market to fully prepare their annual financial statements in Extensible Markup Language format for reporting periods beginning on or after January 1, 2021 (Miścikowska, 2021).

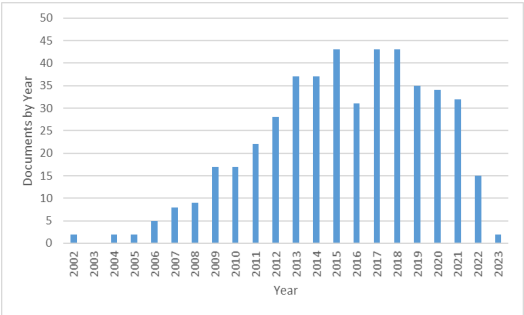



**Figure 3.** Annual scientific production on XBRL
Source: Own calculation based on WoS 2023 data

As shown in Table 1, nearly all the publications (443, 96.77%) were written in English. Three publications were published in Chinese, German, Portuguese, and Spanish languages each, two publications in Russian, and one publication in Arabic.

**Table 1.** Languages used for publication

| Language | No. of publications | Percentage (%) |
|---|---|---|
| English | 443 | 96.77 |
| Chinese | 3 | 0.65 |
| German | 3 | 0.65 |
| Portuguese | 3 | 0.65 |
| Spanish | 3 | 0.65 |
| Russian | 2 | 0.43 |
| Arabic | 1 | 0.22 |
| **Total** | **458** | **100.00** |

Source: Own elaboration based on WoS 2023 data.

Keyword analysis using the VOSviewer was conducted to investigate the predominant themes in XBRL. A total of 1121 keywords were identified in 458 publications. Fig. 4 displays a network visualization of author keywords, employing color, circle size, font size, and line thickness to illustrate their relationships with other keywords. (Sweileh et al., 2017).

According to the analysis in this study, eight clusters in XBRL research have been identified based on author keywords. The first red cluster is related to XBRL, corporate disclosure, e-government, extensible business reporting, financial information, internet, sustainability and transparency. The second green include the keywords of adoption, digital reporting, extensible business reporting, institutional theory, malaysia, standard business reporting and xbrl.

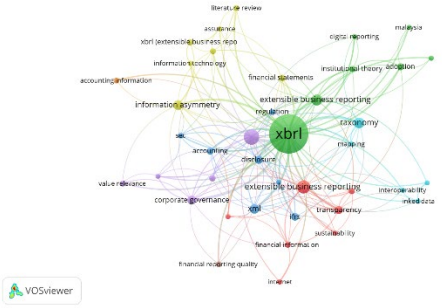



**Figure 4.** Keyword co-occurrence

Source: Own mapping in VOSviewer based on the WoS 2023 data.

The most frequently appearing keywords were financial reporting, taxonomy, information asymmetry, XML, corporate governance, transparency, disclosure and information systems (Table 2).

**Table 2.** Top ten keywords of XBRL based on the occurrence

| Author keywords | Occurrences | Percentage (%) |
|---|---|---|
| Financial Reporting | 40 | 25.81 |
| Taxonomy | 19 | 4.09 |
| Information Asymmetry | 16 | 3.45 |
| XML | 16 | 3.45 |
| Corporate Governance | 14 | 3.02 |
| Transparency | 11 | 2.37 |
| Disclosure | 11 | 2.37 |
| Information Systems | 10 | 2.16 |
| Accounting | 9 | 1.94 |
| Adoption | 9 | 1.94 |

Source: Own elaboration based on WoS 2023 data.

In addition to keywords, it is important to examine the occurrence count in WoS subject categories, which are listed in Table 3, to understand the research area. Research areas are assigned to journals and define the scientific focus of the journal. A journal can be assigned multiple subject categories depending on its scientific orientation. As expected due to the field to which XBRL belongs, which is primarily accounting and auditing, business economics ranks first (65.73%), followed by computer science (20%).

**Table 3.** Number of publications in various research areas

| Research area | No. of publications | Percentage (%) |
|---|---|---|
| Business Economics | 305 | 65.73 |
| Computer Science | 92 | 19.83 |
| Engineering | 30 | 6.47 |
| Information Science Library Science | 22 | 4.74 |
| Operations Research Management Science | 13 | 2.80 |
| Social Sciences Other Topics | 13 | 2.80 |
| Public Administration | 12 | 2.59 |
| Education Educational Research | 9 | 1.94 |
| Science Technology Other Topics | 9 | 1.94 |
| Telecommunications | 9 | 1.94 |

Source: Own elaboration based on WoS 2023 data.

Table 4 presents the leading journals on XBRL. The highest-ranking journal was the *Journal of Information Systems,* with 44 publications and an impact factor (IF) of 2.308.

**Table 4.** Overview of leading journals publishing on XBRL

| Journal name | Publisher | Document type | Total |
|---|---|---|---|
| Journal of Information Systems | American Accounting Association | Journal | 44 |
| Journal of Emerging Technologies in Accounting | American Accounting Association | Journal | 18 |
| International Journal of Accounting Information Systems | Elsevier B. V. | Journal | 16 |
| International Journal of Accounting and Information Management | Emerald Publishing Limited | Journal | 13 |
| International Journal of Disclosure and Governance | Palgrave Macmillan | Journal | 11 |
| Lecture Notes in Computer Science | Springer Nature | Book Series | 10 |
| Issues in Accounting Education | American Accounting Association | Journal | 8 |
| Accounting Horizons | American Accounting Association | Journal | 7 |
| International Journal of E Business Research | IGI Global | Journal | 6 |
| Decision Support Systems | Elsevier B. V. | Journal | 5 |

Source: Own elaboration based on WoS 2023 data.

Table 5 provides an overview of published papers on the topic of XBRL by country. The United States ranks first in terms of the number of papers in this scientific field, with a share of 41.03 % (183 publications), followed by China (72 publications, 15.52%), and Australia (40 publications, 8.62%).

**Table 5.** Overview of publications on XBRL by country

| Country | No. of publications | Percentage (%) |
|---|---|---|
| USA | 183 | 41.03 |
| China | 72 | 15.52 |
| Australia | 40 | 8.62 |
| England | 23 | 4.96 |
| Canada | 22 | 4.74 |
| Italy | 22 | 4.74 |
| Germany | 20 | 4.31 |
| Spain | 16 | 3.45 |
| Indonesia | 13 | 2.80 |
| Malaysia | 12 | 2.59 |
| Netherlands | 12 | 2.59 |
| Taiwan | 11 | 2.37 |

Source: Own elaboration based on WoS 2023 data.

Table 6 shows the leading organizations publishing on XBRL. The most active organization working on XBRL is Rutgers State University at New Brunswick, with the total of 21 publications on XBRL (13.55%). Next on the list are California State University System (17 publications, 3.66%) and The University of Hawaii at Manoa, as well as The University of Hawaii System (each with 16 publications, 3.45%). From the analysis presented, it is evident that all organizations operate within the United States, and research on XBRL is primarily focused on the western side of the world, highlighting a gap between research in the USA and other parts of the world.

**Table 6.** Leading organizations publishing on XBRL

| Organization | No. of publications | Percentage (%) |
|---|---|---|
| Rutgers State University at New Brunswick | 21 | 13.55 |
| California State University System | 17 | 3.66 |
| University of Hawaii at Manoa | 16 | 3.45 |
| University of Hawaii System | 16 | 3.45 |
| Rutgers State University at Newark | 14 | 3.02 |
| State University System of Florida | 13 | 2.80 |

| | | |
|---|---|---|
| University of Arkansas at Fayetteville | 13 | 2.80 |
| University of Arkansas System | 13 | 2.80 |
| University of Adelaide | 12 | 2.59 |
| Bentley University | 10 | 2.16 |
| Delft University of Technology | 10 | 2.16 |

Source: Own elaboration based on WoS 2023 data.

Table 7 listed the most active authors with more than five publications. Among them, Troshani, I. from the University of Adelaide has authored 12 papers predominantly focused on digital reporting. Except him, Farewell, S. M. and Debreceny, R. S. have published more than ten publications on XBRL.

**Table 7.** Top authors publishing on XBRL

| Author name | No. of publications | Percentage (%) |
|---|---|---|
| Troshani, I. | 12 | 13.04 |
| Farewell, S. M. | 11 | 2.37 |
| Debreceny, R. S. | 10 | 2.16 |
| No, W. G. | 9 | 1.94 |
| Vasarhelyi, M. A. | 7 | 1.51 |
| Felden, C. | 7 | 1.51 |
| Boritz, J. E. | 6 | 1.29 |
| Rohde, F. | 6 | 1.29 |
| Wang, T. | 6 | 1.29 |
| Janssen, M. | 6 | 1.29 |
| Alles, M. G. | 6 | 1.29 |
| Wang, D. | 6 | 1.29 |

Source: Own elaboration based on WoS 2023 data.

Fig. 5 shows the bibliographic coupling of authors on the topic of XBRL. Bibliographic coupling occurs when two papers cite a common third paper in their bibliographies. This indicates that there is a likelihood that the two papers address a similar topic. Two documents are bibliographically coupled if they both cite one or more common documents. This method is applied as a complement to co-citation analysis. While co-citation analysis shows that two papers appear together in the reference list of another paper, bibliographic coupling counts the number of references that a group of documents shares together. The diagram illustrates a close collaboration between Debreceny, R. and Farewell, S., who frequently work together on researches, as well as Kim, J. W. and Lim, J. H. Additionally, Vasarhelyi, M. A. collaborates with a group of authors colored in blue, and No, W. G. collaborates with authors colored in purple.

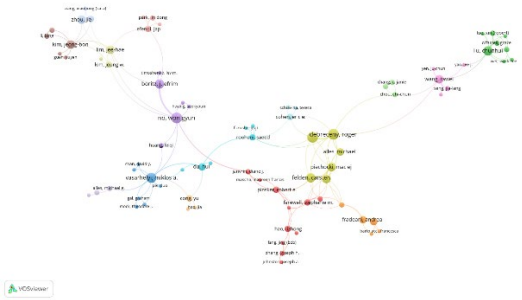


**Figure 5.** Bibliographic coupling of authors on the topic of XBRL

Source: Own mapping in VOSviewer based on the WoS 2023 data.

Fig. 6 shows the bibliographic coupling of co-authorship between countries. Nodes represent countries based on the affiliation addresses of authors, and the thickness of the network lines represents the strength of collaboration (De Iuliis et al., 2024). The two largest nodes are the USA and China. The main partners of the USA are Taiwan, Japan and South Korea (colored green). China has worked closely with Canada and Singapore (colored pink), while England has collaborated with Finland, Jordan and Malta (colored red).

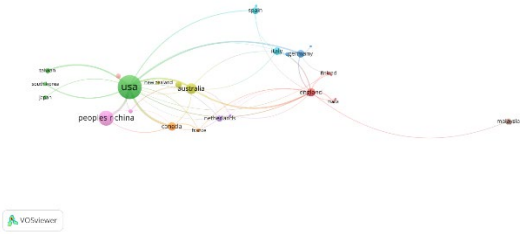



**Figure 6.** Bibliometric coupling of co-authorship between countries

Source: Own mapping in VOSviewer based on the WoS 2023 data.

The citation analysis in Table 8 shows the impact of publications on XBRL. In the field of XBRL over a period of 21 years, 458 papers have been published, which have been cited 4883 times in other works, averaging 221.95 citations per year, 10.48 citations per document, and with an h-index of 34.

**Table 8.** Citation analysis

| Metrics | Data |
|---|---|
| Publication years | 2002-2023 |
| Citation years | 21 |
| Documents | 458 |
| Citations | 4883 |
| Citations per year | 221.95 |

| | |
|---|---|
| Citations per document | 10.48 |
| Hirsch h-index | 34 |

Source: Own elaboration based on WoS 2023 data.

Table 9 lists the most cited articles in the field of XBRL, based on their citation count. The article titled “Does search-facilitating technology improve the transparency of financial reporting?” authored by Hodge, Kennedy & Maines (2004) ranks first in terms of the number of citations (its citation count is 242 or the average per year is 12.74 citations). The second most cited article, with a total of 182 citations (or the average per year is 15.17 citations), was published later, in 2011, under the title "A Review of ERP Research: A Future Agenda for Accounting Information Systems," authored by Grabski, Leech & Schmidt. From the table provided, it can be concluded that among the top ten most cited articles are those with a total of more than 65 citations. Over the twenty-year period observed, the most cited articles were published between 2010 and 2012.

**Table 9.** Most cited articles in the WoS database on XBRL

| Authors | Title | Source | Cites |
|---|---|---|---|
| Hodge, F. D., Kennedy, J. J. & Maines, L. A. (2004) | Does search-facilitating technology improve the transparency of financial reporting? | Accounting Review | 242 |
| Grabski, S. V., Leech, S. A. & Schmidt, P. J. (2011) | A Review of ERP Research: A Future Agenda for Accounting Information Systems | Journal of Information Systems | 182 |
| Debreceny, R., Farewell, S., Piechocki, M., Felden, C. & Graening, A. (2010) | Does it add up? Early evidence on the data quality of XBRL filings to the SEC | Journal of Accounting and Public Policy | 108 |
| Yoon, H., Zo, H. & Ciganek, A. P. (2011) | Does XBRL adoption reduce information asymmetry? | Journal of Business Research | 98 |
| Kim, J. W., Lim, J. H. & No, W. G. (2012) | The Effect of First Wave Mandatory XBRL Reporting across the Financial Information Environment | Journal of Information Systems | 92 |
| Pinsker, R. & Li, S. (2008) | Costs and benefits of XBRL adoption: Early evidence | Communications of the ACM | 80 |
| Blankespoor, E., Miller, B. P. & White, H. D. (2014) | Initial evidence on the market impact of the XBRL mandate | Review of Accounting Studies | 75 |
| Bozanic, Z., Hoopes, J. L., Thornock, J. R. & Williams, B. M. (2017) | IRS Attention | Journal of Accounting Research | 74 |
| Hoitash, R. & Hoitash, U. (2018) | Measuring Accounting Reporting Complexity with XBRL | Accounting Review | 69 |
| Plumlee, R. D. & Plumlee, M. A. (2008) | Assurance on XBRL for Financial Reporting | Accounting Horizons | 67 |

Source: Own elaboration based on WoS 2023 data.

# 5 Discussion and Conclusion

The popularity and development of bibliometric analysis primarily lie in the advancement, availability, and accessibility of bibliometric software and citation databases, which ensure ease and speed in managing structured data. Additionally, the multidisciplinary nature of bibliometric methodology is valuable for handling large volumes of scientific data and generating significant research impact. Citation databases and their online platforms, such as WoS and Scopus, index large volumes of bibliographic data on scientific literature and their references, and measure the citation count of each publication (Buchanan, 2006). Collections of bibliographic data and their corresponding

bibliometric indicators are becoming easily and quickly accessible to scientists for further use.

From this study, 458 publications about XBRL indexed in WoS core collection database were analyzed. Researchers Elliott, R. K., Nutz, A., and Strauss, M. in 2002 have laid the scientific foundations for the research area, and among the most influential authors are Troshani, I., Farewell, S. M. and Debreceny, R. S.

The collaboration structure is revealed through co-authorship analysis, which presents the intellectual collaboration among scientists along with attributes such as organizations and countries. This analysis can provide insights into the connectivity of authors within a specific geographic area, the internationalization of a certain group of authors, and identify authors with extensive collaboration networks. The collaboration network of authors indicates that groups of authors tend to collaborate within smaller groups and have very limited collaboration outside their co-authorship circle. The collaboration network of countries reveals that the most significant scientific collaboration occurs between researchers from the United States and China. Additionally, these two countries collaborate with Taiwan, Japan, South Korea, Canada, and Singapore.

The most frequently used author keywords reveal the thematic areas studied within the field of XBRL, namely financial reporting, taxonomy, information asymmetry, XML, corporate governance, transparency, disclosure and information systems. Journals that have published papers on XBRL are categorized into research areas of various scientific disciplines, such as business economics, computer science, engineering, information science, etc.

Over 95% of documents were published in English and originated from 58 identified countries. With regard to analysis of the citation report from WoS core collection database on XBRL, over a period of 21 years, 458 papers have been published, which have been cited 4883 times in other works, averaging 221.95 citations per year, 10.48 citations per document, and with an h-index of 34.

The documents on XBRL are mostly published in the publications that include business economics, computer science and engineering research area. The analysis of keywords frequency has shown that financial reporting and taxonomy are the most common keywords in the gathered documents.

Despite the comprehensive and objective analysis that bibliometrics enables, there are still some limitations to the conducted research. The first limitation pertains to the fact that the analysis includes papers based on two keywords that appear in the title or abstract of the work (XBRL or extensible business reporting language). In future research, it is recommended to expand the search to include keywords or the entire text. However, in that case, the selection of papers should be based on several keywords that appear in each work. The second limitation relates to the primary source of publication data, which in this case is only the Web of Science database. Therefore, future research could include other relevant databases, such as Scopus. Despite the mentioned limitations, the conducted bibliometric analysis can be beneficial to researchers in the fields of accounting and auditing who are exploring XBRL, aiding in understanding new trends and approaches in researching its application and enhancement for the future.

Bibliometric analysis is a widely applicable tool, so for precise research, it is essential to carefully select the objectives and bibliometric techniques that will provide the best answers.

## Acknowledgments

This work has been fully supported by the Croatian Science Foundation under the project IP-2022-10-6703.

## References

Abdullah, A., Khadaroo, I., & Shaikh, J. (2009), "Institutionalisation of XBRL in the USA and UK", *International Journal of Managerial and Financial Accounting*, Vol. 1, No. 3, pp. 292-304.

Aidi, A. & Rosli, M. (2019), "Bibliometric Analysis of Global Scientific Literature on Web Accessibility", *International Journal of Recent Technology and Engineering,* Vol. 7, No. 6, pp. 250-258.

Backus, C., & Leiser, S. (2021), "An Open Data Standard for Local Government Financial Reporting: How it Could Work in Michigan", *Center for Local, State, and Urban Policy, University of Michigan.* https://closup.umich.edu/sites/closup/files/2021-07/closup-wp-56-Backus-Leiser-An-Open-Data-Standard-for-Local-Government-Financial-Reporting-How-it-Could-Work-in-Michigan.pdf

Bai, Z., Sakaue, M., & Takeda, F. (2014), "The impact of XBRL adoption on the information environment: Evidence from Japan", *The Japanese Accounting Review,* Vol. 4, pp. 49-74.

Bartolacci, F., Caputo, A., Fradeani, A., & Soverchia, M. (2020), "Twenty years of XBRL: what we know and where we are going", *Meditari Accountancy Research,* Vol. 29, No. 5, pp. 1113-1145.

Borgi, H. (2022), "XBRL technology adoption and consequences: A synthesis of theories and

suggestions of future research", *Journal of Accounting and Management Information Systems*, Vol. 21, No. 2, pp. 220-235.

Buchanan, R. A. (2006), "Accuracy of cited references: The role of citation databases", *College & Research Libraries*, Vol. 67, No. 4, pp. 292-303.

Chen, Y. C. (2012), "A comparative study of e-government XBRL implementations: The potential of improving information transparency and efficiency", *Government information quarterly,* Vol. 29, No. 4, pp. 553-563.

de Bellis, N. (2009), "Bibliometrics and citation analysis: from the science citation index to cybermetrics", *Scarecrow press*.

de Bruyn, C., Said, F. B., Meyer, N., & Soliman, M., (2023), "Research in tourism sustainability: A comprehensive bibliometric analysis from 1990 to 2022", *Heliyon*, Vol. 9, No. 8, pp. 1-22.

De Iuliis, M., Cardoni, A., & Cimellaro, G. P. (2024), "Resilience and safety of civil engineering systems and communities: A bibliometric analysis for mapping the state-of-the-art", *Safety Science*, Vol. 174, 106470.

El Ansary, M., & Oubrich, M. (2016), "State of the art and trends of the research on XBRL bibliometric analysis from 2000–2014" in 4th IEEE International Colloquium on Information Science and Technology (CiSt), 2016, IEEE, pp. 243-250.

Elliott, R. K. (2002), "Twenty-first century assurance*", Auditing: A Journal of Practice & Theory,* Vol. 21, No. 1, pp.139-146.

Erkus, H., & Chiu, V. (2014), "On the research contribution of XBRL literature-A bibliometrics analysis", *Eurasian journal of business and economics,* Vol. 7, No. 13, pp. 173-188.

Gatea, A. K., Alnawas, A. A. & Gali, Z. H. (2021), "The effect of XBRL financial reporting on enhancing the transparency of information in the financial statements", *Turkish Journal of Computer and Mathematics Education (TURCOMAT),* Vol. 12, No. 11, pp. 4945-4953.

Goyal, K., & Kumar, S. (2021), "Financial literacy: A systematic review and bibliometric analysis", *International Journal of Consumer Studies*, Vol. 45, No. 1, pp. 80-105.

Ib, C. E., Jide, I., & Zik-Rullahi, A. A. (2016), "The impact of XBRL on financial reporting: A conceptual analysis", *International Journal of Empirical Finance*, Vol. 4, No. 2, pp. 78-85.

Kunisch, S., Menz, M., Bartunek, J. M., Cardinal, L. B., & Denyer, D. (2018), "Feature topic at organizational research methods: How to conduct rigorous and impactful literature reviews?", *Organizational Research Methods*, Vol. 21, No. 3, pp. 519-523.

Lacković Vincek, Z., Gregurec, I., & Kovšca, V. (2023), "Exploring the trend of the research on Logistics 4.0", *LogForum*, Vol. 19, No. 3, pp. 395-409.

Lakovic, T., Rondovic, B., Backovic-Vulic, T., & Ivanovic, I. (2019), "The determinants of xbrl adoption: An empirical study in an emerging economy" in *Information Systems: 15th European, Mediterranean, and Middle Eastern Conference*, *EMCIS 2018, Limassol, Cyprus, October 4-5, 2018, Proceedings 15*, pp. 532-546. Springer International Publishing.

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020), "Conducting systematic literature reviews and bibliometric analyses", *Australian Journal of Management*, Vol. 45. No. 2, pp. 175-194.

Liu, C., Luo, X. R., Sia, C. L., O'farrell, G., & Teo, H. H. (2014), "The impact of XBRL adoption in PR China", *Decision Support Systems*, Vol. 59, pp. 242-249.

Liu, C., Wang, T., & Yao, L. J. (2014), "XBRL's impact on analyst forecast behavior: An empirical study", *Journal of accounting and public policy,* Vol. 33, No. 1, pp. 69-82.

Locke, J., & Lowe, A. (2007), "XBRL: an (open) source of enlightenment or disillusion?", *European Accounting Review,* Vol. 16, No. 3, pp. 585-623.

Miścikowska, D. (2021), "Preparers' readiness for ESEF reporting: Early evidence from the Warsaw Stock Exchange", *Ekonomia Międzynarodowa*, Vol. 35, pp. 194-234.

Nutz, A., & Strauss, M. (2002), "Concept and practical implementation of the eXtensible Business Reporting Language (XBRL) Konzept und praktischer Einsatz", *Wirtschaftsinformatik,* Vol. 44, pp. 447-457.

Roohani, S., Xianming, Z., Capozzoli, E., & Lamberton, B. (2010), "Analysis of XBRL literature: A decade of progress and puzzle", *International Journal of Digital Accounting Research,* Vol. 10, pp. 131-147.

Ruhanen, L., Weiler, B., Moyle, B. D. & McLennan, C. L.J. (2015), "Trends and patterns in sustainable tourism research: A 25-year bibliometric analysis", *Journal of Sustainable Tourism*, Vol. 23, No. 4, pp. 517-535.

Shan, Y. G., & Troshani, I. (2014), "Does XBRL benefit financial statement auditing?", *Journal of*

*Computer Information Systems*, Vol 54, No. 4, pp. 11-21.

Steenkamp, L. P., & Nel, G. F. (2012), "The adoption of XBRL in South Africa: an empirical study", *The Electronic Library,* Vol. 3, No. 30, pp.409-425.

Sweileh, W. M., Al-Jabi, S. W., AbuTaha, A. S., Sa'ed, H. Z., Anayah, F. M. & Sawalha, A. F. (2017), "Bibliometric analysis of worldwide scientific literature in mobile-health: 2006-2016", *BMC Medical Informatics and Decision Making*, Vol. 72, No. 17, pp. 1-12.

Uyob, R., Saad, R. A. J., & Aidi, A. (2019), "A Review of the Study on the Impacts of the eXtensible Business Reporting Language (XBRL)", *International Journal of Scientific & Technology Research,* Vol. 8, No. 9, pp. 2320-2329.

# Phenomenological Inquiry in Entrepreneurship and Small Business with Insights into the Digital Economy Research Context

**Tamara Šmaguc, Marina Klačmer Čalopa**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{tamara.smaguc, marina.klacmer}@foi.unizg.hr

**Mirjana Grčić Fabić**
University of Rijeka
Faculty of Economics and Business
Ivana Filipovića 4, 51000 Rijeka, Croatia
mirjana.grcic.fabic@efri.uniri.hr

***Abstract***. *The paper aims to review phenomenology applications in entrepreneurship and small business research, with a special emphasis on the study of digitalization and digital transformation. In order to achieve research objectives, a narrative literature review was conducted. The collection of literature was carried out by searching Scopus. The results indicate a strong community of phenomenological scholars of entrepreneurship in the United Kingdom. Interpretative Phenomenological Analysis is a popular method, with in-depth and semi-structured interviews being the most common data collection techniques. Regarding the context of digitalization and digital transformation in entrepreneurship and small business, only six phenomenological journal papers were identified. The findings also show that the oldest study was published three years ago, suggesting the novelty and immaturity of the field. To the best of our knowledge, this is the first review paper dealing with the topic. Therefore, it is a good starting point for entrepreneurship and small business scholars interested in this type of qualitative methodology.*

**Keywords.** qualitative methodology, phenomenology, entrepreneurship, digitalization, digital transformation, narrative literature review

## 1 Introduction

Considering the role of digitalization in raising competitiveness and strengthening economic resilience, the interest in the subject has recently grown significantly (Vuković, Korent & Šmaguc, 2024). In the international scientific community, the topic has gained stronger scientific popularity since 2017, and three years later it was experiencing a publication proliferation (Kudelić, Šmaguc & Robinson, 2023; Zhai et al., 2023), primarily in the context of strengthening the firms' resistance to the challenges of the COVID-19 crisis (Kudelić, Šmaguc & Robinson, 2023; Vuković, Korent & Šmaguc, 2024). The importance of issues is also recognized at the level of European Union (EU) policymakers. They declared the period 2021-2030 Europe's Digital Decade, with the digital transformation of companies as one of the key EU priorities (European Commission, 2021).

Despite the exponential growth in popularity, digital transformation and digital entrepreneurship are still relatively new, empirically unsaturated research areas. As with any new research domain, the existing literature is dominated by works of a conceptual nature that offer definitions, systematizations of the area and conceptual frameworks, and implications for somewhat less represented empirical research (see for example Vial (2021), Verhoef et al. (2021), Nambisan (2017), Sussan and Acs (2017), Zaheer, Breyer and Dumay (2019), Reis et al. (2018), Nambisan, Wright and Feldman (2019), Osmundsen, Iden and Bygstad (2018)). In a notable conceptual paper, Verhoef et al. (2021) theorize about possible drivers of digitalization and strategic imperatives of digital transformation, warning about the insufficient focus on empirical studies of digital transformation as a process. Nambisan, Wright and Feldman (2019) note the need for multiple and multilevel empirical analyses that include examining a range of research questions and recommend the integration of ideas and concepts from multiple fields/disciplines. Considering the complex nature of the phenomenon of entrepreneurship (Bruyat & Julien, 2001), such research should be based on ontological pluralism, supplementing the hypothetical-deductive orthodoxy (Drakopoulou-Dodd et al., 2014; Guba & Lincoln, 1994) with an interpretive research framework (Burrell & Morgan, 1979; Cope, 2005). In this context, qualitative research methodology comes to the fore, ensuring the reconstruction of reality shaped from the perspective of real entrepreneurs "on the ground".

Despite the advantages of qualitative methods in the domain, the possibilities of researching digitalization phenomena in entrepreneurship and

small business using qualitative methodology have been poorly explored. Dealing with this research gap, this paper theoretically considers and reviews the application of phenomenology in entrepreneurship and small business, with a special emphasis on the study of digitalization and digital transformation. The latter two terms are distinguished and understood as defined by Verhoef et al. (2021). After a general insight into phenomenology as a research method and the potential of its application in entrepreneurship, the goals and methods of this review paper are presented. The second part of the paper provides a narrative review of phenomenological studies in entrepreneurship and small business in general, and specifically, in the topic of digitalization and digital transformation. The paper ends with concluding considerations, research limitations, and recommendations for future reviews in the field.

# 2 Background of the Study

Section 2.1 elaborates on the characteristics of phenomenology as a philosophical movement and research method. It is followed by a short overview of the advantages of applying phenomenology in entrepreneurship research in Section 2.2.

## 2.1 Phenomenology as a Philosophy and Research Method

The postulates of phenomenology as a research method derive from a broader understanding of it as a philosophical movement. Starting from the latter, Moran (2002, p. 4) views phenomenology as "a radical, anti-traditional style of philosophising, which emphasises the attempt to get to the truth of matters, to describe *phenomena*, in the broadest sense as whatever appears in the manner in which it appears, that is as it manifests itself to consciousness, to the experiencer". The beginnings of the movement are linked to the work of the German philosopher Edmond Husserl in the first half of the 20th century, with roots in the doctrines of Plato, Socrates, and Aristotle (Qutoshi, 2018). The main idea of the philosophy is the establishment of live contact with reality - observation and examination of people's life-worlds or lived experiences, with creating explanations devoid of any prejudice drawn from tradition, experience in advance, common sense, or even existing scientific knowledge (Berglund, 2015; Moran, 2002). The emphasis is on real experiences - getting to know the phenomenon as it really is, in nature, in its essence, capturing it directly from the world (Moran, 2002). Traditional ontological dichotomies (eg agent-structure, consciousness-matter, subject-object) are rejected by phenomenologists claiming that they do not respect the essential postures of the human being. According to them, people naturally perceive situations and objects as meaningful, considering the totality of their experiences. In other words, the meanings attached to phenomena rest on the relationship between parts and the whole, as a result of a person's "historically spun web, or *referential whole*, of interrelated things, background understandings, and practices" (Dreyfus, 1991, pp. 97-99, as cited in Berglund, 2015, pp. 478-479).

Following philosophical foundations, as a qualitative method, phenomenology relies on the interpretative examination of special experiences, events, and states of participants in a specific interaction process (Halmi, 2005). As such, it is based on an empathetic approach to understanding and interpreting individual practices placed in (spatial, industrial, institutional, and other) contexts (Halmi, 2005; Kordeš, 2008). The focus is on the study of critical incidents, from the level of subjectively lived experiences, which enriches the understanding of the investigated phenomena (Cope, 2005).

The paradigmatic framework of phenomenology is interpretivism/constructivism (Burrell & Morgan, 1979), and the researcher's emphasis is on detailed description and interpretation of the investigated phenomena at a deeper level of understanding. From the position of "conceptual silence", the researcher puts "in brackets" theories, hunches, and assumptions, resisting premature judgments based on unreflected ideas (Kordeš, 2008). Such isolation from the fundamental attitudes of the sciences, excluding all transcendent objects and their essence, leads the researcher to a pure transcendental phenomenological experience (Halmi, 2005).

Considering the nature of the method, commonly used data collection strategies are phenomenological interviews, observations, discussions, and focus group meetings. Of course, textual analysis methods are also used, and in certain cases, action research (Qutoshi, 2018). Research participants are selected purposefully, with their eloquence and informativeness being an important selection criterion (Kordeš, 2008). Meaning creation runs parallel to data collection, where a minimum of structure is striven for, with messy analysis due to different ways of connecting different parts of the empirical material. The reporting of findings is usually vigorous in nature, using direct quotes from research participants, with an emphasis on detailed descriptions of phenomena rather than explanations (Qutoshi, 2018).

## 2.2 Phenomenology in Entrepreneurship and Small Business Research

The characteristics of the entrepreneurial process, such as radical uncertainty, creativity and judgment in the situation of unclear goals, make it quite a suitable research arena for phenomenological analyses. In the foreground of entrepreneurial activities are social relations, with the emotional involvement of the individual in the process of birth and development of business. The researcher-phenomenologist captures those social relations and the entrepreneur's lived

experiences with an empathic approach, trying to elucidate and appreciate them, rather than downplay and proclaim them as flaws or anomalies (Qutoshi, 2018). Only such research genres offer results that are sufficiently fruitful from the perspective of the need to create richer and contextually sensitive entrepreneurial policies (Korsgaard et al., 2020).

Phenomenological methods are a kind of niche between research focused on the cognition of entrepreneurs obtained through questionnaires and scales, and narrative and discursive approaches based on in-depth studies of local stories. They focus on lived experiences, the meanings that individuals attach to phenomena and the associated individual strategies. With extensive descriptions and illustrations, phenomenological findings enrich both quantitative cognitive studies and discursive analyses of entrepreneurship, often resulting in the strengthening of existing entrepreneurship theories or the development of new theoretical constructs (Berglund, 2007). As Cope (2005) concludes, although very challenging and demanding, phenomenological research is quite rewarding in the context of insight into the complexity and uniqueness of the perception and practice of entrepreneurial individuals. Thus, entrepreneurship and small business field need a larger number of studies of this type, not only in the context of supplementing quantitative research but also in terms of undertaking complete qualitative research projects, contributing to legitimizing the method.

# 3 Review Objectives and Method

This review is conducted with the following two research objectives:

O1. To give an insight into influential phenomenological studies in entrepreneurship and small business domain with regard to authors, topics, contexts, and methodologies

O2. To give an insight into phenomenological studies on the topic of digitalization and digital transformation in entrepreneurship and small business with regard to authors, contexts, and methodologies.

In order to achieve the objectives of the research, a narrative literature review was conducted (Ferrari, 2015; Green, Johnson & Adams, 2006). Unlike a systematic literature review that is standardized, a narrative review is based on a more informal data extraction process, with data synthesis based on a narrative juxtaposition of material (Xiao & Watson, 2019). The intention is not to summarize the entire structure of an intellectual field, but to find and describe key topics in the literature. Due to its characteristics, a narrative review cannot completely be objectified and isolated from biases arising from the author's experiences and beliefs (Xiao & Watson, 2019). However, to reduce these problems, applying at least some steps of a systematic literature review can be useful (such as defining a literature search database and specifying publication selection criteria) (Ferrari, 2015).

The collection of literature for the review was carried out by searching Scopus as one of the most relevant scientific databases. The search was conducted on 13 May 2024. In accordance with the syntax of the database, two search queries were defined (one for each of the two research objectives). The search was performed within the title, abstract, and keywords. Due to practicality and an effort to include only the most relevant peer-reviewed literature, the following inclusion criteria were applied: 1) Language: English; 2) Document type: Article; 3) Source Type: Journal. It was decided that the O1 would be achieved by reviewing the five most cited empirical studies, and the O2 by reviewing all empirical studies relevant to the field. As a result of the first search query and after applying the specified search filters, 406 journal articles were found. In order to select the most influential articles, they were filtered according to the number of citations in Scopus. After that, screening was performed to identify the first five papers that are empirical research, and at the same time are thematically related to the field. Regarding the O2, a total of ten journal articles were found after applying the filters. Based on the screening of all ten articles, it was determined that four of them are not related to the field. The remaining six articles were included in the narrative review. The Scopus search queries for each research objective and the applied inclusion criteria are shown in Table 1.

A total of eleven selected articles (five for O1 and six for O2) were read and data extraction was performed. The results of the review are presented and discussed in Section 4, accompanied by data on the 11 analyzed publications summarized in Appendix 1.

# 4 Review results

Following the research objectives, this section is divided into two subsections. First, Section 4.1 presents the most influential phenomenological studies in entrepreneurship and small business in general. Afterward, Section 4.2 gives insight into phenomenological studies related to digitalization and digital transformation in entrepreneurship and small business domain.

## 4.1 Influential Phenomenological Studies in Entrepreneurship and Small Business

According to the number of citations in Scopus, the five most influential phenomenological studies of entrepreneurship and small business are Cope (2011), Cope and Watts (2000), Shaw and Carter (2007), Dixon and Clifford (2007), and Doern (2016),

respectively. These studies are significant for phenomenology because they increase the popularity of the method in entrepreneurship and small business. In that sense, they contributed to the broader framework of phenomenology as a research practice.

**Table 1.** Scopus search queries and inclusion criteria with the associated number of publications after applying each filter (search was conducted on May 13, 2024)

| **Search query 1 (for O1)** | **No. of pub.** |
|---|---|
| phenomenolog* AND entrepren* OR "small AND business*" OR "small AND firm*" OR "small AND compan*" OR "small AND enterpris*" OR "sme" OR "smes" OR "msme" OR "msmes" OR "small AND medium*" OR "self-emplo*" OR "start-up" OR "startup" OR "new AND venture*" | Total: 551<br>Language (English): 524<br>Document Type (Article): 409<br>Source Type (Journal): 406<br>Included in the analysis: 5 most cited |
| **Search query 2 (for O2)** | **No. of pub.** |
| phenomenolog* AND entrepren* OR "small AND business*" OR "small AND firm*" OR "small AND compan*" OR "small AND enterpris*" OR "sme" OR "smes" OR "msme" OR "msmes" OR "small AND medium*" OR "self-emplo*" OR "start-up" OR "startup" OR "new AND venture*" AND digital* | Total: 16<br>Language (English): 13<br>Document Type (Article): 10<br>Source Type (Journal): 10<br>Remaining for analysis after screening: 6 |

The most significant phenomenological scientist of entrepreneurship is Jason Cope from the University of Strathclyde. In addition to two empirical studies that lead by the number of citations (Cope, 2011; Cope & Watts, 2000), he is also the author of a very notable conceptual paper in the field (Cope, 2005). The top five most cited publications are signed by a total of eight authors from five institutions (University of Strathclyde, University of Stirling, University of Gloucestershire, Kingston upon Thames, University of London). As can be seen, all institutions are from the United Kingdom, indicating a potentially strong community of phenomenological scholars of entrepreneurship in that country.

When it comes to the thematic focus of the analyzed studies, all of them deal with complex phenomena formed by interweaving the personal and professional lives of entrepreneurs, with immersion in entrepreneurial events filled with a strong dose of emotional engagement of the individual. Specifically, the topics refer to the examination of the process of entrepreneurial learning (Cope, 2011; Cope & Watts, 2000), the investigation of harmonizing the social values of the entrepreneur and the economic goals, in social entrepreneurship (Shaw & Carter, 2007), as well as, ecopreneurship (Dixon & Clifford, 2007), and the research of entrepreneurship and entrepreneurial strategies in the context of crisis management (Doern, 2016). Regarding the methodology, analyzed papers usually use phenomenological case studies and Interpretative Phenomenological Analysis (IPA). Developed by Smith, Jarman and Osborn (1999), IPA is a phenomenological method specific for its "commitment to producing a fine-grained interpretive account that is grounded in, and does justice to, each participant's unique lived experience" (Cope, 2011, p. 609). The aim of IPA-based research is a detailed study of each case individually, before starting to compare cases looking for convergences and divergences. Along with phenomenological philosophy, the intellectual roots of IPA lie in hermeneutics (Eatough & Smith, 2017).

The most common data collection techniques in the analyzed studies are in-depth or semi-structured interviews, sometimes accompanied by document analysis. Participant selection is based on purposeful sampling, with small sample sizes of between 6 and 15 entrepreneurs (the exception is a study by Shaw and Carter (2007) with 80 research participants). Data analysis is based on an inductive approach aimed at developing new theoretical concepts.

In the end, it is relevant to single out the most cited paper *Entrepreneurial learning from failure: An interpretative phenomenological analysis*, published in Journal of Business Venturing by Cope (2011). The study developed a deep and rich understanding of the experience of business failure from the lens of entrepreneurial learning. The emphasis was on examining outcomes and impacts of failure, and the learning process experienced by the entrepreneur as a result of active grieving and recovery. The research methodology was based on IPA. The data was collected by highly unstructured phenomenological interviewing of eight entrepreneurs who experienced business failure. To allow for cross-country comparisons, four participants were from the United Kingdom and the same number came from Silicon Valley, California. The paper provides important contributions to the entrepreneurship literature, especially in the segment of extending Shepherd's (2003) learning theory of grief recovery.

## 4.2 Phenomenological Studies on the Topic of Digitalization and Digital Transformation in Entrepreneurship and Small Business

Regarding the context of digital entrepreneurship and the digitalization of small businesses, only six phenomenological papers were identified. These are the following publications ordered by the number of

citations: Ano and Bent (2022), Guinez-Cabrera and Aqueveque (2022), Allataifeh and Moghavvemi (2021), Imiren et al. (2024), Allataifeh, Moghavvemi and Peerally (2023), and Setiawati, Sutarjo and Primasari (2022). As can be seen, all papers date from the past three years, indicating that the application of phenomenology in this thematic milieu is still in its infancy. This is not surprising considering the immaturity of the digital economy research context in entrepreneurship and small business in general.

In a thematic sense, the papers look at the complex business and social processes associated with digitalization and digital transformation. Ano and Bent (2022) investigate the processes of digital strategy implementation in a family business environment, Allataifeh and Moghavvemi (2021) and Allataifeh, Moghavvemi and Peerally (2023) deal with the phenomenon of digital innovation development, and Imiren et al. (2024) study the creation of legitimacy in the context of digital entrepreneurship. There are also topics on entrepreneurial motives and paths of social media influencers (Guinez-Cabrera & Aqueveque, 2022), and marketing communication strategies of women small business owners in the digital age (Setiawati, Sutarjo & Primasari, 2022).

As is typical of phenomenology, the most frequent data collection methods are in-depth or semi-structured interviews, with a purposeful sample that is very small (a few participants) or small (eg 35 participants in the study by Guinez-Cabrera and Aqueveque (2022)). The methodology of the most cited study by Ano and Bent (2022) is based on IPA, including eight semi-structured interviews within five case studies from France. Data analysis was performed by manual coding and thematic analysis, following the approach developed by Yin (2010). The research deals with the role of human and cultural capital in the processes of shaping and implementing digital strategy in five family firms. The focus was on understanding the company's idiosyncrasies and uniqueness, by diving into the perceptions and personal experiences of individuals. The contributions of the paper are important since it is one of the first studies of the digital transformation of family firms from the lens of human capital. The findings are relevant for entrepreneurs from practice, pointing to the possibilities of identifying and cultivating specific forms of capital to achieve a market advantage based on digital transformation (Ano & Bent, 2022). In this context, they are also important for policymakers aiming at designing measures to support the digitalization of entrepreneurship.

# 5 Concluding Remarks

Entrepreneurship and small business research has been predominantly characterized by employing the deductive research approach, using the existing theory to shape the research process and data analysis. This refers primarily to the application of quantitative research methodology with an emphasis on setting explicit hypotheses to be confirmed or rejected, and testing already existing theories, postulates and assumptions, leaving a very narrow space for detailed interpretation and understanding of the complex nature of the phenomenon. That especially refers to the importance of looking into the context specificities and variation in unfolding the processes, practices and experiences of entrepreneurial individuals, organizations and the environment within which they operate.

The qualitative research methodology, which assumes an inductive research logic and draws conclusions based on observations and data, opens up opportunities to gain insights into the multifaceted nature and more detailed contextual data that reveal vivid perceptions and experiences that entrepreneurs face in their specific environments.

The necessity of this approach is emphasized even more strongly in the area of examining entrepreneurial initiatives within the framework of digital transformation. Digital entrepreneurship as a relatively new research context where affordances of digital technologies represent a novel framework for examining and interpreting current understanding of processes and effects of entrepreneurial activity, sets the need for implementing qualitative research and strengthening existing and building new theoretical insights.

Phenomenology as a qualitative research approach has great potential for exploring and revealing the lived experiences of individuals involved in entrepreneurial endeavors, especially in the context of digital entrepreneurship. The research results with the given inclusion criteria, which point to only a few published papers in the last three years on the topic of phenomenological studies in the field of digitalization in entrepreneurship and small businesses, further, confirm the great prospect of applying this methodology.

This review paper represents a good initial framework for readers who have no experience with phenomenology but are interested in the method application in entrepreneurship and small business domain. The insight into the most cited empirical publications, authors, and methodological settings provided by the paper is a relevant guideline for further familiarization with the field. However, some limitations must be acknowledged. The review is based on a small number of publications, using a narrative approach with a somewhat lower level of research rigor. Future studies should consider a larger number of publications from several scientific databases and include publications in languages other than English. Minimizing data restrictions can also refer to including not only journal articles but also proceeding papers and book chapters. Finally, depending on the research objectives, a more rigorous methodology should be applied, such as a systematic literature review and/or bibliometric analysis, which can provide insight into

the current state and trends in phenomenological research within this area.

Finally, attention should be paid to potential specific topics and emerging research questions in the field. Phenomenology as a method could yield fruitful results in researching the experiences of entrepreneurs in the context of COVID-19 conditions, including new digitalization opportunities arising from the crisis. It is also very suitable for situational examinations of obstacles and drivers of digitalization at the level of entrepreneurs, companies and the environment. Phenomenological research would make it possible to retrieve digitalization drivers and obstacles from the level of lived experience, enriching existing knowledge. Multiple phenomenological studies conducted in different industrial contexts would contribute to the understanding of how digitalization practices are shaped in different environments.

## Acknowledgments

This work has been fully supported by the Croatian Science Foundation under the project IP-2022-10-6703.

## References

Allataifeh, H., & Moghavvemi, S. (2021). The individual dimension of digital innovation: The altered roles of innovation agents and market actors. *Sustainability*, 13(16), 8971.

Allataifeh, H., Moghavvemi, S., & Peerally, J. A. (2023). How does the digital innovation process unfold in practice? A novel third-generation and empirical-based need–solution pairing model. *European Journal of Innovation Management*, 26(3), 730-754.

Ano, B., & Bent, R. (2022). Human determinants influencing the digital transformation strategy of multigenerational family businesses: a multiple-case study of five French growth-oriented family firms. *Journal of Family Business Management*, 12(4), 876-891.

Berglund, H. (2007). Researching entrepreneurship as lived experience. In H. Neergaard & J. P. Ulhøi (Eds.). *Handbook of qualitative research methods in entrepreneurship* (pp. 75-93). Edward Elgar.

Berglund, H. (2015). Between cognition and discourse: Phenomenology and the study of entrepreneurship. *International Journal of Entrepreneurial Behavior & Research*, 21(3), 472-488.

Bruyat, C., & Julien, P. A. (2001). Defining the field of research in entrepreneurship. *Journal of business venturing*, 16(2), 165-180.

Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life*. Burlington: Ashgate Publishing Company.

Cope, J. (2005). Researching Entrepreneurship Through Phenomenological Inquiry: Philosophical and Methodological Issues, *International Small Business Journal*, 23(2), 163-189.

Cope, J. (2011). Entrepreneurial learning from failure: An interpretative phenomenological analysis. *Journal of business venturing*, 26(6), 604-623.

Cope, J., & Watts, G. (2000). Learning by doing–an exploration of experience, critical incidents and reflection in entrepreneurial learning. *International Journal of Entrepreneurial Behavior & Research*, 6(3), 104-124.

Dixon, S. E., & Clifford, A. (2007). Ecopreneurship–a new approach to managing the triple bottom line. *Journal of Organizational Change Management*, 20(3), 326-345.

Doern, R. (2016). Entrepreneurship and crisis management: The experiences of small businesses during the London 2011 riots. *International small business journal*, 34(3), 276-302.

Dodd Drakopoulou, S., Mconald, S., Mcelwee, G., & Smith, R. (2014). A bourdieuan analysis of qualitative authorship in entrepreneurship scholarship. *Journal of Small Business Management*, 52(4), pp. 633–654.

Dreyfus, H. (1991). *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*. Cambridge: MIT Press.

Eatough, V., & Smith, J. A. (2017). Interpretative phenomenological analysis. In C. Willing & W. S. Rogers (Eds.), *The Sage handbook of qualitative research in psychology* (pp. 193-209). London: Sage.

European Comission - Europska komisija (2021). Digitalno desetljeće Europe: digitalni ciljevi za 2030. Retrieved from https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_hr

Ferrari, R. (2015). Writing narrative style literature reviews. *Medical writing*, 24(4), 230-235.

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of chiropractic medicine*, 5(3), 101-117.

Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. In N. K. Denzin & S. Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks: Sage Publications.

Guinez-Cabrera, N., & Aqueveque, C. (2022). Entrepreneurial influencers and influential entrepreneurs: two sides of the same coin. *International Journal of Entrepreneurial Behavior & Research*, 28(1), 231-254.

Halmi, A. (2005). *Strategije kvalitativnih istraživanja u primijenjenim društvenim znanostima*. Jastrebarsko: Naklada Slap.

Imiren, E., Lassalle, P., Mwaura, S., & Nicolopoulou, K. (2024). Understanding legitimacy building in contexts through digital entrepreneurship. *International Journal of Entrepreneurial Behavior & Research*, 30(2/3), 548-571.

Kordeš, U. (2008). Fenomenološko istraživanje. In N. Koller-Trbović & A. Žižak (Eds.), *Kvalitativni pristup u društvenim znanostima* (pp. 13–35). Zagreb: Edukacijsko-rehabilitacijski fakultet Sveučilišta u Zagrebu.

Korsgaard, S., Hunt, R. A., Townsend, D. M., & Ingstrup, M. B. (2020). COVID-19 and the importance of space in entrepreneurship research and policy. *International Small Business Journal*, 38(8), 697-710.

Kudelić, R., Šmaguc, T., & Robinson, S. (2023). Artificial Intelligence in the Service of Entrepreneurial Finance: Knowledge Structure and the Foundational Algorithmic Paradigm. arXiv preprint arXiv:2311.13213.

Moran, D. (2002). *Introduction to phenomenology*. London: Routledge.

Nambisan, S. (2017). Digital entrepreneurship: Toward a digital technology perspective of entrepreneurship. *Entrepreneurship theory and practice*, 41(6), 1029-1055.

Nambisan, S., Wright, M., & Feldman, M. (2019). The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes. *Research Policy*, 48(8), 103773.

Osmundsen, K., Iden, J., & Bygstad, B. (2018). Digital Transformation: Drivers, Success Factors, and Implications. In *MCIS 2018 Proceedings*. The 12th Mediterranean Conference on Information Systems (MCIS), Corfu, Greece. Retrieved from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1004&context=mcis2018

Qutoshi, S. B. (2018). Phenomenology: A philosophy and method of inquiry. *Journal of Education and Educational Development*, 5(1), 215-222.

Reis, J., Amorim, M., Melão, N., & Matos, P. (2018). Digital transformation: a literature review and guidelines for future research. In *World conference on information systems and technologies* (pp. 411-421). Cham: Springer.

Setiawati, S. D., Sutarjo, M. A. S., & Primasari, I. (2022). Self-Existence in Social Media as a Communication Strategy for Women Entrepreneurs. *Res Militaris*, 12(2), 1927-1938.

Shaw, E., & Carter, S. (2007). Social entrepreneurship: Theoretical antecedents and empirical analysis of entrepreneurial processes and outcomes. *Journal of small business and enterprise development*, 14(3), 418-434.

Shepherd, D. A. (2003). Learning from business failure: Propositions of grief recovery for the self-employed. *Academy of Management Review*, 28(2), 318-328.

Smith, J.A., Jarman, M., & Osborn, M. (1999). Doing interpretative phenomenological analysis. In M. Murray & K. Chamberlain, (Eds.), *Qualitative health psychology: Theories and methods* (pp. 218-240). London: Sage.

Sussan, F., & Acs, Z. J. (2017). The digital entrepreneurial ecosystem. *Small Business Economics*, 49(1), 55-73.

Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, 122, 889-901.

Vial, G. (2021). Understanding digital transformation: A review and a research agenda. In A. Hinterhuber, T. Vescovi & F. Checchinato (Eds.), *Managing Digital Transformation* (pp. 13-66). London: Routledge.

Vuković, K., Korent, D., & Šmaguc, T. (2024). *Poduzetničko društvo u europskom kontekstu*. Varaždin: Sveučilište u Zagrebu Fakultet organizacije i informatike.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of planning education and research*, 39(1), 93-112.

Yin, R. K. (2010). *Qualitative Research from Start to Finish*. New York: Guilford Publications.

Zaheer, H., Breyer, Y., & Dumay, J. (2019). Digital entrepreneurship: An interdisciplinary structured literature review and research agenda. *Technological Forecasting and Social Change*, 148, 119735.

Zhai, Y., Yang, K., Chen, L., Lin, H., Yu, M., & Jin, R. (2023). Digital entrepreneurship: global maps and trends of research. *Journal of Business & Industrial Marketing*, 38(3), 637-655.

# Appendix 1

**Table 2.** Publications included in the review (*number of citations in Scopus)

| Publication | Cit.* | Research objective | Methodology |
|---|---|---|---|
| *Influential phenomenological studies in entrepreneurship and small business domain (in general)* | | | |
| Cope, J. (2011). Entrepreneurial learning from failure: An interpretative phenomenological analysis. Journal of business venturing. | 639 | to create a theoretical model of learning from entrepreneurial failure with key stages in the entrepreneurial learning process | Interpretative Phenomenological Analysis (IPA); Phenomenological interviewing; 8 participants (4 from the USA, 4 from the UK) |
| Cope, J., & Watts, G. (2000). Learning by doing–an exploration of experience, critical incidents and reflection in entrepreneurial learning. International Journal of Entrepreneurial Behavior & Research. | 614 | to examine the entrepreneurial learning process and its relations with the personal and career development of an entrepreneur | Phenomenological case study approach; 6 participants (small business owners); emphasis on critical incident analysis |
| Shaw, E., & Carter, S. (2007). Social entrepreneurship: Theoretical antecedents and empirical analysis of entrepreneurial processes and outcomes. Journal of small business and enterprise development. | 499 | to examine entrepreneurial processes in social entrepreneurship (financial risk, network embeddedness, reward and profit, innovation, etc.) | Phenomenological approach; 80 in-depth interviews with social entrepreneurs from the UK; inductive analysis, and grounded theory |
| Dixon, S. E., & Clifford, A. (2007). Ecopreneurship–a new approach to managing the triple bottom line. Journal of Organizational Change Management. | 229 | to investigate the practices of ecopreneurs by which they create an economically sustainable business, aligned with their social and environmental values | Single case study of Green-Works from the UK; semi structured interviews, micro-ethnography, and document analysis; inductive analysis |
| Doern, R. (2016). Entrepreneurship and crisis management: The experiences of small businesses during the London 2011 riots. International small business journal. | 184 | to examine the impact of the crisis on small businesses and strategies to respond (the context of the London 2011 riots) | Phenomenological approach; in-depth interviews with 15 owner-managers of small businesses in London |
| *Phenomenological studies on the topic of digitalization in entrepreneurship and small business domain* | | | |
| Ano, B., & Bent, R. (2022). Human determinants influencing the digital transformation strategy of multigenerational family businesses: a multiple-case study of five French growth-oriented family firms. Journal of Family Business Management. | 19 | to investigate the impact of human and cultural capital on the creation and implementation of the digital transformation strategy of family businesses | Interpretative Phenomenological Analysis (IPA); 8 semi-structured interviews with family business owners (different generational cohorts); 5 case studies from France |
| Guinez-Cabrera, N., & Aqueveque, C. (2022). Entrepreneurial influencers and influential entrepreneurs: two sides of the same coin. International Journal of Entrepreneurial Behavior & Research. | 10 | to investigate entrepreneurial motives and paths of social media users who became social media influencers (SMIs) | Phenomenological approach; 35 semi-structured face-to-face interviews with SMIs |
| Allataifeh, H., & Moghavvemi, S. (2021). The individual dimension of digital innovation: The altered roles of innovation agents and market actors. Sustainability. | 3 | to examine the processes in the development of digital innovation and the roles of a heterogeneous set of actors in these processes | Phenomenological approach; semi-structured interview with 21 SMEs entrepreneurs from the ICT sector in Malaysia |
| Imiren, E. et al. (2024). Understanding legitimacy building in contexts through digital entrepreneurship. International Journal of Entrepreneurial Behavior & Research. | 0 | to explore the challenges of building the legitimacy of entrepreneurs operating in the digital context | Phenomenological approach to analysis; semi-structured interviews with 21 entrepreneurs in Nigeria |
| Allataifeh, H. et al. (2023). How does the digital innovation process unfold in practice? A novel third-generation and empirical-based need–solution pairing model. European Journal of Innovation Management. | 0 | to develop a model explaining the digital innovation process from practice | Exploratory and phenomenological study; 21 Malaysian SMEs from ICT sector |
| Setiawati, S. D. et al. (2022). Self-Existence in Social Media as a Communication Strategy for Women Entrepreneurs. Res Militaris. | 0 | to examine the marketing communication strategies of women small business owners in the digital age | Phenomenological approach; in-depth interviews with 5 women SMEs |

# Exploring the trend of scientific research of eye-tracking for digital advertising insight – A Bibliometric Analysis

**Iva Gregurec, Magdalena Kuštelega**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{iva.gregurec, mkustele2}@foi.unizg.hr

**Goran Vlašić**
University of Zagreb
Faculty of Economics and Business
Trg J.F. Kennedy 6, 10000 Zagreb, Croatia
{gvlasic}@efzg.hr

***Abstract***. *Eye-tracking technology continues to evolve, playing an increasingly significant role in digital advertising. As a result, scientists have been motivated to explore this topic further. In this paper, desk research was conducted to analyze publications related to eye-tracking usage in digital advertising, using data from the WoS database. The analysis includes standard bibliometric indicators, examining research productivity, paper and source types, language of publications, distribution by countries, active affiliations, authorship, keywords, and citation analysis. The data reveal that eye-tracking technology in the digital advertising context has been researched since 1994, with a notable increase in research activity since 2020. A total of 49 publications related to eye-tracking in digital advertising are visible in the WoS database. All these papers are published in English, and most of them are original research articles, authored by 120 researchers from 20 different countries and 90 affiliations. Notably, this paper represents one of the few attempts to explore eye-tracking as a future trend using bibliometric analysis, particularly leveraging the VOSviewer software.*

**Keywords.** eye-tracking, digital marketing, research, bibliometric analysis, VOSviewer

# 1 Introduction

Following the adage that “seeing is believing” and “believing is buying” (Maslowska et al., 2020) marketing can benefit from a better understanding of how consumers use sight in purchase situations. Therefore, eye-tracking technology plays an important role in marketing as it enables precise estimation of eye movement patterns and provides insights into consumers' attention (Maslowska et al., 2020; Schmidt & Maier, 2022). By utilizing eye-tracking technology, eye movements and fixation patterns can be accurately and discreetly observed (Russell et al., 2017). Furthermore, eye-tracking can offer a fresh and unique perspective to researchers and marketers to understand how online consumers make purchasing decisions (Maslowska et al., 2020).

Over time, eye-tracking technology's ability to evaluate attention without drawing focus to the stimuli being analyzed provides new, fresh insights into the “black box”, or inner mechanisms of product placements and advertising messages (Russell et al., 2017).

Eye-tracking is increasingly used particularly in print ads, to capture both individual occurrences and the overall context of the ad (Fox et al., 1998). By tracking visual attention of users this technology can be used to help determine one’s focus on an ad (Fox et al., 1998) as well as brand placement disclosure (Boerman et al., 2015). Smit et al. (2015) highlighted that eye-tracking can be challenging for research participants, but can offer crucial information which can enhance marketing campaigns' strategic planning.

Over time, eye-tracking technologies significantly improved. In 2013 a study by Hartmann et al. was carried out, and it showed the differences in analyzing eye-tracking data obtained through automated tabulation, registration, and tracking of eye-fixation. The authors divided the data into picture and text areas within experimental ads using Tobii T60 eye-tracking technology. The goal was to determine the number of visits to each area. They concluded that viewers are more drawn to text-based advertising messaging than to the image itself (Hartmann et al., 2013). Three years later, a new technology emerged, it was electroencephalography–eye-tracking (EEG/ET). This innovative technique allows precise identification of both the emotional valence (from the EEG) and the source of a reaction (from the ET) (Matukin et al., 2016). With the help of this technique, marketers can gain insights into the emotional responses and attention patterns of consumers according to which they can create more visually appealing brochures, internet ads, packaging, outdoor ads, and booklets (Matukin et al., 2016).

The upcoming chapter outlines the research methodology. Subsequently, the third chapter presents the research findings, encompassing descriptive statistical results, visualization analyses, research

implications and limitations. Finally, the paper ends with a summary of the conducted research.

# 2 Methodology

A review of the literature published until April 24th, 2024, which dealt with the research of eye-tracking usage in digital advertising is the main aim of this paper. Desk research was conducted to fulfill the set research goal. A bibliometric analysis was made on the collected secondary data to summarize the existing literature in the Web of Science core collection database. Usual bibliometric analysis with descriptive statistics helps to identify authors, keywords, citations, journals, organizations, countries, all their links, and other criteria to reveal their connection at the scientific level (Lacković Vincek et al., 2023). Data for bibliometric analysis can be obtained using various databases. Most often, it includes relevant ones such as Scopus and Web of Science (WoS) and the authors decided to use the Web of Science core collection database based on several facts. WoS was selected due to its reputation as the most discerning database, boasting high impact factors and comprehensive citation analysis. It stands as one of the world's foremost relevant databases (Pavlić, 2024) so the reason for choosing WoS database is evident. According to the author's knowledge, so far in the WoS database, bibliometric analysis has been done on topics related to this research, such as those that cover neuromarketing research (Alsharif et al., 2021), usage of eye-tracking technologies for vision screening (Ali et al., 2021) and eye-tracking technologies in marketing in general (Munoz Leiva et al., 2022). However, in the case of those papers, the same search query was not used to search the WoS and also VOSviever as a visualization tool was used in two of them (Alsharif et al., 2021; Ali et al., 2021), but not directly connected with the aim of this paper. Also, it should be pointed out that database Alsharif et al. (2021) used the Scopus database for their research. Given that this specific type of paper is not currently present in the mentioned database, the significance of this paper's contribution becomes apparent.

The research process of bibliometric analysis consisted of the following stages:

1. Research design – During their research in the field of digital marketing, the authors constantly encountered the term "eye-tracking", which is often associated with the field of digital advertising. Considering the significant increase in digital sales after the Covid-19 pandemic and the projected future growth of the same (Statista, 2024), it has become obvious that the analysis of existing papers on eye-tracking technology is essential for companies operating in the digital environment. Additionally, the authors noticed a lack of bibliometric analysis papers covering this topic, which indicated the need for further research, so they decided to fill that gap and make a bibliometric analysis of existing papers visible in the WoS database.
2. Search strategy definition and selection of the appropriate database – After connecting several keywords (synonyms) and phrases that are associated with the idea of the paper itself, it was determined which are the primary keywords for literature search ("eye-tracking" OR "eye movement" OR "eye monitoring" OR "eye motion"). By selecting keywords (search query), several scientific databases were searched, including the two most relevant - Scopus and WoS and it was decided, as stated earlier, that the bibliometric analysis would be done on papers published in the WoS database. Based on the search query 52 235 papers were obtained in all research areas in the entire WoS database until April 24th, 2024. The search query included variations of all terms in all fields. However, a more detailed filtering of the "Business Economics" research area 712 papers was obtained. Additional filtering of that research area to the area that only covers "Communication" which is connected the most with digital advertising which is the main topic of this paper, resulted in 49 papers.
3. Refining the initial search results – In this phase, all 49 papers were scanned to filter the data according to the given criteria, which were that the papers investigating eye-tracking technology are related to digital advertising. By reviewing the titles, abstracts and primary texts of each article, it was determined that all the papers meet the given criteria. Information for documents that met the threshold included year of publication, language, journal, title, author, affiliation, keywords, document type, and number of citations, all of which were exported in plain text format for the purpose of the next phase – data analysis.
4. Data analysis – In this phase, descriptive statistics in WoS bibliometric mapping in the VOSviewer software (version 1.6.10) was used to perform and to visualize the obtained results, i.e. a network of papers co-occurrences, co-citations, and bibliographic matching.

# 3 Research results

## 3.1 Results of descriptive statistics

The main goal of this paper is to give a detailed review of the literature published until April 24th, 2024 within the Web of Science core collection database. The focus of this research is on eye-tracking usage in the context of digital advertising. The paper achieved its goal through conducted desk research.

In the WoS database, according to the research results after the second phase of the analysis, 49 papers covering the research topic were filtered. Of the 49 papers published in the WoS database between 1994 and 2024, all papers were published in English by 120 authors from 20 different countries and 90 affiliations. These publications included 44 original research articles, 6 proceeding articles, 1 book chapter, 1 review article and 1 early access article. The first scientific paper on this topic was published in 1994 and it was a paper entitled "Do Adolescents Attend to Warnings in Cigarette Advertising - An Eye-Tracking Approach" written by the authors Krugman, D.M., Fox, R.J., Fletcher, J.E., Fischer, P.M. and Rojas, T.H. and published in Journal of Advertising Research. From then until April 24th, 2024, 49 papers on this topic were published and 44 of them have been cited, and the total number of citations of all papers is 1 142 (without self-citations 1 089) of average 40,79 per item and H-index 18. The highest number of analyzed papers was published in 2017 and 2019, 7 of them. The highest level of citations was achieved in 2022, when the number of citations reached 178. The most cited paper entitled "Consumer eye movement patterns on yellow pages advertising" by Lohse, G.L. was published in year 1997. Its number of citations is 212 (average per year is 17,57). However, it is important to point out that in 2017 there was a noticeable increase in the number of published papers on this topic. Namely, since 2020, 21 papers have been published. This increase in research results can probably be attributed to the impact of the Covid-19 pandemic, as well as the increasing interest of scientists in researching this area due to the increase in digital sales. Table 1. shows the ten most cited papers in the WoS database that dealt with the research of eye-tracking usage in digital advertising. And, the distribution of the number of all published papers as well as the number of citations and the total number of citations in the period from 1994 to 2024 is shown graphically in Figure 1.

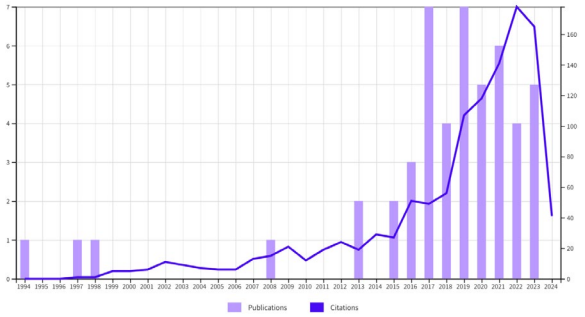



**Figure 1.** Distribution of the number of published papers and citations by year (Source: Authors's work based on Clarivate Web of Science, April 24th, 2024)

Most papers, 13 of them are published in the International Journal of Advertising, 12 in the Journal of Advertising, 12 in the Journal of Advertising Research, and 4 in the Journal of Current Issues and Research in Advertising. Two papers each were published in the journals European Advertising Academy, Marketing Identity and Public Relations Review. In other journals is published one paper (Advances in Advertising Research, Elgar Handbooks in Political Science, International Journal of Business Communication, Marketing Identity Online Rules Pt I and II and Research Handbook on Visual Politics). 40 published papers are indexed in journals with the Social Sciences Citation Index (SSCI), 6 in the Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH) category, 4 of them in the Emerging Sources Citation Index (ESCI) and 1 in Book Citation Index – Social Sciences & Humanities (BKCI-SSH) category. Based on the research area, and bearing in mind the set search query, the results show that all scientific papers are related to business economics and communication (49 papers) and one in government law. It should be noted that this one is included in the other two mentioned research areas as well.

**Table 1.** The ten most cited papers in the WoS database that dealt with the research of eye-tracking usage in digital advertising

| RANK | PAPER | AUTHOR(S) | JOURNAL | YEAR | Total WoS citations |
|---|---|---|---|---|---|
| 1 | Consumer eye movement patterns on yellow pages advertising | Lohse G.L. | Journal of Advertising | 1997 | 212 |
| 2 | Do Adolescents Attend to Warnings in Cigarette Advertising - An Eye-Tracking Approach | Krugman et al. | Journal of Advertising Research | 1994 | 116 |
| 3 | Using Eye Tracking to Understand the Effects of Brand Placement Disclosure Types in Television Programs | Boerman et al. | Journal of Advertising | 2015 | 108 |
| 4 | Adolescents' attention to beer and cigarette print ads and associated product warnings | Fox et al. | Journal of Advertising | 1998 | 94 |
| 5 | Nature imagery in advertising Attention restoration and memory effects | Hartmann et al. | International Journal of Advertising | 2013 | 58 |
| 6 | Understanding which cues people use to identify influencer marketing on Instagram: an eye tracking study and experiment | Boerman & Muller | International Journal of Advertising | 2022 | 41 |
| 7 | Face Presence and Gaze Direction In Print Advertisements How They Influence Consumer Responses-An Eye-Tracking Study | Adil et al. | Journal of Advertising Research | 2018 | 30 |
| 8 | How consumers attend to online reviews: an eye-tracking and network analysis approach | Maslowska et al. | International Journal of Advertising | 2020 | 28 |
| 9 | The Development and Testing of a Child-inspired Advertising Disclosure to Alert Children to Digital and Embedded Advertising | De Jans et al. | Journal of Advertising | 2018 | 28 |
| 10 | Toward a Better Understanding of Advertising Stimuli Processing Exploring the Link between Consumers' Eye Fixation and Their Subconscious Responses | Matukin et al. | Journal of Advertising Research | 2016 | 27 |

Source: Authors's work based on Clarivate Web of Science, April 24th, 2024

A total of 20 countries were involved in the publications of the researched topic of this paper. The most active country with 20 publications is USA. It is followed by the Netherlands (6 publications), Australia (5 publications), France (4 publications), Belgium, PRC and Spain with 3 publications. Austria, Czech Republic, Denmark, Germany, Greece, India, New Zealand, Oman, Poland, Slovakia, South Africa, Taiwan and United Arab Emirates have one publication on the topic of eye-tracking in connection with digital advertising. The most active academic affiliations are the University of Amsterdam with 6 publications, and Ghent University, Texas Tech University, Texas Tech University System, University of Georgia and University System of Georgia with 3 publications. American University, Macquarie University, Northwestern University, Uniformed Services University of the Health Sciences USA, University of Minnesota System, University of Minnesota Twin Cities and the University of South Australia have 2 papers visible in WoS database, and other 77 institutions had one publication until April 24th, 2024.

## 3.2 Visualization of bibliometric analysis

This section describes the bibliometric analysis of keyword co-occurrence, bibliographic coupling of documents to see the links between published papers and their authors, as well as citation metrics.

A total of 12 keywords, in the papers included in the bibliometric analysis on the topic of eye-tracking in connection with digital advertising, that occurred more than five times in the WoS database are distributed in 2 clusters of different colors (red and green) with 50 links and total link strength of 116 (see Figure 2). The first cluster, colored in red, includes 7 keywords (attention, eye-tracking, impact, memory, persuasion knowledge, recall and responses). The second cluster, which is colored green, includes 5 keywords (brand, information, model, movements and pictorial). The most frequent keyword, also the core keyword related to search queries, was "eye-tracking" with 15 occurrences, 10 links and a total link strength of 29. This keyword is followed by keyword "attention" with an occurrence level of 13, 11 links between other keywords and a total link strength of 30. "Brand" as the third most frequently used keyword (12 occurrences) has 9 links and a total link strength of 29. "Information" and "impact" are the fourth keyword with 10. It is followed by the keywords "memory" with 9 occurrences, "movements" with 7 occurrences, "recall", "responses", "persuasion knowledge", "model" with 6 occurrences and "pictorial" with 5 occurrences. All of them have a lower level of occurrence, links with other keywords and link strength than the three main ones (eye-tracking, attention and brand).

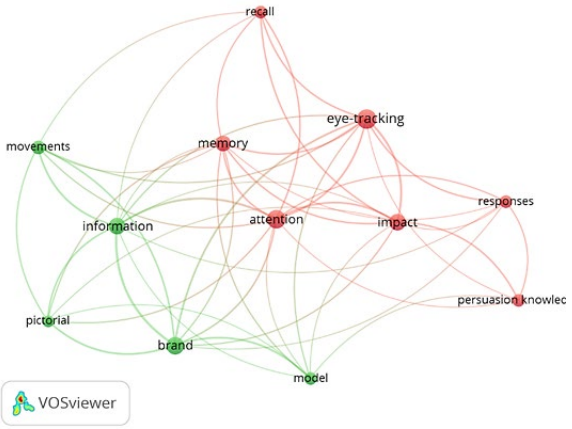



**Figure 2.** Visualization map of keywords co-occurrence (Source: Author's work based on collected data, April 24th, 2024)

From Figure 3, it can be seen that when matching the available papers 7 clusters exist. In the first cluster, colored in red, 11 papers were found. All papers in this cluster are published between 2017 and 2023. In the second cluster, which is colored green, 8 papers were found. They are all papers published in the last five years (from 2019 to 2024). The third cluster, which includes 8 papers, is marked in blue. The fourth cluster, colored yellow, consists of 7 papers. Within this cluster are the oldest papers, published in the years 1994, 1997, and 1998. Papers from this cluster are also in the five most cited papers, the papers from authors Lohse (1997), Krugman, et al. (1994) and Hartmann et al. (2013). 7 papers are in the fifth cluster, colored purple. Boerman et al. (2015) and Boerman and Muller (2022) are the authors of the papers that are in five most cited papers of this research field. Those papers are in the sixed cluster which has 5 papers and is marked in light blue. The last and the smallest one, which is colored orange, consists of 3 papers. This visualization shows that the paper "Consumer eye movement patterns on yellow pages advertising" written by Lohse and published in 1997 represents a research base for all other papers published in the field of research on eye-tracking as a trend in digital advertising together with paper "Do adolescents attend to warnings in cigarette advertising - An eye-tracking approach" from the author Krugman and his associates (1994).

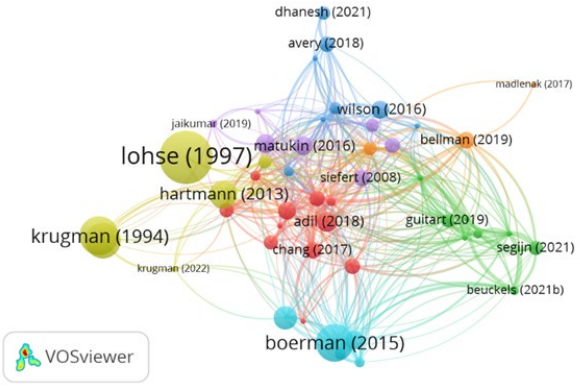



**Figure 3.** Visualization map of bibliographic coupling of documents (fractional counting) (Source: Author's work based on collected data, April 24th, 2024)

### 3.3 Research discussion and limitations

Eye-tracking has an undisputable role in today's digital advertising which can be seen from the theoretical background of this paper as well as from bibliometric analysis that was made for the purpose of achieving the main goal set before research was conducted. After all phases of the research process, which is listed in the methodological part of this paper, have been produced it is evident that according to keywords occurrence the link between eye-tracking technology and digital advertising is strong and very important. According to the papers published since 2020, it can be concluded that this field is gaining in importance each day. Perhaps the reasons for this are the development of digital technology or the increase in sales in the digital environment during the lockdown caused by the Covid-19 pandemic, but perhaps the reasons are something else, time will certainly show. Currently, only a limited number of papers, authors, institutions, and countries are actively researching the mentioned field. However, there is an expectation that this situation will evolve in the future. Until then, the scientific impact of this paper remains unquestionable.

Nevertheless, the conducted research has certain limitations. Firstly, the fundamental limitation lies in the fact that the bibliometric analysis relies solely on the WoS database. While WoS is considered a very comprehensive database, future studies should broaden their scope by incorporating additional scientific databases. This would enable a more comprehensive analysis and facilitate comparisons across different databases. Additionally, the research's limitations extend to used search query, as it inherently restricts the authors. In response to the initial query according to which a substantial number of papers, specifically 52 235 of them, the authors set criteria that they considered the most appropriate for the topic of the paper, eye-tracking technology in digital advertising. This approach potentially resulted in fewer papers for bibliometric analysis which can be considered as a research limitation. However, upon reviewing all 49 selected papers, it was confirmed that the established conditions were appropriate and that each paper corresponded to the researched topic.

## 4 Conclusion

As this article shows, eye-tracking is gaining popularity as a crucial technology for measuring and capturing the attention of consumers, yet it is still not reaching its full potential. However, the development of eye-tracking technology has created a wealth of new opportunities for study and growth in the area of digital advertising, which was the primary subject of this paper. And with the daily development of eye-tracking technology, in the future digital advertising, regardless of web or mobile version, could gain more and more importance.

Apart from providing a theoretical overview of the term "eye-tracking" technology in the digital advertising context, this paper presents the results collected from the WoS database by conducting desk research. The use of a bibliometric analysis, which includes a data set of 49 publications, reveals the strong and significant relationship between eye-tracking technology and digital advertising based on keyword occurrence. Considering the theoretical overview of the term "eye-tracking" from its inception in 1994 to 2024 (the most ancient and most recent article from the observed data set), the previously mentioned outcome was predicted. So, a greater increase in scientific papers on this topic can be expected.

However, it is crucial to remember that the success of bibliometric analysis done for the purpose of writing this paper is primarily the result of high-quality data preparation, which consists of four stages: research design, search strategy definition and selection of the appropriate database, refining the initial search results and data analysis.

## References

Adil, S., Lacoste-Badie, S., & Droulers, O. (2018). Face Presence and Gaze Direction In Print Advertisements How They Influence Consumer Responses-An Eye-Tracking Study. In *JOURNAL OF ADVERTISING RESEARCH* (Vol. 58, Issue 4, pp. 443–455). ADVERTISING RESEARCH FOUNDATION. https://doi.org/10.2501/JAR-2018-004

Ali, Q., Heldal, I., & Helgesen, C. G. (2021). A Bibliometric Analysis and Visualization of the Use of Eye-Tracking Technologies for Vision Screening. *2021 INTERNATIONAL CONFERENCE ON E-HEALTH AND BIOENGINEERING (EHB 2021), 9TH EDITION.* 9th IEEE International Conference on E-Health and Bioengineering (EHB), New York. https://doi.org/10.1109/EHB52898.2021.9657547

Alsharif, A. H., Md Salleh, N. Z., Baharun, R., & Hashem, E. A. R. (2021). Neuromarketing research in the last five years: A bibliometric analysis. *COGENT BUSINESS & MANAGEMENT*, *8*(1), 1978620. https://doi.org/10.1080/23311975.2021.1978620

Boerman, S. C., & Muller, C. M. (2022). Understanding which cues people use to identify influencer marketing on Instagram: An eye tracking study and experiment. In *INTERNATIONAL JOURNAL OF ADVERTISING* (Vol. 41, Issue 1, pp. 6–29). ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD. https://doi.org/10.1080/02650487.2021.1986256

Boerman, S. C., van Reijmersdal, E. A., & Neijens, P. C. (2015). Using Eye Tracking to Understand the Effects of Brand Placement Disclosure Types in Television Programs. In *JOURNAL OF ADVERTISING* (Vol. 44, Issue 3, pp. 196–207). ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD. https://doi.org/10.1080/00913367.2014.967423

De Jans, S., Vanwesenbeeck, I., Cauberghe, V., Hudders, L., Rozendaal, E., & van Reijmersdal, E. A. (2018). The Development and Testing of a Child-inspired Advertising Disclosure to Alert Children to Digital and Embedded Advertising. In *JOURNAL OF ADVERTISING* (Vol. 47, Issue 3, pp. 255–269). ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD. https://doi.org/10.1080/00913367.2018.1463580

Fox, R., Krugman, D., Fletcher, J., & Fischer, P. (1998). Adolescents' attention to beer and cigarette print ads and associated product warnings. In *JOURNAL OF ADVERTISING* (Vol. 27, Issue 3, pp. 57–68). CTC PRESS/JOA. https://doi.org/10.1080/00913367.1998.10673563

Hartmann, P., Apaolaza, V., & Alija, P. (2013). Nature imagery in advertising Attention restoration and memory effects. In *INTERNATIONAL JOURNAL OF ADVERTISING* (Vol. 32, Issue 2, pp. 183–210). ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD. https://doi.org/10.2501/IJA-32-2-183-210

Krugman, D., Fox, R., Fletcher, J., Fischer, P., & Rojas, T. (1994). Do adolescents attend to warnings in cigarette advertising—An eye-tracking approach. In *JOURNAL OF ADVERTISING RESEARCH* (Vol. 34, Issue 6, pp. 39–52). ADVERTISING RESEARCH FOUNDATION.

Lacković Vincek, Z., Gregurec, I., & Kovšca, V. (2023). Exploring the trend of the research on Logistics 4.0. *Logforum*, *19*(3), 359–409. https://doi.org/10.17270/J.LOG.2023.859

Lohse, G. (1997). Consumer eye movement patterns on yellow pages advertising. In *JOURNAL OF ADVERTISING* (Vol. 26, Issue 1, pp. 61–73). ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD. https://doi.org/10.1080/00913367.1997.10673518

Maslowska, E., Segijn, C. M., Vakeel, K. A., & Viswanathan, V. (2020). How consumers attend to online reviews: An eye-tracking and network analysis approach. In *INTERNATIONAL JOURNAL OF ADVERTISING* (Vol. 39, Issue 2, pp. 282–306). ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD. https://doi.org/10.1080/02650487.2019.1617651

Matukin, M., Ohme, R., & Boshoff, C. (2016). Toward a Better Understanding of Advertising Stimuli Processing Exploring the Link between Consumers' Eye Fixation and Their Subconscious Responses. In *JOURNAL OF ADVERTISING RESEARCH* (Vol. 56, Issue 2, pp. 205–216). ADVERTISING RESEARCH FOUNDATION. https://doi.org/10.2501/JAR-2016-017

Munoz Leiva, F., Rodriguez Lopez, M. E., & Garcia Marti, B. (2022). Discovering prominent themes of the application of eye tracking technology in marketing research. *CUADERNOS DE GESTION*, *22*(1), 97–113. https://doi.org/10.5295/cdg.211516fm

Pavlić, I. (2024). ELECTRODERMAL ACTIVITY IN TOURISM RESEARCH – BIBLIOMETRIC ANALYSIS. *CroDiM : International Journal of Marketing Science*, *7*(1), 99–116.

Russell, C. A., Russell, D., Morales, A., & Lehu, J.-M. (2017). Hedonic Contamination of Entertainment How Exposure to Advertising in Movies and Television Taints Subsequent Entertainment Experiences. In *JOURNAL OF ADVERTISING RESEARCH* (1; Vol. 57, Issue 1, pp. 38–52). ADVERTISING RESEARCH FOUNDATION. https://doi.org/10.2501/JAR-2017-012

Schmidt, L. L., & Maier, E. (2022). Interactive Ad Avoidance on Mobile Phones. In *JOURNAL OF ADVERTISING* (4; Vol. 51, Issue 4, pp. 440–449). ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD. https://doi.org/10.1080/00913367.2022.2077266

Smit, E. G., Boerman, S. C., & van Meurs, L. (2015). The Power of Direct Context As Revealed by Eye Tracking A Model Tracks Relative Attention To Competing Editorial and Promotional Content. In *JOURNAL OF ADVERTISING RESEARCH* (Vol. 55, Issue 2, pp. 216–227). ADVERTISING RESEARCH FOUNDATION. https://doi.org/10.2501/JAR-55-2-216-227

Statista. (2024, February). *Global retail e-commerce sales 2014-2027*. Statista. https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

# Intelligent Energy

Zvonimir Šimić, Danijel Topić, Marina Dubravac, Hrvoje Kurkutović
**Using of Battery Storage to Match the Supply and Demand of Electricity in Smart Grids**

Jakub Perička, Tomáš Meravý, Juraj Ďuďák
**Sustainable Microgrid Networks: Advances and Applications**

# Using of Battery Storage to Match the Supply and Demand of Electricity in Smart Grids

**Zvonimir Šimić, Danijel Topić, Marina Dubravac, Hrvoje Kurkutović**
Josip Juraj Strossmayer University of Osijek
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Kneza Trpimira 2B, 31000 Osijek, Croatia
{zvonimir.simic, danijel.topic, marina.dubravac, hrvoje.kurkutovic}@ferit.hr

***Abstract**. Distributed electricity generation from renewable energy sources is in the increase in distribution grids. This gives opportunity to local generation and consumption of electricity. To match the supply and demand of electricity, battery storage (BS) is required. It is required to measure different parameters like electricity consumption, photovoltaic (PV) generation, battery state of charge, and temperature and exchange them in real-time with the utility grid. This is the concept of smart grids where different sensors are used for real-time measures and fast communication is required to ensure a fast response to unpredictable changes in PV generation and electricity consumption. In this paper, the application of BS in smart grids is presented. According to results, 21.42 % less electricity from the utility grid is required with BS which reduces electricity cost by 25.66 %. The amount of electricity exported to the utility grid is reduced by 89.83 % which increases the share of PV generation used for supplying own electricity consumption by 26.92 %.*

**Keywords.** Renewable energy sources, smart grid, battery storage, electricity demand

## 1 Introduction

The share of generation from renewable energy sources is on the rise and the main focus is on wind and solar. Wind power plants are connected to transmission grids while PV power plants can be connected to different grid levels depending on the installed power. PV power plants are very common in distribution grids where the main idea is to generate free electricity and use it at the same location which has a significant contribution to improving energy efficiency and sustainability.

Electricity generation from PV is variable, intermittent, and unpredictable which makes it a challenge to match local consumption with generation. Production from PV is high over the day when people are usually at work and electricity demand is low. During this period, there is an excess of PV generation which is exported to the utility grid. In the evening people are at home and load demand increases. In the evenings, the PV generation is no longer available and electricity for supplying demand is imported from the utility grid (Radhakrishnan & Selvan, 2015).

One possible solution is shifting electricity consumption from periods without PV generation to periods when the PV generation is available. There are time-shiftable devices in households (Pawar, Sampath, Ghosh, & Vittal, 2018) that can be used in any part of the day. As periods with a lot of sun are over days when people are at work, home energy management systems are required to send signals to devices when to turn on.

Another solution to match the supply and demand of electricity is the implementation of BS (Reka & Ramesh, 2015). The idea is to store PV generation excess in periods with more generation than consumption of electricity and then to use stored electricity in periods with lower generation than consumption of electricity (Tutkun, Burgio, Jasinski, Leonowic, & Jasinska, 2021). To decide when to use BS, real-time measurements of the generation and consumption of electricity are required. Different parameters like temperature, humidity, and electricity price should also be measured to control electricity demand and improve electricity use. There is a requirement for fast communication between sensors, control units, and the system operator to enable the real-time match of supply and demand of electricity.

Smart grids use digital modern technologies, including sensors and software to match electricity supply with electricity demand in real-time. The concept of the smart grid is presented in Fig. 1. In the smart grid in addition to energy flow are data flows. Energy flows are one-way energy flows from renewable energy sources to the smart grid and from the smart grid to consumers and two-way energy flows between battery storage and the utility grid. Data flows are two-way between the smart grid and consumers, renewable energy sources, battery storage, and utility grid. Data flows enable the exchange of real-time consumption data, generation data, electricity price, temperature, and other information important for the operation of the smart grid.

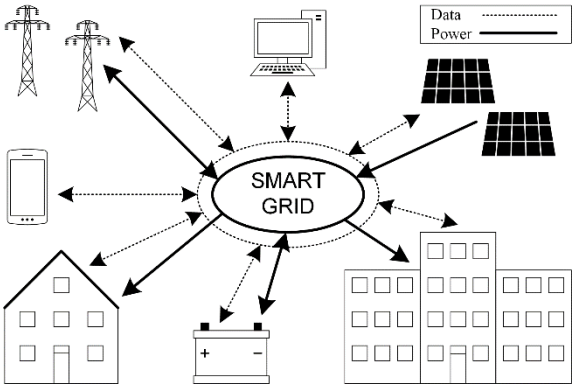



**Figure 1.** Smart grid concept

In this paper, a smart grid with electricity generation from the PV power plant, electricity consumption of the residential building, and BS is simulated. Simulations are done without and with BS to see the influence of using BS to help match electricity demand with electricity generation.

This paper is organized as follows. The first chapter is an introduction to smart grids and the importance of BS systems. In Chapter 2, the methodology of this work is explained with emphasis on the used algorithm and input data. In Chapter 3 results are presented and discussed. The last chapter is the conclusion.

# 2 Methodology

In this chapter, the algorithm for battery storage and required input data is explained.

## 2.1 Battery storage algorithm

The algorithm for deciding when to charge and discharge BS is explained. The algorithm requires the amount of electricity generation and electricity demand in each hour to decide how to use battery storage. The flow chart diagram of the used battery storage algorithm is presented in Fig. 2.

### 2.1.1 BS charging

When the amount of PV generation is higher than the electricity demand in the observed hour, there is an excess of electricity that can be stored in the BS. Battery charging is possible if battery storage is not full. If battery charging is not possible, electricity excess needs to be exported to the utility grid.

### 2.1.2 BS discharging

When the amount of PV generation is lower than the electricity demand in the observed hour, there is a lack of energy that can be delivered from BS if there is enough stored electricity. If there is not enough electricity in battery storage, the required electricity is imported from the utility grid.

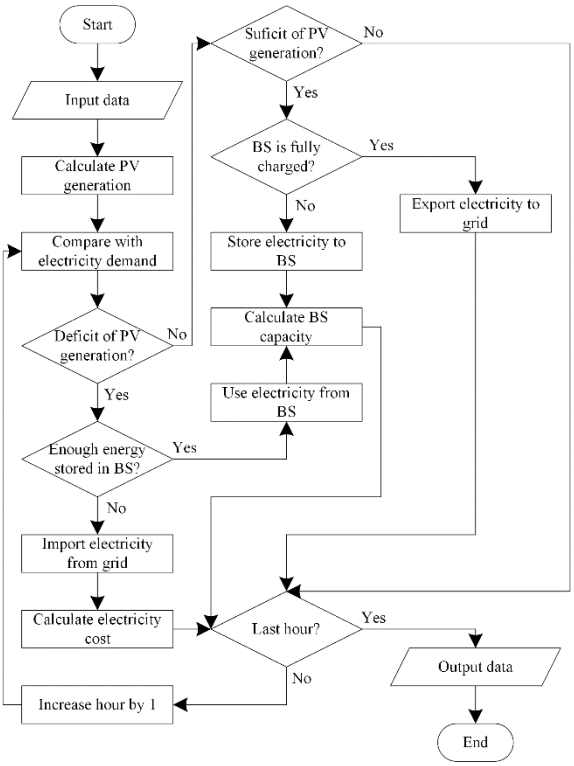



**Figure 2.** Flow chart diagram of BS algorithm

## 2.2 Electricity generation from PV power plant

Electricity generation from the PV power plant is modeled according to the solar irradiation data available from PV GIS (European Commission, 2022) for the selected location with hourly resolution. The amount of PV generation in each hour is calculated according to eq. 1.

$$P_{pv}(t) = G(t) \cdot n_m \cdot A_m \cdot \eta_m \cdot \eta_{inv} \qquad (1)$$

Where:

- $P_{pv}(t)[W]$ – output power of a PV power plant in each hour *t*;
- $G(t)[W/m^2]$ – solar irradiation in each hour *t*;
- $n_m$ – number of PV modules;
- $A_m[m^2]$ – the square area of a PV module;
- $\eta_m$ – efficiency of a PV module;
- $\eta_{inv}$ – efficiency of the PV inverter.

The PV module used in this model is a module with a rated power equal to 500 kW and an efficiency in the amount of 21.3 % (Bluesun Solar, 2022). Dimensions of a PV module are 2.056 m in length and 1.140 m in width which results in a square area equal to 2.34 $m^2$.

The installed power of a PV power plant is 20 kW which means that there are 40 PV modules. The efficiency of the PV inverter is 98 % and the PV output power is calculated. The PV generation during one month is presented in Fig. 3.

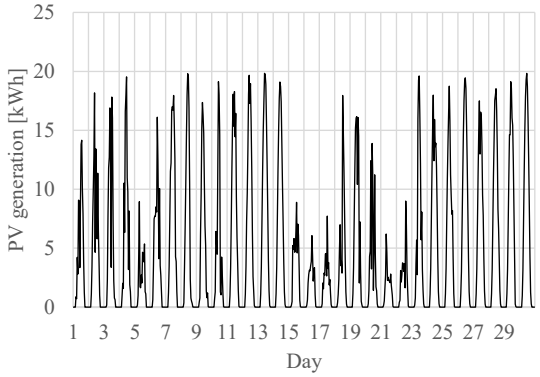



**Figure 3.** The PV generation

## 2.3 Electricity demand

Electricity demand is modeled according to the measured consumption data. Electricity consumption data is measured for one week with a 10-minute time interval and averaged to electricity consumption data with an hourly time interval.

The electricity consumption data is then scaled to building size. Electricity consumption data during one month is presented in Fig. 4.

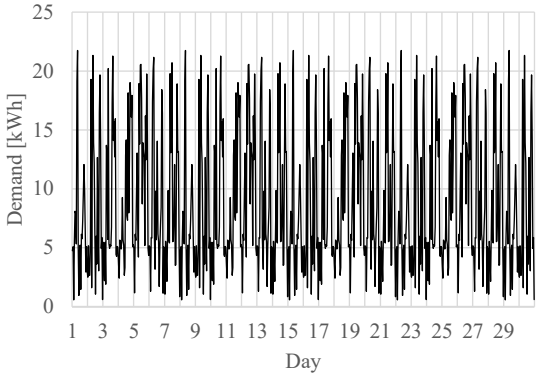



**Figure 4.** Electricity demand

## 2.4 Electricity price

The electricity pricing model used in this work is a tariff billing model for electricity buyers from the household category in Croatia. The electricity billing items for electricity buyers from the household category are presented in Table 1.

VAT for electricity in Croatia is 13 %. The high tariff is from 08:00 to 22:00 and the low tariff is from 22:00 to 08:00 during summer. During winter, the high tariff is from 07:00 to 21:00 and the low tariff is from 21:00 to 07:00 (HEP Elektra, 2023), (HEP ODS, 2022). The assumption is that the smart grid is simulated for one month in summer.

**Table 1.** Billing items for buyers from the household category in Croatia

| Billing item | EUR/kWh |
|---|---|
| Electricity - high tariff | 0.074789 |
| Electricity - low tariff | 0.036697 |
| Grid using fee - high tariff | 0.051762 |
| Grid using fee - low tariff | 0.022563 |
| RES incentive fee | 0.013239 |

## 2.5 Battery storage model

Battery storage in smart grids is required to help match the supply and demand of electricity. Absorbing electricity excess from PV generation helps to increase the share of local electricity generation used to cover local electricity consumption. Using stored energy in periods without available PV generation and charging BS with electricity from the utility grid at a low price helps to reduce electricity costs.

During charging and discharging of battery storage, the state of charge needs to be monitored. The SoC in each hour is calculated according to eq. 2.

$$SoC(t) = SoC(t-1) + P_c(t) \cdot \eta_c - P_d(t)/\eta_d \qquad (2)$$

Where:

- $SoC(t)$– battery $SoC$ in hour $t$;
- $SoC(t-1)$ – battery $SoC$ in previous hour $t-1$;
- $P_c(t)[W]$ – battery charging power in hour $t$;
- $\eta_c$– battery charging efficiency;
- $P_d(t)[W]$ – battery charging power in hour $t$;
- $\eta_d$– battery discharging efficiency.

Battery storage used in this work has a battery capacity of 60 kWh and the rated power is 20 kW which is equal to the installed power of a PV power plant. The battery storage technology is assumed to be Lithium iron phosphate ($LiFePO_4$) with cycle efficiency equal to 92 % and maximum depth of discharge (DoD) equal to 80 %.

# 3 Results

In this chapter results of using BS in a smart grid through simulation are presented. Simulation is done in a Python environment. Results for two different cases are presented. The first case is a simulation without BS and the second case is a simulation with BS. Two different days are observed, a sunny day with a lot of available PV generation and a cloudy day, without enough PV generation. The PV generation and electricity demand during a sunny day are presented in Fig. 5 and during a cloudy day in Fig. 6.

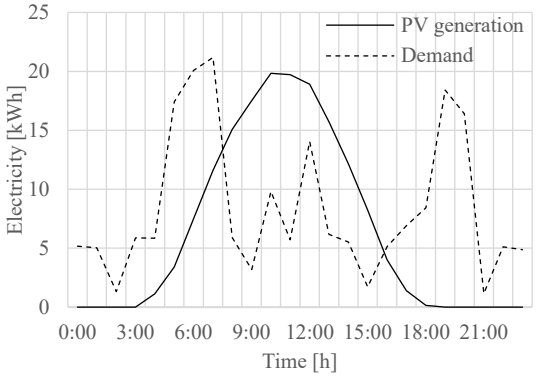



**Figure 5.** PV generation and electricity demand during a sunny day

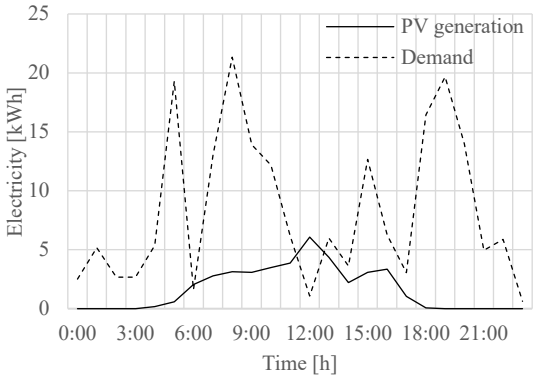



**Figure 6.** PV generation and electricity demand during a cloudy day

## 3.1 Daily results without BS in smart grid

When BS is not available in the smart grid, the entire PV generation excess occurs in hours in which the amount of electricity from the PV power plant is higher than the electricity demand is exported to the utility grid. The high electricity demand, which occurs in periods without enough electricity generation from the PV power plant, is covered with electricity imported from the utility grid, as shown in Fig. 7 for a sunny day.

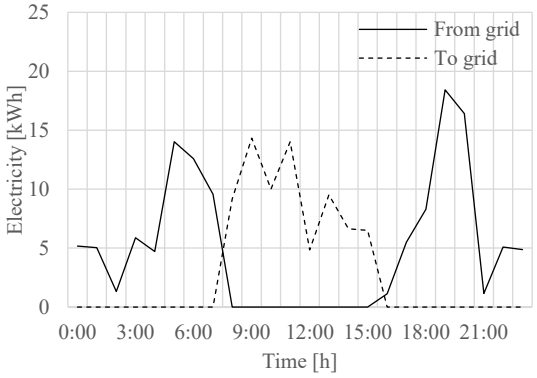



**Figure 7.** The exchange of electricity with the utility grid during a sunny day without BS

During a cloudy day, only a small amount of electricity is exported to the utility grid. Most of the time, the electricity required to supply electricity demand is imported from the utility grid as shown in Fig. 8.

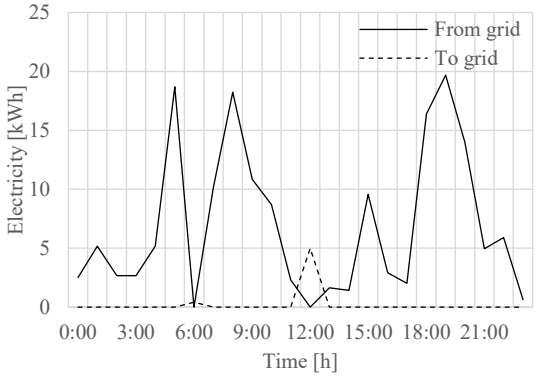



**Figure 8.** The exchange of electricity with the utility grid during a cloudy day without BS

## 3.2 Daily results with BS in smart grid

Using the BS in the smart grid enables to storage excess of PV generation for later use instead of exporting it to the utility grid. BS is charged from 8:00 to 11:00 hours when the PV generation is available during a sunny day and discharged from 16:00 to 22:00 when the amount of PV generation is not enough to supply the entire electricity demand, which can be seen from Fig. 9.

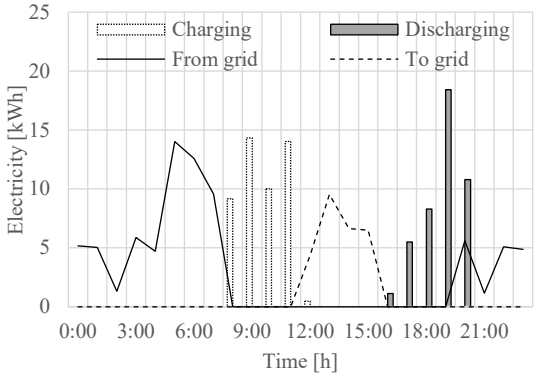



**Figure 9.** The exchange of electricity with the utility grid and the BS during a sunny day

The BS is not charged from 12:00 to 15:00 because the BS is full, which can be seen from Fig. 10 which represents the battery state of charge (SoC). During this period the PV generation excess is exported to the utility grid.

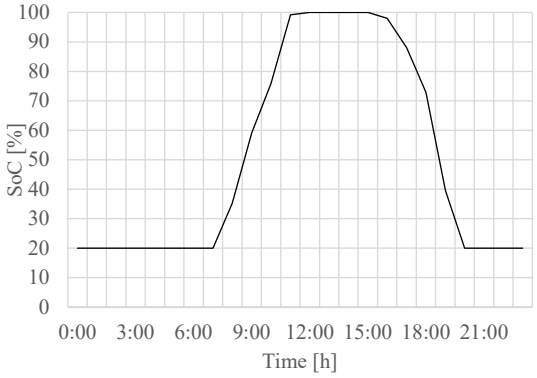



**Figure 10.** The battery SoC during a sunny day

A small excess of PV generation during a cloudy day is now stored to BS, as it is shown in Fig. 11 but this energy is not significant and electricity demand is supplied from the utility grid. The battery SoC is at the minimum level due to low PV generation, as it is shown in Fig. 12.

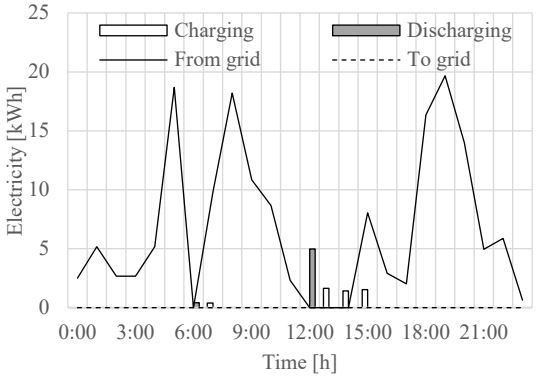



**Figure 11.** The exchange of electricity with the utility grid and the BS during a cloudy day

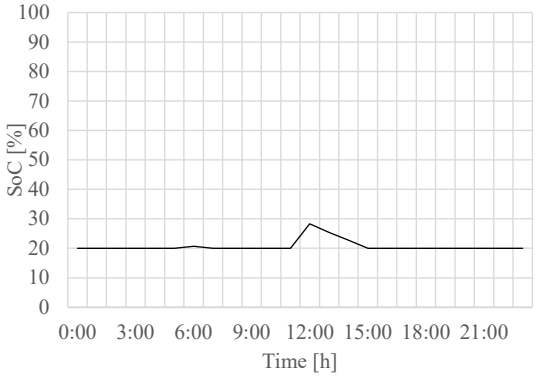



**Figure 12.** The battery SoC during a cloudy day

## 3.3 Monthly results

Total monthly simulation results are presented in Table 2. Electricity demand in the observed month is 6259.66 kWh while estimated PV generation is 3241.72 kWh. The amount of electricity imported from the utility grid is 4073.70 kWh without BS, while the amount of electricity imported with BS is 3201.14 kWh which is a decrease of 872.56 kWh or 21.42 %. The amount of electricity exported to the utility grid without BS is 1055.76 kWh and 107.32 kWh with BS which is the decrease in the amount of 948.44 kWh or 89.83 %.

**Table 2.** Monthly results

| | **Without BS** | **With BS** |
|---|---|---|
| Demand [kWh] | 6259.66 | |
| PV generation [kWh] | 3241.72 | |
| Import from grid [kWh] | 4073.70 | 3201.14 |
| Export to grid [kWh] | 1055.76 | 107.32 |
| Stored to BS [kWh] | - | 948.44 |
| Used from BS [kWh] | - | 872.56 |
| Electricity cost [EUR] | 532.97 | 396.18 |
| PV share [%] | 67.43 | 94.35 |

The amount of PV generation stored in BS is 948.44 kWh and the amount of 872.56 kWh is used from BS to match electricity demand. The amount of usable electricity is lower due to battery cycle efficiency in the amount of 92 %. Without BS, the electricity cost is 532.97 EUR and with BS, the electricity cost is 396.18 EUR. The electricity cost decrease is 136.79 EUR or 25.66 %. The share of PV generation used for supplying electricity demand has increased from 67.43 % without BS to 94.35 % with BS which is an increase of 26.92 %.

# 4 Conclusion

In this work, the importance of using BS to match the supply and demand of electricity in smart grids is explained and confirmed with an example. Without BS, it is not always possible to match the supply of electricity from intermittent and variable renewable RES like PV power plants with the electricity demand. One solution is shifting electricity consumption from periods without PV generation to periods when the supply of electricity from the PV power plant is available. There are time-shiftable devices in households that can be used in any part of the day. As periods with a lot of sun are over days when people are at work, home energy management systems are required to send signals to devices when to turn on.

When BS is available in the smart grid, the free electricity from the PV power plant can be stored in BS in periods with a lot of sun and with low electricity demand and used in periods without sun and with high electricity demand.

According to the results, 21.42 % less electricity is imported from the utility grid with BS than without BS. This leads to a reduction of electricity cost by

136.79 EUR of 25.66 % for one observed month. The amount of electricity exported to the utility grid is reduced by 89.83 % because this difference is stored in BS and used later in periods without enough PV generation. This leads to the increase of the share of PV generation used for supplying the own electricity consumptions for 26.92 %.

In future work, the BS algorithm can be improved to find the optimal time for charging and discharging. In addition to charging BS when there is available PV generation, BS can be charged with electricity from the utility grid when the electricity price is low.

## Acknowledgments

The research was supported by the European Union's Horizon Europe Framework Programme: HORIZON-WIDERA-2023-ACCESS-04, Pathways to Synergies – Coordination and Support Actions - under the project name SynGRID - Creating synergies in Widening countries on the topic of low-voltage grid management (grant number 101160145). The funding bodies had no involvement in the preparation of the manuscript.

## References

Bluesun Solar. (2022). BSM500PM5-72SB PV module datasheet. Retrieved from https://www.bluesunpv.com/uploadfile/downloads/BSM500PM-72SB%20500W%20.pdf

European Commission. (2022). Photovoltaic Geographical Information System (PV-GIS). Retrieved from https://joint-research-centre.ec.europa.eu/photovoltaicgeographical-information-system-pvgisn_en

HEP Elektra. (2023). Electricity prices for households from Apr 1, 2024. Retrieved from https://www.hep.hr/elektra/kucanstvo/tarifne-stavke-cijene/1547

HEP ODS. (2022). Grid using fees for households from Apr 1, 2022. Retrieved from https://www.hep.hr/ods/kupci/kucanstvo/tarifne-stavke-cijene/160

Pawar, P., Sampath, S., Ghosh, T., & Vittal, K. (2018). Load Scheduling Algorithm Design for Smart Home Energy Management System. *7th International Conference on Power and Energy (PECon)* (pp. 304-309). Kuala Lumpur: IEEE.

Radhakrishnan, A., & Selvan, M. P. (2015). Load scheduling for smart energy management in residential buildings with renewable sources. *18th National Power Systems Conference (NPSC)*. Guwahati: IEEE.

Reka, S. S., & Ramesh, V. (2015). A novel integrated approach of energy consumption scheduling in smart grid environment with the penetration of renewable energy. *International Journal of Renewable Energy Research-IJRER*, 1196-1205.

Tutkun, N., Burgio, A., Jasinski, M., Leonowic, Z., & Jasinska, E. (2021). Intelligent Scheduling of Smart Home Appliances Based on Demand Response Considering the Cost and Peak-to-Average Ratio in Residential Homes. *Energies*.

# Sustainable Microgrid Networks: Advances and Applications

**Jakub Perička, Tomáš Meravý, Juraj Ďuďák**
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava
Jana Bottu 2781/25, Trnava, Slovakia
{jakub.pericka, tomas.meravy, juraj.dudak}@stuba.sk

**Abstract.** *This paper explores the development and implementation of distributed control strategies within AC microgrids, emphasizing their role in renewable energy integration and enhancing environmental sustainability. The research discusses the motivation for integrating renewable energy sources to reduce the carbon footprint and the limitations of conventional distribution networks. Microgrids are introduced as a viable solution to these challenges, offering improved resilience and efficiency in distributed energy systems. The paper classifies microgrids and details their essential components, including energy loads, distributed energy resources, and energy storage systems. The study further delves into various control methods, highlighting the hierarchical control structure with primary, secondary, and tertiary control layers. By examining recent advancements and addressing challenges in microgrid technology, the paper aims to present a comprehensive overview of how modern microgrids can contribute to a more sustainable energy future.*

**Keywords.** microgrid, energy efficiency, control algorithms, distributed energy sources

# 1 Introduction

As the world's population continues to grow, and technological progress accelerates with it, the need for energy for humanity to function is increasing at a staggering rate. This growing demand puts great pressure on the existing energy infrastructure and requires, in addition to its modernization, a transition to other, more sustainable and resilient energy systems. According to the International Energy Agency (IEA), global electricity demand is expected to grow by around 3% per year until 2025, which would mean an increase of around 2500TWh. This is more than double Japan's current annual electricity consumption (Electricity Market Report, 2023). The transition to renewable energy sources(RES) is important to mitigate the adverse effects of climate change and reduce greenhouse gas emissions, however, their intermittent and variable nature, such as with solar and wind power, complicates their integration and poses significant technical challenges for traditional distribution systems (Denholm, 2013). Microgrids(MGs), i.e. decentralized energy systems that can operate as a stand-alone energy island or in cooperation with the main distribution grid, have emerged as a possible solution to these challenges. The integration of MGs achieves increased network reliability and efficiency, as well as better opportunities for the integration of RES. By localizing energy production and consumption, line losses are reduced and energy security is increased, especially in isolated or remote areas (Shahbazitabar, 2021) and areas with high risk of natural disasters. Despite their advantages, the complexity of their control algorithms hinders the spread of MGs. Effective MG management requires strategies for balancing power generation and connected load, managing energy storage, and also ensuring synchronization and coordination with the main grid. Continuous refinement of these algorithms is needed to maximize the benefits of MGs and accelerate the transition to RES.

## 1.1 Legislative motivation for developmentof microgrids

Greenhouse gas(GHG) emissions are a serious problem that needs to be addressed effectively. Within the European Union, this is addressed by the 'European Green Deal' plan, which aims to achieve GHG neutrality by 2050 and to make Europe the first carbon-neutral continent (A European Green Deal, 2019). Although much attention has been paid to reducing emissions in the transport sector, a large part of the emissions is caused by the energy sector, or the production, transport and consumption of energy. In 2012, the European Union adopted Directive 2012/27/EU, which aimed to increase the efficiency of energy use (and thus reduce energy consumption) by 20% compared to 1990, i.e. to a maximum of 959 Mtoe - Million Equivalent Tonnes of Oil - (EU Directive, 2023). The final energy consumption in 2020 was at 907 Mtoe, marking the success of this directive. Also, this success motivated the creation of a new directive, which was recently put into force in October 2023. The EU/2023/1791 Energy Efficiency Directive aims to reduce energy consumption by a further 11.7% by 2030 compared to 2020, i.e. to a

maximum of 763 Mtoe of final consumption. This also includes a high promotion of decolonization of heating and cooling systems, maximizing the use of waste heat and RES (Energy Efficiency Directive, 2023). This directive is part of the 'REPowerEU' plan and one of the objectives of this plan is to increase the overall share of RES to 45% by 2030, up from 23% in 2022. The share of electricity consumption in total energy consumption has been fairly stable over the last 11 years, averaging 22.93% with a standard deviation of 0.18% . This sector also shows the most significant progress in the implementation of RES with an average annual increase of 1.405%(Eurostat, 2024). In 2022, up to 41.17% of the electricity in the EU was from RES, while if we consider the share of electricity in total energy consumption and the targets of the "REPowerEU" plan, by 2030 this share should be at 80.55% (REPowerEU, 2023). Such a change will have a major impact on the aforementioned GHG emissions. Emissions due to electricity consumption (in grams of CO2 equivalent emissions per kilowatt-hour [gCO2eq/kWh]) have seen a significant decline over the past decades, which has had a positive impact on the overall reduction of emissions even though total consumption has not seen sharp declines (Eurostat, 2024). This has been a significant motivation in the deployment of RES, and it can be expected that efforts to deploy these technologies will only increase in the coming years.

# 2 Microgrids overview

Since the term microgrid was first introduced in 2001, several definitions have been developed for this type of energy network and its functional principle. B. Lasseter (Lasseter, 2001) first defined a MG as a group of micro-energy sources, energy storage systems and loads that represents a single entity that has central control. According to Lasetter, the basis of the MG concept is the idea of a controlled interconnection between the microgrid and the main distribution network. This definition is followed by the generally accepted definition (Shahbazitabar, 2021):" Microgrids are energy distribution systems comprising loads and distributed energy resources (such as distributed generators, batteries or controllable loads) that are controllable and operate in a coordinated manner, whether connected to the main electricity grid and/or as an energy island." MGs have the added economic potential enabled by ancillary services for energy trading between the MG and the main grid. The concept of "energy islands" was explored by Rettig et al. (Rettig, 2023) and laid the foundations of a theoretical framework defining this concept, which is often associated with MGs. The types of energy islands are defined by their boundaries, which can be geographical, as in the case of real islands, political in the case of isolated states or provinces, and technological for the separation of available energy services. While many characteristics are shared with MGs, such as decoupling from larger grids, local energy management, easier integration of renewables, and the use of battery storage units, energy islands are largely isolated on a permanent basis in the context of their geopolitical and physical conditions. This difference shows the adaptability of MGs not only in remote locations but also in the context of easily interconnected grids (e.g. cities, towns, households).

## 2.1 Microgrid structure

The elements of MGs can be divided into several groups:

**Energy loads** are all devices that draw electricity from the grid and convert it into other types of energy. We divide loads into two subgroups:

- Non-controllable loads, which include a wide range of appliances that cannot be controlled by the grid. These can be household appliances, production machines, directly controlled electric motors, computing equipment, etc. These devices are a challenge for optimizing control in the MG and their consumption must be monitored.
- Controllable loads whose power draw can be controlled based on the needs of the MG or by external intervention. Examples of these systems include air conditioning units with adjustable target temperature, controllable lighting systems, and frequency-controlled electric motors. This group also includes chargers for electric vehicles, which are a rapidly expanding technology and bring new opportunities and challenges to the MG system. Controllable loads are the basis of supply-demand management strategies and allow control systems to balance demand to some degree to ensure grid stability.

**Distributed energy resources (DERs)** are small-scale units for local power generation located close to energy loads. They may include (and in most use cases do include) RES. However, fossil fuel consuming generators can also be used as distributed resources, which can be used as a redundant source when renewable energy is scarce. The choice of DERs for use in a MG depends on local resources as well as environmental and economic factors.

**Energy storage systems (ESS)** are an essential element of the MG and enable the penetration of RES that have intermittent energy supply characteristics. There can be several types of these systems, each with its own advantages and disadvantages of use, as well as specific application cases.

**Power electronics** layer interfaces are devices that provide conversion between different types of distribution networks (alternating-current(AC) and direct-current(DC)) and also adjust network characteristics (e.g. voltage and frequency). They play a fundamental role in controlling the flow of power within the MG, together with monitoring functions. The basic elements

are AC-to-DC converters and vice versa, which allow the connection of different types of sources and loads (Abbasi, 2023).

**Control systems**, which perform the task of monitoring MG parameters and the status of all elements, communicating with MG elements, and executing control algorithms based on data acquired in real time. They are key to ensure the functionality of the network as they control the power flow within the network and the network mode (grid connected/isolated).

## 2.2 Classification of microgrids

MGs are a broad concept and therefore further classification is needed to specify their size, purpose, mode of operation and other parameters (Shahbazitabar, 2021). Classification by type:

- **Campus MGs:** usually in schools and on university campuses where they are used for research of energy management, distributed energy resources or controllable loads. Such a grid is also used, for example, by the University of California, San Diego, where it supplies electricity to a 450 hectar campus from two 13.5MW gas turbines, a 3MW steam turbine and 1.2MW of solar panels. Together, these sources cover 85% of the electrical consumption, 95% of the heating and 95% of the cooling within the campus (San Diego Microgrid, 2024).
- **Military MGs:** mainly used for energy security and resilience of military bases. In 2019, a military MG was completed at Kirkland Air Force Base (USA) (Department of Energy, 2020) and is composed of modular parts, each with power conversion, control, protection, and storage functions. These modular parts are interconnected by a single bus with a bipolar DC voltage of $\pm$375 V. The MG connects photovoltaic panels, batteries and gas generators together with controllable loads (electric vehicle chargers).
- **Residential MGs:** These include single households or residential communities and provide them with increased reliability, potential cost savings and increased energy independence. They help reduce the carbon footprint of households and maximize the use of RES. Examples of such energy communities can be found all over the world, one of them is the village of Feldheim in Germany, which covers 100% of its own consumption from RES (Euronews Energy, 2022). Another example is the Isle of Eigg energy island in Scotland, which is completely decoupled from the main energy grid and combines solar, wind and hydro power in its MG, thus reducing the island's dependence on distributed diesel generators and using this energy to heat community centres in case of surplus energy production (Electric, 2024).
- **Commercial MGs:** specialized for use in the commercial sphere to increase the reliability and maintainability of shopping centers, corporate campuses or hospitals. They enable energy management in the hands of businesses and other economic benefits such as reduced energy costs and the ability to sell excess energy to the national grid. As computing capacity moves to the Cloud space, the energy demands of data centers are also increasing, which is why US retailer eBay decided to connect the "Topaz Data Center" in the US state of Utah to the MG to eliminate their dependence on uninterruptible power supplies (UPS). This network is based on a group of biogas generators, which also reduced the energy consumption of local coal-fired power plants (Data Center Knowledge, 2012).
- **Industrial microgrids:** Usually larger than other types of MGs. These grids allow industrial companies to comply with ever-tightening legislation on the energy intensity of production while reducing their own energy costs. Another reason may be insufficient supply capacity from the main grid, as is the case with the Tritax Symmetry Logistics Park in the UK. This MG contains enough renewable energy and energy storage to keep the logistics center running in islanded mode, while at times of low energy consumption the entire center can be powered solely from batteries (Microgrid Knowledge, 2020).

According to their size, MGs are divided into:

- Small networks typically with a capacity in the order of hundreds of kW.
- Large networks with capacity in MW

In terms of use they can be divided into:

- Networks for high quality energy that ensure stable voltage levels without interference and fluctuations. This type of networks is mainly intended for industry and other sectors where even small fluctuations in power quality can lead to service interruptions and financial losses (e.g., data centers) (Azar, 2000).
- Resilience-oriented networks designed to maximize the reliability of power supply even in crisis situations. These networks can operate in grid-connected mode and also in fully isolated mode, and provide rapid restoration of supply during "low frequency, high impact" events such as natural disasters (Chi Xu, 2017).

Division in terms of type of electrical voltage:

- **AC microgrids(ACMGs)** use low and medium AC voltages to transmit power from sources to consumers. The advantage is easier interoperability with the external distribution network and compatibility with appliances operating at traditional low voltage levels. The DC distributed sources and batteries are connected to a common AC bus via DC-to-AC converters, with some of the energy being lost in this conversion and transmission. The AC distribution network can be single-phase, three-phase with a neutral conductor, and three-phase without a neutral conductor. To improve the network parameters and better behavior during faults, it is important to control the loads connected to the MG system by, for example, simply disconnecting them from the power sup-

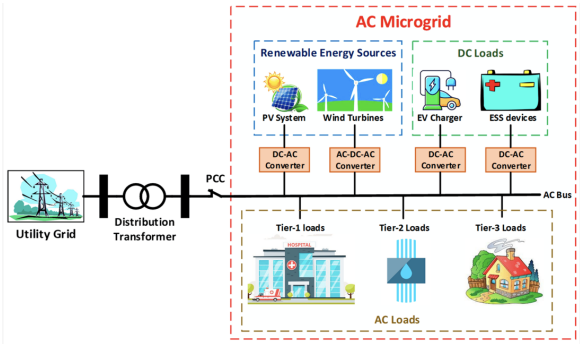



**Figure 1:** Diagram of a typical ACMG (Abbasi, 2023).

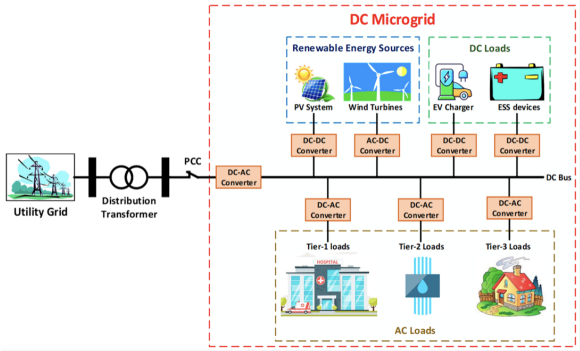



**Figure 2:** Diagram of a typical DCMG (Abbasi, 2023).

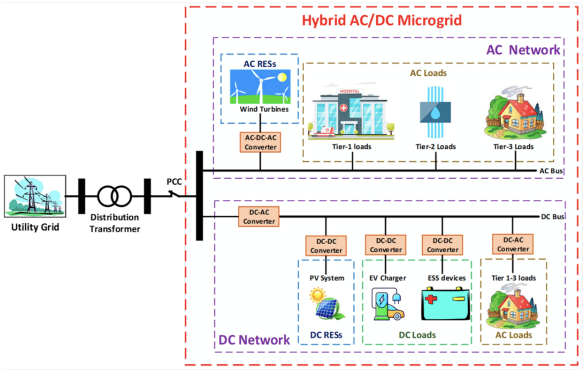



**Figure 3:** Diagram of a typical HMG (Abbasi, 2023).

ply. Moran (Moran, 2016) defines the division of AC loads into three levels according to the possibility of disconnection from the grid:

- Tier-1 loads are the most critical infrastructure that must never be disconnected as part of load shedding. This category includes hospitals, ambulance services, emergency dispatching, and the like.
- Tier-2 system's peak energy load or to shift it over time. These loads may be disconnected for the purpose of starting additional generators during outages. These loads include heating, air conditioners, pool filters, etc.
- Tier-3 loads can only be disconnected in an emergency, for example to prevent a complete failure. Residential and commercial buildings with backup generators are in this category. Some loads in this category can also be used to generate power to the grid to start after a complete outage.

The structure of the AC microgrid is shown in Fig.1

- **DC microgrids(DCMGs** are a new, emerging technology that enables direct connection of DC power supplies, batteries and DC loads. Advantages are lower line losses, easy connection to DC appliances, and fewer synchronization interfaces with AC networks. The DC voltage distribution network can be monopolar (one conductor, the role of the neutral conductor is fulfilled by the earth), unipolar/homopolar (one conductor for the supply, one for the return path of the current) and bipolar (two conductors with opposite voltage polarity, the centre point is grounded). The structure of the DC microgrid is shown in Fig.2

- **Hybrid microgrids(HMGs)** combine the advantages of the two types described above for increased flexibility, efficiency and reliability. They include both AC and DC buses interconnected by rectifiers and inverters. The advantage of this approach is the ability to easily interconnect DC elements with DC elements and greater functional flexibility. The disadvantage is the higher complexity of control systems and algorithms for synchronizing both types of distribution systems (Sinha, 2024) The structure of the hybrid microgrid is shown in Fig.3

# 3 Control of microgrids

MGs present a cyber-physical system, with the cyber part providing services like inverter control, economical dispatch or power flow management based on measured and collected data (Shan, 2022). Control mechanisms within the MG are usually segmented according to a hierarchical control structure consisting of three layers: primary, secondary and tertiary. Each subsequent layer creates setpoints and references for the layer below. As we go higher in this hierarchical model, the period between control signals (control period) increases as well from the primary control working in miliseconds up to the tertiary control methods generating commands in periods ranging from several minutes up to hours (Abhishek, 2020).

## 3.1 Primary control

Primary control is responsible for regulating each inverter output to ensure power sharing between DGs. For this purpose, droop control is usually used to droop frequency and voltage in relation to active and reactive power, respectively. This control method is decen-

tralized, so it operates locally on the level of a single inverter-interfaced DG using droop method given as:

$$\omega = \omega^* - m_p(P - P^*) \quad (1)$$

$$E = E^* - n_p(Q - Q^*) \quad (2)$$

With $\omega$ and $E$ representing the output frequency and voltage command, respectively. $\omega^*$ represents the nominal grid frequency, $P$ as measured active power and $P^*$ being the active power reference. The voltage control equation follows the same structure with reactive power measurement $Q$ and reference $Q^*$. $m_p$ and $n_p$ are droop coefficients which adjust droop characteristic angle. It achieves good results in terms of power sharing between DGs, however with higher droop coefficients power sharing between DGs improves, but it also introduces voltage and frequency deviation from nominal grid values. Droop control outpus serves as a setpoint for inner control voltage(external) and current(internal) control loops. This method can be further augmented with methods for enhancing power stability and compensating power issues like negative sequence currents (Zhao, 2017) or transmission line impedance (Peng, 2019).

### 3.2 Secondary control

To compensate the primary network output deviation and increase grid stability during transients, secondary control is employed. There are several methods used on this control layer, which we can further divide depending on the system architecture. *Centralized control methods* are computed using a single centralized controller, which collects data from the whole MG and generates references for primary control in DGs. This approach allows to contain the whole control layer in a single device with easier synchronization. However, the main issue is a single point of failure, which can negatively influence primary controllers in case of expected or unexpected power issues (Dadi, 2023). The opposite to this is fully *decentralized system* in which the secondary controllers operate on each DG inverter. Absence of communication makes this design attractive, because of the potential issues that come with packet transfer like communication delay or packet loss. This approach however, is infeasible in case of secondary control because it prevents the controllers to cooperate in reaching a common optimization goal and may cause instability during transients (Singh, 2019). A very promising compromise between these two architectures is *distributed control*, which gained interest of many scientists in the recent years. This communication architecture consists of secondary controllers operating on local DG level, but employing communication links between neighboring nodes (Moradi, 2023). This not only helps distribute computation load between the controllers, it also allows cooperation in optimization problems and converge on nominal values for both voltage and frequency. To describe a distributed communication network, graph theory is commonly used (Liu, 2024). As mentioned before, by introducing communication between nodes, issues associated with packet loss and communication delays are also introduced. Researchers in (Huang, 2024) used a Lyapunov-Krasovskii functional to provide adaptive delay-independent secondary control with stability proved by extensive testing including load variations and plug-and-play scenarios. Another approach presented in (Wu, 2019) employs lead-lag compensation and gain regulation blocks to introduce phase lead and compensate for communication delays with real-time experimental validation.

### 3.3 Tertiary control

The last control layer in the hierarchical control model is tertiary control. The purpose of this layer is to provide energy management functionalities, control power flow within the MG and optimize the economic dispatch (Shan, 2023). As we go higher in the hierarchical model, the control period lengthens and with tertiary control, the time between commands can be minutes, even hours depending on the optimization goal and prediction horizon. There is a general discussion on whether the tertiary control needs to be employed in islanded MGs as some of the mentioned functionalities can be associated with secondary control as well (Shan, 2022). In grid-connected mode tertiary control also handles electricity market engagement and cooperation in the case of multi-MG networks (Rashidi, 2021). To achieve the stated goals, optimization techniques are used, such as mixed-integer linear or quadratic programming (Vilaisarn, 2022; Jabr, 2021), optimal power flow methods (Brandao, 2020; Chopra, 2022), genetic algorithms (Hernandez, 2023; Kothakotla, 2021) or particle swarm optimization(PSO) (Shan, 2022; Phommixay, 2020).

## 4 Conclusion

MGs are a technology area with great research and application potential and represent one of the few options for 100% integration of RES, mainly through the use of ESS, which form the core of the MG concept. MGs are based on the separation of services into small, self-managed units, which improves the reliability and flexibility of energy for end consumers. A fundamental prerequisite for proper operation and economic viability is the use of sophisticated algorithms at different levels of MG management, which take into account the instantaneous characteristics of RES, the prediction of energy load, the parameters of energy storage and also the price of energy. The objective of this article is to provide a research platform based on the conducted problem analysis of the studied area.

The follow-up research will be aimed mainly on tertiary control algorithms for MGs with a focus on better

utilization of battery storage and improvement of the economic efficiency of MG operation using prediction and optimization algorithms, such as PSO and genetic algorithms.

# References


Electricity Market Report 2023: Executive Summary. (2023). Retrieved from https://www.iea.org/reports/electricity-market-report-2023/executive-summary

Paul Denholm, Maureen Hand, Maddalena Jackson, and Sean Ong. Grid Integration of Renewable Energy: Flexibility, Innovation, and Experience. 2013. Technical Report NREL/TP-6A20-60451. National Renewable Energy Laboratory (NREL).

Maryam Shahbazitabar, Amjad Anvari-Moghaddam, and Josep M. Guerrero, "An Introduction to Microgrids, Concepts, Definition, and Classifications." 2021. In *Microgrids: Advances in Operation, Control, and Protection*, Ed. Amjad Anvari-Moghaddam et al., Cham: Springer International Publishing, pp. 3-16. ISBN: 978-3-030-59750-4. doi: 10.1007/978-3-030-59750-41.

European Commission. A European Green Deal. (2019). Retrieved December 20, 2023, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

*Directive (EU) 2023/1791 of the European Parliament and of the Council of 13 September 2023 on energy efficiency and amending Regulation (EU) 2023/955 (recast).* (2023).

European Union. Energy Efficiency Directive. (2023). Retrieved December 20, 2023, from https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-directive_en

Eurostat. Final energy consumption by sector. (2024). Retrieved December 20, 2023, from https://ec.europa.eu/eurostat/databrowser/product/view/ten00123?lang=en

European Commission. REPowerEU: affordable, secure and sustainable energy for Europe. (2024). Retrieved December 21, 2023, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/repowereu-affordable-secure-and-sustainable-energy-europe_en

Eurostat. Air emissions accounts by NACE Rev. 2 activity. (2024). Retrieved December 21, 2023, from https://ec.europa.eu/eurostat/databrowser/product/view/env_ac_ainah_r2?lang=en

B. Lasseter, "Microgrids [distributed power generation]," in *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, vol. 1, 2001, pp. 146-149 vol.1. doi: 10.1109/pesw.2001.917020.

E. Rettig, I. Fischhendler, and F. Schlecht, "The meaning of energy islands: Towards a theoretical framework," *Renewable and Sustainable Energy Reviews*, vol. 187, 2023, p. 113732. ISSN: 1364-0321. doi: 10.1016/j.rser.2023.113732.

UC San Diego Microgrid. Retrieved January 5, 2024, from Microgrid Knowledge Symposiums. https://microgrid-symposiums.org/microgrid-examples-and-demonstrations/uc-san-diego-microgrid/

Department of Energy. "Kirtland Air Force Base DC Microgrid is Fully Operational." (2020). Retrieved January 5, 2024, from https://www.energy.gov/oe/articles/kirtland-air-force-base-dc-microgrid-fully-operational

Euronews. Energy: The self-sufficient German town that gives back to the grid. (2022). Retrieved January 5, 2024, from https://www.euronews.com/2022/09/30/energy-the-self-sufficient-german-town-that-gives-back-to-the-grid

Eigg Electric. Isle of Eigg. Retrieved January 5, 2024, from http://isleofeigg.org/eigg-electric/

Data Center Knowledge. eBay: Bloom Boxes Will Power Utah Data Center. (2012). Retrieved January 6, 2024, from https://www.datacenterknowledge.com/archives/2012/06/21/ebay-bloom-boxes-will-power-utah-data-center

Microgrid Knowledge. Owning Your Energy: The Potential of Microgrids for Industry. (2020). Retrieved January 9, 2024, from https://www.microgridknowledge.com/sponsored/article/33014277/rolls-royce-solutions-america-owning-your-energy-the-potential-of-microgrids-for-industry

K. Azar, "Power consumption and generation in the electronics industry. A perspective," in *Sixteenth Annual IEEE Semiconductor Thermal Measurement and Management Symposium (Cat. No.00CH37068)*, 2000, pp. 201-212. doi: 10.1109/stherm.2000.837085.

Yuan Chi and Yan Xu, "Resilience-oriented microgrids: A comprehensive literature review,"

in *2017 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, 2017, pp. 1-6. doi: 10.1109/ISGT-Asia.2017.8378346.

Bill Moran, "Microgrid load management and control strategies," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, 2016, pp. 1-4. doi: 10.1109/tdc.2016.7520025.

Rishi Ratan Sinha and Neeraj Kanwar, "Chapter Four - Hybrid microgrids: architecture, modeling, limitations, and solutions," in *Modeling and Control Dynamics in Microgrid Systems with Renewable Energy Resources*, Ed. Ramesh C. Bansal, Jackson J. Justo, and Francis A. Mwasilu, Academic Press, 2024, pp. 65-82. ISBN: 978-0-323-90989-1. doi: 10.1016/B978-0-323-90989-1.00012-9.

Maysam Abbasi et al., "Review on the Microgrid Concept, Structures, Components, Communication Systems, and Control Methods," *Energies*, vol. 16, no. 1, 2023. ISSN: 1996-1073. doi: 10.3390/en16010484.

Shan, Y., Pan, A., & Liu, H. (2022). A switching event-triggered resilient control scheme for primary and secondary levels in AC microgrids. *ISA Transactions*, 127, 216-228. doi:10.1016/j.isatra.2022.02.039.

Abhishek, A., Ranjan, A., Devassy, S., Kumar Verma, B., Ram, S.K., & Dhakar, A.K. (2020). Review of hierarchical control strategies for DC microgrid. *IET Renewable Power Generation*, 14, 1631-1640. 10.1049/iet-rpg.2019.1136.

Zhao, X., Guerrero, J. M., Savaghebi, M., Vasquez, J. C., Wu, X., & Sun, K. (2017). Low-Voltage Ride-Through Operation of Power Converters in Grid-Interactive Microgrids by Using Negative-Sequence Droop Control. *IEEE Transactions on Power Electronics*, 32(4), 3128-3142. doi:10.1109/TPEL.2016.2570204.

Peng, Z., Wang, J., Bi, D., Wen, Y., Dai, Y., Yin, X., & Shen, Z. J. (2019). Droop Control Strategy Incorporating Coupling Compensation and Vir-tual Impedance for Microgrid Application. *IEEE Transactions on Energy Conversion*, 34(1), 277-291. doi:10.1109/TEC.2019.2892621.

Dadi, R., Meenakshy, K., & Damodaran, S. (2023). A Review on Secondary Control Methods in DC Microgrid. *Journal of Operation and Automation in Power Engineering*, 11(2), 105-112. doi:10.22098/joape.2022.9157.1636.

Singh, P., Paliwal, P., & Arya, A. (2019). A Review on Challenges and Techniques for Secondary Control of Microgrid. *IOP Conference Series: Materials Science and Engineering*, 561(1), 012075. doi:10.1088/1757-899X/561/1/012075.

Moradi, M., Heydari, M., & Zarei, S. F. (2023). An overview on consensus-based distributed secondary control schemes in DC microgrids. *Electric Power Systems Research*, 225, 109870. doi:10.1016/j.epsr.2023.109870.

Liu, Z., Li, J., Su, M., Liu, X., & Yuan, L. (2024). Stability Analysis of Equilibrium of DC Microgrid Under Distributed Control. *IEEE Transactions on Power Systems*, 39(1), 1058-1067. doi:10.1109/TPWRS.2023.3266244.

Huang, X., et al. (2024). Distributed Secondary Control for Islanded Microgrids Considering Communication Delays. *IEEE Access*, 12, 64335-64350. doi:10.1109/ACCESS.2024.3393770.

Wu, X., Xu, Y., He, J., Shen, C., Chen, G., Vasquez, J. C., & Guerrero, J. M. (2019). Delay-Dependent Small-Signal Stability Analysis and Compensation Method for Distributed Secondary Control of Microgrids. *IEEE Access*, 7, 170919-170933. doi:10.1109/ACCESS.2019.2955090.

Shan, Y., Ma, L., & Yu, X. (2023). Hierarchical Control and Economic Optimization of Microgrids Considering the Randomness of Power Generation and Load Demand. *Energies*, 16(5503). doi:10.3390/en16145503.

Rashidi, R., Hatami, A., & Abedini, M. (2021). Multi-microgrid energy management through tertiary-level control: Structure and case study. *Sustainable Energy Technologies and Assessments*, 47, 101395. doi:10.1016/j.seta.2021.101395.

Vilaisarn, Y., Moradzadeh, M., Abdelaziz, M., & Cros, J. (2022). An MILP formulation for the optimum operation of AC microgrids with hierarchical control. *International Journal of Electrical Power & Energy Systems*, 137, 107674. doi:10.1016/j.ijepes.2021.107674.

Jabr, R. A. (2021). Mixed-Integer Convex Optimization for DC Microgrid Droop Control. *IEEE Transactions on Power Systems*, 36(6), 5901-5908. doi:10.1109/TPWRS.2021.3083678.

Brandao, D. I., Ferreira, W. M., Alonso, A. M. S., Tedeschi, E., & Marafão, F. P. (2020). Optimal Multiobjective Control of Low-Voltage AC Microgrids: Power Flow Regulation and Compensation of Reactive Power and Unbalance. *IEEE Transactions on Smart Grid*, 11(2), 1239-1252. doi:10.1109/TSG.2019.2933790.

Chopra, S., Vanaprasad, G. M., Tinajero, G. D. A., Bazmohammadi, N., Vasquez, J. C., & Guerrero, J. M. (2022). Power-flow-based energy management of hierarchically controlled islanded

AC microgrids. *International Journal of Electrical Power & Energy Systems*, 141, 108140. doi:10.1016/j.ijepes.2022.108140.

Hernandez-Mayoral, E., Madrigal-Martinez, M., Mina-Antonio, J.D., Iracheta-Cortez, R., Enriquez-Santiago, J.A., Rodriguez-Rivera, O., Martinez-Reyes, G., & Mendoza-Santos, E. (2023). A Comprehensive Review on Power-Quality Issues, Optimization Techniques, and Control Strategies of Microgrid Based on Renewable Energy Sources. *Sustainability*, 15, 9847. doi: 10.3390/su15129847.

Kothakotla, V., & P, B. K. (2021). Analysis and Design of Robust PID Controller for Grid Voltage Control of Islanded Microgrid Using Genetic Algorithm. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2021, pp. 719-726. doi:10.1109/ICICCS51141.2021.9432094.

Shan, Y., Hu, J., & Liu, H. (2022). A Holistic Power Management Strategy of Microgrids Based on Model Predictive Control and Particle Swarm Optimization. *IEEE Transactions on Industrial Informatics*, 18(8), 5115-5126. doi:10.1109/TII.2021.3123532.

Phommixay, S., Doumbia, M. L., & Lupien St-Pierre, D. (2020). Review on the cost optimization of microgrids via particle swarm optimization. *International Journal of Energy and Environmental Engineering*, 11(1), 73-89. doi:10.1007/s40095-019-00332-1.

# Intelligent Systems

Jože M. Rožanec, Dunja Mladenić
**WaAUROCC: measuring how steep the ROC AUC curve is**

Barbara Kovačić, Anja Brnjaković, Gaurish Thakkar
**Beyond Words: Sentiment Analysis of Croatian Language Attitudes**

Gaurish Thakkar, Vanja Štefanec, Daša Farkaš, Marko Tadić
**Building a Large Language Model for Moderately Resourced Language: A Case of Croatian**

Oliver Lohaj, Ján Paralič, Ruslana Buinytska
**Exploring usability in combining data sources for detecting toxic behavior in social media**

Ana Meštrović
**A Review of Methods for Analyzing Online Media Texts Using Large Language Models / Pregled metoda za analizu online medijskih tekstova primjenom velikih jezičnih modela**

Lucia Načinović Prskalo
**Text Analysis Methods for Identifying and Recognizing Gender Bias, Inequality, and Violence in Language: An Overview / Metode analize teksta za identifikaciju i prepoznavanje rodne pristranosti, nejednakosti i nasilja u jeziku: Pregled**

Dino Vlahek, Dijana Oreški
**Data Spaces in the Republic of Croatia: Potential and Opportunities**

Ivana Ružić, Igor Balaban
**Social Robots: Applications, Challenges, and Future Directions - literature review**

# WaAUROCC: measuring how steep the ROC curve is

**Jože M. Rožanec**
Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
joze.rozanec@ijs.si

**Dunja Mladenić**
Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
dunja.mladenic@ijs.si

**Abstract.** *The Area Under the Receiver Operating Characteristic Curve (ROC AUC) is a widely used performance measure summarizing classifiers' discriminative power. Among its strengths is that it does not depend on threshold settings. Nevertheless, machine learning models scoring the same ROC AUC value can display a different behavior along the curve - a characteristic that may be relevant to model selection. This paper introduces WaAUROCC, an alternative measure that builds upon the ROC AUC and mitigates some of its well-known weaknesses. In particular, it allows for assessing how steep the curve is without visually inspecting it. It thus enables the identification of models where the highest recall is achieved with the lowest False Positive Rate. Furthermore, the metric allows for contrasting the models' performance against performance acceptance criteria, providing insights on whether such criteria are met and how much is outperformed. The approach followed when creating the WaAUROCC metric can be followed with other AUC metrics, such as the Precision-Recall AUC. We validate the usefulness of the proposed metric on three real-world datasets. In addition, we illustrate its usefulness with four synthetic scenarios.*

**Keywords.** Metrics, Classification, Operational Performance, MlOps

## 1 Introduction

**Background** The ROC AUC (Bradley, 1997) is commonly used to assess the discriminative power of machine learning classifiers (Halligan et al., 2015) and select the best ones Muschelli III (2020); Carrington et al. (2021) by comparing how the performance of a model varies across the True Positive Rate (TPR) and the False Positive Rate (FPR). The goal of the metric is to achieve a perfect TPR with $FPR = 0$. Among the metric strengths is that it is scale-invariant (measures how predictions are ranked, regardless of their absolute values) and provides an aggregate measure of performance across all possible classification thresholds Hernández-Orallo et al. (2012). ROC AUC is unaffected by dataset skew but may mask poor performance (Jeni et al., 2013; Cook and Ramadas, 2020). This could be because ROC AUC considers the four quadrants of a confusion matrix (in contrast to PR AUC, which does not consider true negatives) Sofaer et al. (2019). Multiple authors have documented detailed analysis behind such behavior (Fawcett and Flach, 2005; Webb and Ting, 2005; Cook and Ramadas, 2020).

**Motivation** In use cases such as defect detection or medical diagnosis and treatment, machine learning models are used along with a manual evaluation to produce an outcome Luzio et al. (2024); Klawonn et al. (2011); Moons et al. (1997). For example, when dealing with defect detection, we may prefer models that allow for business rules, such as automatically rejecting (or accepting) products for which we are highly confident that they are defective (or not). On top of that, we expect to have a reasonable subset of cases that can be manually inspected where defect occurrence is highly probable so that upon human inspection, a decision is made whether to accept or reject the transaction. We expect that models achieving a high Recall at a low FPR enable greater automation, given (i) high Recall means a high proportion of True Positives has been identified while keeping the number of False Negatives low and (ii) the proportion of False Positives identified is low w.r.t. the True Negatives. While the ROC AUC is a single scalar value and doesn't directly tell about specific thresholds, the ROC curve (from which the AUC is derived) can be used to compare models at particular points. Furthermore, the ROC AUC provides no information on whether a particular model achieves higher Recall at a lower FPR when compared to others. Therefore, a different metric must be considered to account for such differences and enable model selection that takes such model behavior into account.

**Limitations of the State of the Art** To account for the different weights between sensitivity and specificity, Hand (2009) proposed an alternative metric that considers the relative misclassification cost distribution based on the assumption that cost should dominate specificity in the choice of measure. Such assumption contrasts with machine learning models bound to a manual revision setting, where decisions are made based on how many cases can undergo manual revision. Halligan et al. (2015) followed a different rationale and proposed a metric evaluating change in sen-

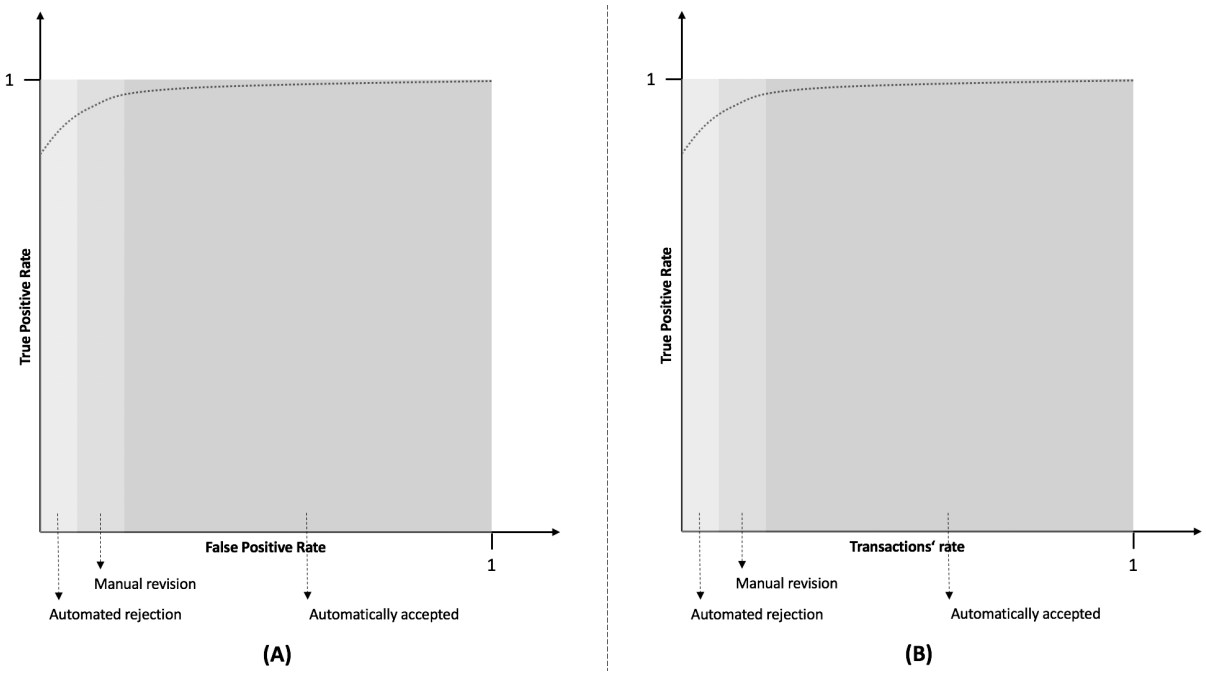



**Figure 1:** The images represent two plots that guide the evaluation of machine learning models for defect inspection: (A) ROC AUC, contrasting True Positive Rate against False Positive Rate, and (B) contrasting True Positive Rate against the ratio of items processed.

sitivity and specificity at clinically relevant thresholds. This approach is similar to the $Precision@Recall_K$ used by Luzio et al. (2024). Nevertheless, both fail to provide a global perspective of the models' performance (summarize in a score the behavior observed across all thresholds and observed in plots) - a key advantage of ROC AUC.

**Key insights** This paper introduces WaAUROCC [1] (/'wauɹɑk/) - *Wa*sserstein *A*rea *U*nder the *R*eceiver *O*perating *C*haracteristic *C*urve). It measures the Wasserstein distance between a given ROC AUC and the ideal case. By doing so, even when the ROC AUCs have the same value, WaAUROCC allows us to identify models with better operational performance (achieving high TPR and low FPR while considering the fewest samples possible).

# 2 WaAUROCC - how far is our model from ideal performance?

## 2.1 Comparing distributions

Two cases are known to us beforehand: ideal classifier performance (ROC AUC=1) and the random classifier case (ROC AUC=0.5). In addition, we can plot the ROC of the classifier we aim to evaluate. The ROC curve can be contrasted with the best and worst-case scenarios to obtain a score between zero and one, reflecting how close the performance of the classifier under consideration is to the best case. To do so, we follow a similar approach as presented in Rožanec et al. (2023): we can build bar plots that mirror the shape of the ROC curves for the three abovementioned cases and measure the effort required to turn one distribution into another. The bar plots are created considering a fixed number of equal-sized bins representing FPR values. When considering the effort required to turn one distribution into another, we look into the Optimal Transport mathematical problem, which aims to find the most efficient way to move mass between distributions. In this context, the Wasserstein distance measures the similarity between two distributions. For the WaAUROCC, we measure the distance between the actual ROC AUC curve and the optimal one.

## 2.2 Metric definition

We define the WaAUROCC metric for a multiclass scenario in Eq. 1. Classifiers close to an ideal performance will result in WaAUROCC values close to one, while models with little or no discriminative power will result in values close to zero. The cases that maximize TPR while minimizing FPR are rewarded.

## 2.3 WaAUROCC vs. Partial ROC plots

**ROC AUC interpretation** The ROC AUC has at least three interpretations: (i) the average Recall over all possible FPR values between zero and one, (ii) the

[1] A scikit-learn compatible implementation will be made available upon the paper's acceptance.

$$WaAUROCC = \sum_{i=1}^{n} \frac{1 - \frac{W_i(b_i, b_{best})}{W_i(b_{worst}, b_{best})}}{n} \quad (1)$$

**Equation 1:** $W_i(b_i, b_{ref})$ is the Wasserstein distance between the histogram $h_i$ (representing normalized ROC AUC values for class $i$) and the reference bar plots ($b_{best}$ and $b_{worst}$ for best and worst cases, respectively), and $n$ is the number of classes.

probability of correctly ranking a particular class based on the observed cases in the test set, and (iii) given a set of classes, how different the distributions of the predicted values are for them (relates to the Mann-Whitney statistic) Bamber (1975).

**ROC AUC drawbacks** Nevertheless, given Recall and FPR evaluate complementary aspects of the model performance, a classifier is frequently expected to have a high Recall and a low FPR. Furthermore, there are many cases (e.g., medical diagnosis) where only a limited range of FPR must be considered (either due to the values observed or some restrictions), making it unreasonable to use a metric that reflects the performance of higher FPR values Fahey et al. (1995); Scheidler et al. (1997); Carrington et al. (2022). Therefore, attention must be paid to the upper-left area of a ROC AUC diagram Yang et al. (2021).

**Overcoming ROC AUC limitations with pROCAUC** To address these ROC AUC flaws, the partial ROC AUC has been proposed, restricting the ROC AUC metric to a particular FPR range Walter (2005). Nevertheless, given the pAUC requires defining FPR ranges, comparisons between tests may lack homogeneity. Furthermore, the pROCAUC lacks the symmetry property of the ROCAUC, and the effective use of less information has been the source of concern about whether it results in a loss of statistical precision compared to the ROC AUC Obuchowski and McClish (1997); Ma et al. (2013). To mitigate such issues, several variations have been proposed. E.g., Ma et al. (2013) suggested a standardized pAUCROC by dividing the partial area of the curve by the partial area of the random case. While such a solution can estimate the quality of the predictor between 0.5 to 1 regardless of the FPR interval, the actual value of the AUC could increase or decrease when changing the interval size. A different approach was suggested by Carrington et al. (2020), who suggests not only performing the integration over the x-axis but a horizontal integration over the y-axis, thereby capturing the area at the top-left corner. Such an area provides insight into the improvement opportunity for a particular model compared to the best case. Based on the horizontal and vertical AUC, they define the concordant partial AUC as half the sum of the vertical partial area under the ROC curve pAUC and the horizontal partial area under the ROC curve.

**Toward a better metric: WaAUROCC** As described above, pROCAUC and the derivative metrics suffer from several shortcomings. Nevertheless, most of them could be overcome if a metric could (i) measure not only the ROC AUC but also how steep the curve is through FPR on a continuous range without defining particular segments and (ii) if a custom baseline curve could be specified so that a comparison could be drawn between the baseline and the performance of the model under evaluation. WaAUROCC achieves both by computing the Wasserstein distance between (i) the distributions that emulate the ROC AUC curve and the ideal case and (ii) the distributions that emulate the ROC AUC curve and the custom baseline curve. Through (i), the metric takes into account that lower FPRs are preferred and penalizes curves that achieve higher Recall at higher FPRs. Through (ii), domain knowledge and business requirements can be introduced, allowing us to measure how far the models' performance is satisfying a particular Recall and FPR levels.

## 2.4 WaAUROCC: measuring compliance with acceptance criteria

**Intuition** In Eq. 1, we show how a single metric can summarize the classifiers' ROC behavior, considering the AUC and how steep such a curve is over all of the False Positive Rate values. Similarly, acceptance criteria could be established considering AUC and the minimal expected curve steepness across False Positive Rate sections. Models whose WaAUROCC would be equal or greater than the WaAUROCC of such curve would be guaranteed to satisfy or supersede the acceptance criteria behavior.

**How is this useful?** In many real-world cases, knowing the models' overall performance is insufficient. More fine-grained insights are required to decide how such a model should be deployed and used in production environment settings. E.g., (i) what threshold interval should be considered for automated decision-making?, (ii) what threshold interval should be considered to seek complementary evaluation (e.g., with a model whose inference has a higher cost)?, (iii) what threshold interval should be considered for decision-making that requires human intervention? WaAUROCC enables drafting ROC AUC curves that reflect the expected performance across the whole FPR continuum. The WaAUROCC score of such curves establishes a baseline against which the models can be evaluated.

| Synthetic case | Predicted values | Ground truth |
|---|---|---|
| 1 | [0.1, 0.1, 0.1, 0.5, 0.7, 0.2, 0.2, 0.5, 0.55, 0.75] | [0, 0, 0, 0, 0, 1, 1, 1, 1, 1] |
| 2 | [0.25, 0.25, 0.45, 0.55, 0.50, 0.5, 0.55, 0.35, 0.55, 0.75] | [0, 0, 0, 0, 0, 1, 1, 1, 1, 1] |
| 3 | [0.1, 0.1, 0.1, 0.3, 0.1, 0.3, 0.1, 0.1, 0.7, 0.5] | [0, 0, 0, 0, 0, 1, 1, 1, 1, 1] |
| 4 | [0.1, 0.4, 0.2, 0.25, 0.55, 0.35, 0.25, 0.75, 0.8, 0.35] | [0, 0, 0, 0, 0, 1, 1, 1, 1, 1] |

**Table 1:** Synthetically crafted samples of predictions and ground truths. The examples result in the same ROC AUC but display different curve shapes.

# 3 Operational meaning of the proposed metric

Let's consider the ROC AUC curve, which can be discretized into bins (like a bar chart), and compare two different ROC AUC bar charts. The Wasserstein distance between both distributions measures the effort required to transform one distribution into another (e.g., a given ROC AUC into a perfect ROC AUC (the distribution for the latest one is always the same)). Such effort conveys information on how early the model can achieve a perfect True Positive Rate: the lower the distance, the smaller the difference (between our ROC AUC and a perfect model), and the closer we are to the perfect model. We considered an arbitrary cost (e.g., a cost of "one" per bin) to compute the effort required to move cases between bins. It must be noted that the ROC AUC is a monotonically non-decreasing function: the True Positive Rate values can remain the same (e.g., a perfect ROC curve) or increase when the False Positive Rate increases.

Computing the proposed metric is not equivalent to computing the ROC AUC: the synthetic examples presented in Section 6 show that curves with the same ROC AUC result in different values of the proposed metric, which can be associated with a particular behavior (lower metric values correspond to models achieving higher True Positive Rate at a lower False Positive Rate) as confirmed by visual analysis.

To compute the abovementioned metric, we propose normalizing the ROC AUC histograms to ensure the area under the curve is always the same. By doing this, we lose information regarding the original area under the curve. Nevertheless, we keep information regarding the curve shape - a key aspect of our metric. We can estimate the effort required to reshape the distribution to the ideal case by keeping the shape. The lower the required effort, the better the model: the model will achieve a higher True Positive Rate at a low False Positive Rate.

## 3.1 Putting it all together: WaAUROCC explained through the defect detection use case

When evaluating machine learning models for defect detection, we are interested in identifying the models that help us capture the highest percentage of True Positives while keeping False Positives low to avoid a bad user experience. This can be done by comparing the ROC AUC across models: models with a higher ROC AUC score display such characteristics. Nevertheless, there may be cases where models with an identical ROC AUC score display a different behavior along the curve. For such cases, we propose the WaAUROCC metric, where at an equal ROC AUC score, a higher WaAUROCC score will be assigned to models with a higher True Positive Rate at a lower False Positive Rate. Therefore, the WaAUROCC metric score could be used to select a subset of models. Once such selection is performed, we consider there are at least two plots of interest (see Fig. 1):

(A) ROC AUC helps us understand how the model performs along different thresholds when comparing the tradeoff between the True Positive Rate (Recall) and the False Positive Rate. This plot is of primary importance in defining the threshold for automated rejections: for a set of items, we may reject the ones we are most confident are defective, tolerating a small number of false positives. The amount of false positives to be tolerated is a business criterion.

(B) Recall vs. inspection rate, which helps us understand the number of items that fall within a particular Recall range. This plot is particularly relevant to defining the thresholds for cases undergoing manual revision. Usually, the manual revision capacity is fixed; therefore, we seek models that allow the widest possible Recall range for the same capacity.

Considering the same rationale as the one followed when creating the partial ROC plots and the fact that some business criteria may fix a False Positive Rate threshold, we may evaluate the models considering a partial ROC AUC and the corresponding WaAUROCC score. Furthermore, an analogous metric could be computed to identify the most promising models for the threshold range for manual revision.

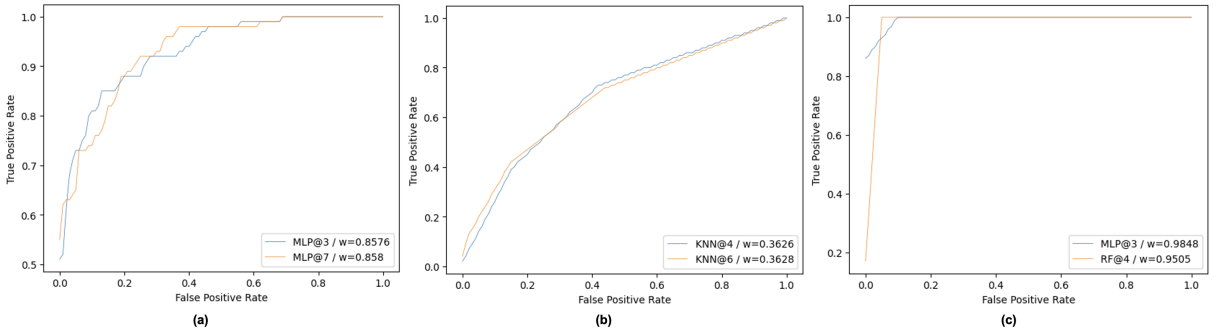



**Figure 2:** The plots correspond to the three cases reported in Table 2 where the measured ROC AUC values for the models are the same, but the resulting curves are different. In particular, the cases correspond to the following datasets: (a) DEFECTS, (b) SUPPORT2, and (c) CANCER.

| USE CASE | MODEL | FOLDS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| DEFECTS | KNN | 0,8794 | 0,9073 | 0,8592 | 0,9054 | 0,8249 | 0,8943 | 0,8554 | 0,8833 | 0,9096 | 0,8496 |
| | MLP | 0,9076 | 0,9304 | **0,9306** | 0,9568 | 0,8984 | 0,9118 | **0,9306** | 0,9364 | 0,9480 | 0,9039 |
| | RF | 0,8653 | 0,8836 | 0,8980 | 0,9395 | 0,8364 | 0,8949 | 0,8856 | 0,8406 | 0,9186 | 0,8626 |
| SUPPORT2 | KNN | 0,6732 | 0,6753 | 0,6445 | **0,6830** | 0,6291 | **0,6830** | 0,6697 | 0,6157 | 0,6523 | 0,6856 |
| | MLP | 0,7605 | 0,7718 | 0,7912 | 0,7704 | 0,7506 | 0,7581 | 0,7971 | 0,7755 | 0,7824 | 0,7908 |
| | RF | 0,7416 | 0,7538 | 0,7643 | 0,7631 | 0,7456 | 0,7418 | 0,7799 | 0,7696 | 0,7721 | 0,7598 |
| CANCER | KNN | 0,9414 | 0,9750 | 0,9930 | 0,9792 | 0,9554 | 0,9784 | 0,9992 | 0,9836 | 0,9864 | 0,9465 |
| | MLP | 0,9801 | 0,9850 | **0,9937** | 0,9973 | 0,9622 | 0,9972 | 1,0000 | 0,9943 | 0,9946 | 0,9924 |
| | RF | 0,9587 | 0,9866 | 0,9854 | **0,9937** | 0,9527 | 0,9914 | 0,9973 | 0,9972 | 0,9924 | 0,9960 |

**Table 2:** ROC AUC scores obtained for the use cases presented by machine learning model and fold. Bolded values correspond to cases where the models resulted in the same ROC AUC score for that particular dataset.

| USE CASE | MODEL | FOLDS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| DEFECTS | KNN | 0,7535 | 0,8111 | 0,7135 | 0,8071 | 0,6451 | 0,7826 | 0,7059 | 0,7602 | 0,8133 | 0,6945 |
| | MLP | 0,8103 | 0,8564 | **0,8576** | 0,9101 | 0,7925 | 0,8188 | **0,8580** | 0,8683 | 0,8931 | 0,8038 |
| | RF | 0,7265 | 0,7640 | 0,7923 | 0,8756 | 0,6693 | 0,7855 | 0,7671 | 0,6770 | 0,8345 | 0,7212 |
| SUPPORT2 | KNN | 0,3432 | 0,3473 | 0,2865 | **0,3626** | 0,2558 | **0,3628** | 0,3362 | 0,2295 | 0,3018 | 0,3677 |
| | MLP | 0,5162 | 0,5390 | 0,5770 | 0,5358 | 0,4966 | 0,5113 | 0,5883 | 0,5459 | 0,5594 | 0,5766 |
| | RF | 0,4788 | 0,5032 | 0,5236 | 0,5214 | 0,4867 | 0,4790 | 0,5547 | 0,5341 | 0,5392 | 0,5150 |
| CANCER | KNN | 0,8749 | 0,9420 | 0,9848 | 0,9505 | 0,9032 | 0,9489 | 0,9982 | 0,9671 | 0,9711 | 0,8850 |
| | MLP | 0,9572 | 0,9665 | **0,9867** | 0,9941 | 0,9202 | 0,9941 | 1,0000 | 0,9881 | 0,9887 | 0,9832 |
| | RF | 0,9083 | 0,9713 | 0,9699 | **0,9861** | 0,9036 | 0,9828 | 0,9941 | 0,9941 | 0,9832 | 0,9911 |

**Table 3:** WaAUROCC scores obtained for the use cases, presented by machine learning model and fold. Bolded values correspond to cases where the models resulted in the same ROC AUC score for that particular dataset.

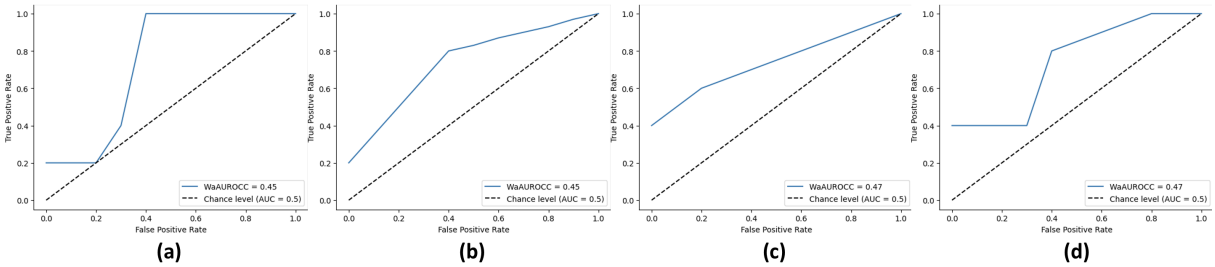



**Figure 3:** The images correspond to four synthetic scenarios with the same ROC AUC (0.74) but different WaAUROCC. The plots are sorted in ascending order based on WaAUROCC values. The dotted line indicates the chance level.

# 4 Methodology and experimental setup

## 4.1 Real-world datasets

We considered ten-fold stratified cross-validation, devoting nine folds to train K-Nearest Neighbors (KNN),

Multi-Layer Perceptron (MLP), and Random Forest (RF) classifiers, and one fold to test them. We experimented on three datasets: (i) a dataset introduced by Connors et al. (1995), predicting death in hospital (SUPPORT2), (ii) the breast cancer dataset Street et al. (1993) (CANCER), and (iii) a real-world dataset of printed logos on shaver images provided by an industrial partner within project cooperation to research automated defect inspection (DEFECTS). Images were converted into feature vectors by leveraging ResNet-18 embeddings. Non-numerical features were removed, and features selection considering top K ranking features based on mutual information and a maximum of $\sqrt{N}$, with $N$ equal to the number of instances in the training set. We executed the abovementioned setup multiple times considering different seed numbers, until getting cases where the models' performance would result in the same ROC AUC values. We then focused on those cases and analyzed whether did the WaAUROCC provide a better intuition behind them.

### 4.2 Synthetic samples

In addition to the real-world datasets, we crafted four synthetic samples of predictions and expected ground truth to showcase different scenarios in which the ROC AUC would result in the same value despite a different curve shape, but the WaAUROCC would detect such discrepancies. We showcase the examples in Table 1.

## 5 Results

### 5.1 Real-world datasets

We detail the results obtained in Table 2 and Table 3. The tables present the results obtained by measuring ROC AUC and WaAUROCC across datasets and machine learning models over ten folds of a cross-validation setting. When performing the experiments, we could find one case per dataset where the machine learning models resulted in the same ROC AUC score, but the curves were noticeably different. For ease of analysis, those cases are bolded in the abovementioned tables and we plot their ROC curves in Fig. 2.

**Same ROC AUC, different operational performance.** For the DEFECTS dataset, we found that the MLP model reported the same ROC AUC value at third and seventh fold, but the WaAUROCC scores showed the model tested at the seventh fold would achieve a higher Recall at a lower FPR. This is validated with the plot at Fig. 2 (a). A similar case is observed for the SUPPORT2 dataset but for the KNN model: while the ROC AUC score for the fourth and sixth folds is the same, the WaAUROCC score indicates the model tested at the sixth fold achieves a higher TPR for the same FPR at least until a FPR of 0.2 (see Fig. 2 (b)). Finally, for the CANCER dataset, the same ROC AUC is reported by two different kinds of models at different test folds: MLP at the third fold shows the same performance as the RF model at the fourth fold. Nevertheless, the WaAUROCC score shows a preference for the MLP model tested at the third fold - which is consistent with the ROC curves in 2 (c).

### 5.2 Synthetic samples

We detail the results obtained for the synthetic samples in Fig. 3. By analyzing the plots, we observe that (a) has a low initial TPR but gradually improves and reaches 0.8 TPR at 0.4 FPR. In contrast, (b) and (c) have a higher initial TPR. While both reach a 0.8 TPR at 0.4 FPR, (b) has a flat 0.4 TPR between 0 and 0.4 FPR, while (c) shows a mild increasing slope in that same range. Among the three models, (c) is preferable, given it achieves the highest TPR at the lowest FPR.

**Same ROC AUC, different operational performance.** Four synthetic scenarios were considered, ensuring the same ROC AUC values but different operational performances (see Fig. 3). Considering overall performance, (c) and (d) are preferable, given that they achieve the highest TPR at the lowest FPR when considering the whole FPR range.

**Partial ROC plots.** Nevertheless, (c) and (d) display a different operational performance if considering a specific region of the ROC plot. If we constrain the metric computation to the FPR $[0 - 0.2]$ range, the WaAUROCC values change ((a)=0.11, (b)=0.28, (c)=0.44, and (d)=0.33), showing that model (c) is preferred to the rest and that model (b) is preferred to (a). Such insights could be particularly relevant, e.g., if interested in automating decisions for ranges with a high TPR and FPR below a certain threshold.

## 6 Discussion and conclusion

**Contribution** The ROC AUC can sometimes mask critical performance differences between models, especially regarding Recall and False Positive Rates. To address this issue, we propose the WaAUROCC metric. By framing the ideal and actual ROC AUC curves as two distributions, we can compute the Wasserstein distance between the ideal scenario and the actual one. Doing so provides insights into the ROC curve steepness (higher Recall at a lower FPR), enabling a better model selection based on operational performance criteria. We showcase the usefulness of the proposed metric by evaluating machine learning models on three real-world datasets. Furthermore, we also crafted four synthetic scenarios where each ROC curve has a distinct shape but results in the same ROC AUC scores. In all cases, the WaAUROCC metric correctly identifies which curves are steeper at the beginning of the curve.

Finally, we show the WaAUROCC metric could also be applied for partial ROC curves, leading to scores that are analogous to pROCAUC while correctly preferring curves with higher initial steepness.

**Limitations** The proposed metric has certain limitations, among which we should mention the fact that (i) the ROC plot is discretized into bins and (ii) an arbitrary cost is established to compute the effort required to move cases between bins, under the assumption that the same cost is incurred to move an instance to a better score as to a lower score. Nevertheless, it must be noticed that strongly misclassified instances will finally incur a high cost, given they must be moved a longer way to guarantee a perfect Recall without False Positives. Another limitation of this work is the assumption that the classifiers operate within the convex hull of their ROC curve. While it is possible to find nonlinear combinations of classifiers producing ROC curves that exceed their particular convex hulls Scott et al. (1998), we consider such ensembles to be a different model on their own.

**Future work** We plan to contribute the implementation of this metric to the scikit-learn repository. Our future work will explore means to mitigate the limitations of the proposed metric, such as the costs of the misclassified cases being moved through bins.

## Acknowledgments

This work was supported by the Slovenian Research Agency, the European Union's Horizon 2020 project STAR under grant agreement H2020-956573, and the Horizon Europe project HumAIne, Grant Agreement No. 101120218. We want to express our gratitude to all of the project partners for their contributions and collaboration.

## References

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of mathematical psychology*, 12(4):387–415.

Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145 – 1159.

Carrington, A. M., Fieguth, P. W., Qazi, H., Holzinger, A., Chen, H. H., Mayr, F., and Manuel, D. G. (2020). A new concordant partial auc and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. *BMC medical informatics and decision making*, 20:1–12.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., et al. (2022). Deep roc analysis and auc as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):329–341.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., McInnes, M., Magwood, O., et al. (2021). Deep roc analysis and auc as balanced average accuracy to improve model selection, understanding and interpretation. *arXiv preprint arXiv:2103.11357*.

Connors, A. F., Dawson, N. V., Desbiens, N. A., Fulkerson, W. J., Goldman, L., Knaus, W. A., Lynn, J., Oye, R. K., Bergner, M., Damiano, A., et al. (1995). A controlled trial to improve care for seriously iii hospitalized patients: The study to understand prognoses and preferences for outcomes and risks of treatments (support). *Jama*, 274(20):1591–1598.

Cook, J. and Ramadas, V. (2020). When to consult precision-recall curves. *The Stata Journal*, 20(1):131–148.

Fahey, M. T., Irwig, L., and Macaskill, P. (1995). Meta-analysis of pap test accuracy. *American journal of epidemiology*, 141(7):680–689.

Fawcett, T. and Flach, P. A. (2005). A response to webb and ting's on the application of roc analysis to predict classification performance under varying class distributions. *Machine Learning*, 58:33–38.

Halligan, S., Altman, D. G., and Mallett, S. (2015). Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. *European radiology*, 25:932–939.

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the roc curve. *Machine learning*, 77(1):103–123.

Hernández-Orallo, J., Flach, P., and Ferri Ramírez, C. (2012). A unified view of performance metrics: Translating threshold choice into expected classification loss. *Journal of Machine Learning Research*, 13:2813–2869.

Jeni, L. A., Cohn, J. F., and De La Torre, F. (2013). Facing imbalanced data–recommendations for the use of performance metrics. In *2013 Humaine association conference on affective computing and intelligent interaction*, pages 245–251. IEEE.

Klawonn, F., Höppner, F., and May, S. (2011). An alternative to roc and auc analysis of classifiers. In

*International Symposium on Intelligent Data Analysis*, pages 210–221. Springer.

Luzio, E., Ponti, M. A., Arevalo, C. R., and Argerich, L. (2024). Decoupling decision-making in fraud prevention through classifier calibration for business logic action. *arXiv preprint arXiv:2401.05240*.

Ma, H., Bandos, A. I., Rockette, H. E., and Gur, D. (2013). On use of partial area under the roc curve for evaluation of diagnostic performance. *Statistics in medicine*, 32(20):3449–3458.

Moons, K. G., Stijnen, T., Michel, B. C., Büller, H. R., Van Es, G.-A., Grobbee, D. E., and Habbema, J. D. F. (1997). Application of treatment thresholds to diagnostic-test evaluation: an alternative to the comparison of areas under receiver operating characteristic curves. *Medical Decision Making*, 17(4):447–454.

Muschelli III, J. (2020). Roc and auc with a binary predictor: a potentially misleading metric. *Journal of classification*, 37(3):696–708.

Obuchowski, N. A. and McClish, D. K. (1997). Sample size determination for diagnostic accuracy studies involving binormal roc curve indices. *Statistics in medicine*, 16(13):1529–1542.

Rožanec, J. M., Bizjak, L., Trajkova, E., Zajec, P., Keizer, J., Fortuna, B., and Mladenić, D. (2023). Active learning and novel model calibration measurements for automated visual inspection in manufacturing. *Journal of Intelligent Manufacturing*, pages 1–22.

Scheidler, J., Hricak, H., Kyle, K. Y., Subak, L., and Segal, M. R. (1997). Radiological evaluation of lymph node metastases in patients with cervical cancer: a meta-analysis. *Jama*, 278(13):1096–1101.

Scott, M. J., Niranjan, M., and Prager, R. W. (1998). Realisable classifiers: Improving operating performance on variable cost problems. In *BMVC*, pages 1–10. Citeseer.

Sofaer, H. R., Hoeting, J. A., and Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10(4):565–577.

Street, W. N., Wolberg, W. H., and Mangasarian, O. L. (1993). Nuclear feature extraction for breast tumor diagnosis. In *Biomedical image processing and biomedical visualization*, volume 1905, pages 861–870. SPIE.

Walter, S. D. (2005). The partial area under the summary roc curve. *Statistics in medicine*, 24(13):2025–2040.

Webb, G. I. and Ting, K. M. (2005). On the application of roc analysis to predict classification performance under varying class distributions. *Machine learning*, 58:25–32.

Yang, Z., Xu, Q., Bao, S., He, Y., Cao, X., and Huang, Q. (2021). When all we need is a piece of the pie: A generic framework for optimizing two-way partial auc. In *International Conference on Machine Learning*, pages 11820–11829. PMLR.

# Beyond Words: Sentiment Analysis of Croatian Language Attitudes

**Barbara Kovačić**
Ludwig-Maximilians-University
Center for Information and Language Processing
Oettingenstrasse 67, 80538 Munich, Germany
Barbara.Kovacic@campus.lmu.de

**Anja Brnjaković, Gaurish Thakkar**
University of Zagreb
Department of Information and Communication Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
{abrnjako, gthakkar}@m.ffzg.hr

**Abstract.** *Sociolinguistic research can benefit heavily from natural language processing (NLP) tools. Especially language attitude research can be facilitated by using sentiment analysis. By using existing tools for Bosnian-Croatian-Montenegrin-Serbian (BCMS), we try to develop a pipeline that can be used to collect sentiment towards specific attitudes by the example of the Croatian variety, henceforth referred to as the Croatian language. Therefore, we add sentiment labels at the sentence level to a dataset containing comments on language-related newspaper articles. Then we look at the automation of sentiment annotation for this resource, training different transformer models on that task by using the Machine Learning Toolkit MaChAmp. The experiments demonstrate varied performance across transformer models, with the BERTić model achieving the highest accuracy (0.71) in sentiment analysis of Croatian newspaper comments.*

**Keywords.** Sentiment analysis, Croatian language, Language attitudes

## 1 Introduction

The Croatian language is distinguished by its three primary dialects: Čakavian, Kajkavian, and Štokavian. Each dialect uses a different term for "what", specifically "ča", "kaj," and "što". The Štokavian dialect underpins the Standard Croatian language, which experienced a revival post-Yugoslavian war in the 1990s. In the aftermath, Croatia introduced a broad language policy (Mønnesland, 1997) aimed at maximising differentiation from the Serbian Standard language. On January 26, 2024, for the first time in history, the Croatian government legally regulated the use of the official and public use of the Croatian language. As announced on the official website of the Croatian government (Vlada Republike Hrvatske/Hina, 2024), the Štokavian dialect has been chosen as the model for the standard. The law prescribes the basic rules of the Croatian standard language in official and public use. The public use of the Croatian language is any oral or written language communication in a space intended for the public or a wider audience, including electronic ones. In media content, in addition to the use of the Croatian language, the Croatian dialect idiom can be used, especially when there are artistic reasons for this.

One unintended outcome of the Croatian language policy is the reduction in the number of Čakavian and Kajkavian dialect speakers. The changing attitudes of Croatian speakers towards the standard language, who increasingly believe it to be more correct than the dialects, may help to explain this phenomenon (Kalogjera, 2001). The concept of "language attitudes" refers to how individuals react to different language varieties and their users, influenced by psychological and sociocultural factors. These attitudes are a blend of cognitive aspects involving beliefs and stereotypes, affective aspects related to evaluations, and behavioural aspects that reflect observable actions and responses (Garrett, 2006).

Several approaches have been developed to investigate language attitudes, such as the societal treatment approach. Garrett (2006) defines this approach as an extensive research category that examines public domain sources, including government discourse, media, and literature. This method focusses on the sociocultural and political context of attitudes, aiming to offer a thorough understanding of the relationship between societal discourses, political events, and individual attitudes. In previous research, computational methods have been applied to study societal treatment, such as Durham's analysis of tweets about the Welsh

English Accent (Durham, 2016). While data collection was automated, the categorisation and analysis were done manually, making the process both time-consuming and resource-intensive.

Sentiment analysis presents a potential solution to this problem by reducing both the time consumption and resource intensity involved in manual categorisation and analysis by facilitating high-level sentiment classification of documents, sentences, or words (Pang et al., 2008). Therefore, we first annotate sentences containing language-related reader comments from different news websites (Bogetić and Batanović, 2020). By doing so, the dataset can be further analysed for language attitudes towards the Čakavian and Kajkavian dialect and their speakers. Finally, we test automatic annotation of sentiment with the current state-of-the-Art transformer models for Croatian, aiming to contribute the infrastructure for further automation of language attitude research.

## 2 Related Work

Durham is the first to combine sentiment analysis and language attitude research by the example of the Welsh English accent (Durham, 2016). His study aimed to analyse attitudes towards the Welsh accent expressed on Twitter over a nine-month period from September 2012 to May 2013. Data collection involved searching for tweets containing the terms “Welsh" and “accent", resulting in 87,165 extracted tweets. To manage the large number of tweets, only four days each month were selected for analysis. On those days, tweets were coded based on whether their attitudes were mostly positive, negative, or neutral. Categories included tweets showing love or appreciation, negativity, ambiguous attitudes, performance elements, and those discussing British dialects broadly. The analysis excluded retweets and replies to maintain focus. Monthly tweet counts varied between 9,000 to nearly 12,000, with peaks coinciding with events like the premiere of “The Valleys," a reality TV show featuring Welsh accents. A secondary coding phase refined categories based on thematic frequency, aligning findings with established frameworks for discussing attitudes towards accents.

In addition to that, sentiment annotation has been conducted on various BCMS linguistic resources in the past, resulting in a range of BCMS datasets available for sentiment analysis. The Croatian film review dataset, focusing on Croatian, offers approximately 10,000 sentences from film reviews, annotated at the sentence level and sourced from Thakkar et al. (2023). Another Croatian dataset, the dataset for stance and sentiment analysis from user comments, comprises 54 articles and 904 user comments from the newspaper “24 sata”, annotated at the comment level, published by Bošnjak and Karan (2019). The corpora of Thakkar et al. (2023) and Bošnjak and Karan (2019) are particularly relevant to our task because they include texts published by private users on the internet. Given that much of the input on language attitudes comes from social media and news site comment sections, these corpora can be valuable resources. We can use them either as raw data for analysing language attitudes or as a basis for extending our annotated corpus through data synthesis methods.

Furthermore, different technical approaches have been tested on improving existing methods. Babić et al. (2021) developed a framework for characterisation of COVID-19-related tweets in Croatian. Their research methodology involves several key steps focused on enhancing sentiment analysis using Croatian language data related to COVID-19. Initially, the CroCoV-cseBERT model was trained to embed COVID-19 tweets in Croatian as vectors, compared against a fastText baseline model. Four machine learning (ML) models - naive Bayes, random forest, support vector machine (SVM), and multilayer perceptron - were then trained for sentiment classification, using both embedding models. Following this, clustering techniques were applied to identify themes in the tweets, analysing sentiment trends and retweeting behaviours across different pandemic waves and clusters. The methodology developed by Babić et al. (2021) offers valuable insights into sentiment analysis within specific topics, like COVID-19, by employing specialised models and clustering techniques. This approach can be adapted to analyse language attitudes in Croatian texts by identifying dominant themes and sentiment trends in discussions about language policies, regional dialects, or linguistic debates. The ability to detect and categorise sentiment trends could reveal underlying language attitudes across different communities.

Another milestone of sentiment analysis for Croatian is the multitask learning approach, developed by Thakkar et al. (2021). This research employs cross-lingual knowledge transfer between Slovenian and Croatian datasets for sentiment analysis, leveraging their high mutual intelligibility among South Slavic languages. The study hypothesises Slovenian as an effective hub language for this purpose, given both datasets’ shared text type (news articles). Using a shared encoder based on the CroSloEngual model, trained on Slovenian, Croatian, and English texts, the methodology includes a multitask learning setup across document-level, paragraph-level, and sentence-level tasks. The experimental setup encompasses five scenarios: zero-shot learning on Slovenian data for Croatian sentiment analysis at the document-level; multitask learning on Slovenian data across all three levels independently of Croatian data; traditional fine-tuning on Croatian data alone; simultaneous training on both Croatian and Slovenian data at the document-level; and training on both datasets but focusing only on document-level classification. The multitask learning approach by Thakkar et al. (2021) is highly relevant

for our research as it leverages cross-lingual knowledge transfer, particularly between Slovenian and Croatian, to improve sentiment analysis. This method could help us better understand language attitudes in multilingual contexts or regions where multiple South Slavic languages are spoken. Moreover, both use cases leverage the CroSloEngual BERT model, which has demonstrated strong performance, establishing it as the state-of-the-art model for Croatian sentiment analysis.

# 3 Dataset

This section details the dataset used for our sentiment analysis. We explain the steps taken to prepare and process the data, including the extraction of metadata and the selection of sentences for annotation. The section also describes our annotation process, where annotators classified sentiment and identified targets. Finally, we discuss the inter-annotator agreement, noting the challenges faced, and the measures implemented to ensure the dataset's reliability.

## 3.1 Original Dataset

We require a source containing natural language from public discussions about language itself, which is why we utilise an existing dataset provided by Bogetić and Batanović (2020) called MetaLangNEWS-COMMENTS-Hr[1]. The corpus is a collection of reader comments from 738 news articles, consisting of 21,533 texts and 823,459 tokens, published in the five-year period of January 1, 2015 - January 1, 2020. It was designed to facilitate research on metalanguage, as well as to provide insight into linguistic ideologies, language policy and planning and contemporary debates on language defining, naming, and standardisation. The corpus is available in .txt, .xml and CoNLL-U format. The mentioned formats were used for data preparation, extraction and processing.

## 3.2 Data Preparation, Extraction and Preprocessing.

The following is a brief description of our data collection and annotation process. We begin by extracting data from .xml files, where each file represents a news article. From these files, we extract metadata such as document ID, source name, article title, author, and text. Next, we process the corresponding CoNLL-U files, which contain the article comments split into sentences. We focus exclusively on sentences from the comments, as sentiment in news articles and titles tends to be neutral, while sentiment within reader comments can vary significantly depending on the topic. We then merge the metadata with the sentences by matching them to the document ID. This approach allows us to filter out news articles without comments and enhance data analysis possibilities, such as analysing sentiment by source or author. As a result, we generate a .csv file containing 57,257 sentences from 465 articles, containing the following metadata:

- global-id: Global document ID, in the form 'source-id'-'local-id'
- source-name: Full name of the source website
- article-title: Article title in its original script
- article-time: Date on which the article was published
- article-author: Name or initials of the article author
- article-text: Main text of the article in its original script
- sentence-id: ID of the sentence in the CONLL-u file
- sentence: target sentence

Due to time limitations, we pick the first 3,000 sentences to be annotated.

## 3.3 Annotation Environment

For our annotation environment, we use the Google Drive ecosystem. We create a folder on Google Drive for this purpose, where each annotator has their own Google spreadsheet. This spreadsheet includes the dataset from the aforementioned .csv file and features additional columns for "sentiment" and "target". In the "sentiment" column, we implement a drop-down menu allowing annotators to select the tags "positive", "neutral", "negative" or "mixed", the latter being used for sentences exhibiting both positive and negative sentiments, such as those with nested clauses. This setup aims to minimise spelling errors in the annotations.

## 3.4 Sentiment Annotation

For manual sentiment annotation, three annotators are involved in the process. The annotators are graduate students of information and communication sciences who are native speakers of the Croatian language and the Štokavian dialect. They have no prior experience in data annotation. Before annotation, annotation guidelines, describing each label with examples, were created. the guidelines were defined following the works from Mohammad (2016) and (Kralj Novak et al., 2015).

The annotators' task is to manually determine the polarity of each sentence and identify the aspect. This task is defined as a sentence-level sentiment analysis. As shown in the example below, annotators classify polarity into one of four categories: "positive", "neutral", "negative" and "mixed". For each sentence, they also identify the target object (aspect) - such as a person, place, subject, or other entity-toward which the opinion is directed.

[1] http://hdl.handle.net/11356/1370

- **Hr**: “ne pravite se ludi - ćirilica je u ovom slučaju negativan simbol koljača”
- **En**: “don’t pretend to be crazy - the Cyrillic alphabet in this case is a negative symbol of the butcher”
- **Label**: negative
- **Target**: Cyrillic alphabet

Each sentence is annotated three times by three different annotators. During the annotation phase, annotators must report any uncertainties and document them in the Notes column of their assigned spreadsheet. Difficult cases are discussed and incorporated into the annotation guidelines.

We also create a Google Sheets Dashboard that aggregates values from each annotation spreadsheet and colour-codes discrepancies in the labels: red for total disagreement, yellow for partial agreement, and green for full agreement. This overview facilitates discussions on cases where annotations differ and helps determine the final annotation. Once each annotator has completed 3,000 sentences in their spreadsheet, the entire dataset is re-annotated to produce the final annotated dataset. During the annotation process, it is noted that some sentences were poorly split, so these examples are removed from the dataset, resulting in a total of 2,938 annotated sentences. After the annotation and cleaning, The dataset includes 18 authors, 26 articles, 25 dates, and 5 sources. The final data characteristics for the first task is shown in Table 1.

**Table 1: Distribution of Sentiment Labels in the Dataset.** The table presents the distribution of sentiment labels within the final dataset, showcasing the number of instances for each sentiment category: Positive, Neutral, Negative, and Mixed. The total number of instances in the dataset is 2,938.

| Label | Number of instances |
|---|---|
| Positive | 189 |
| Neutral | 939 |
| Negative | 1662 |
| Mixed | 148 |
| Total | 2938 |

### 3.5 Inter-Annotator-Agreement

For calculating inter-annotator agreement, we use Fleiss’ Kappa, a statistical measure developed by Joseph Fleiss (1971). Fleiss’ Kappa measures agreement among multiple annotators by comparing their classifications against chance expectations. It involves calculating observed and expected agreement rates and normalising them. The Kappa values range from -1 (complete disagreement) to 1 (perfect agreement), with 0 indicating chance-level agreement. Values from 0.01 to 1.00 indicate increasing levels of agreement: slight (0.01-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80), and almost perfect (0.81-1.00). Using the Google Sheets Dashboard, Fleiss’ Kappa is calculated as 0.47, indicating a moderate level of agreement (0.41-0.60) among annotators. The moderate Kappa suggests that while there is a reasonable level of consensus among annotators, some variability remains. This variability can impact the reliability of the final results, indicating that further refinement or additional training of the annotators may be needed to improve consistency. It may also suggest that the dataset is ambiguous, which can impact the accuracy and performance of the language model trained on it.

## 4 Methodology

For our experiments, we use the machine learning toolkit MaChAmp (Van Der Goot et al., 2021). MaChAmp uses a pre-trained contextualised model as its initial encoder, fine-tuning the layers with an inverse square root learning rate decay and linear warm-up. Each task has a specific decoder trained for that task. For text classification, MaChAmp predicts labels using the embedding of the first token, inserting special tokens for tasks involving multiple sentences to delineate sentence boundaries. By default, the MaChAmp toolkit uses multilingual BERT (mBERT) (Devlin et al., 2019) as embeddings. mBERT supports over 100 languages with the largest Wikipedia, including Croatian and other varieties of BCMS. But the toolkit also allows training on other models.

Another multilingual transformer model including Croatian is CroSlo-Engual BERT (Ulčar and Robnik-Šikonja, 2020), hereinafter abbreviated as cseBERT. This model is trained on Croatian, Slovenian, and English corpora to support less-resourced languages. The training involved 5.9 billion tokens from monolingual corpora, including news articles and general web crawls. These corpora were preprocessed and deduplicated, ensuring high-quality data. Word piece vocabularies were created with the bert-vocab-builder tool, resulting in 49,601 tokens for cseBERT. The model employs the bert-base architecture, featuring a 12-layer bidirectional transformer encoder with 768 hidden units and 110 million parameters. Whole word masking was used for the masked language model task, and case information was preserved. This model is publicly available and provides a significant resource for enhancing NLP applications in Croatian, Slovenian, and English. Initially developed for part-of-speech tagging, named entity recognition (NER), and dependency parsing, cseBERT has also been applied to sentiment analysis projects like analysing COVID-19 communication on Twitter (Babić et al., 2021) and newspaper articles (Thakkar et al., 2021).

To be able to compare multilingual with monolingual approaches, we finally use MaChAmp in combination with BERTić (Ljubešić and Lauc, 2021). BERTić was trained over 8 billion tokens in BCMS. The model uses several existing datasets: the hrWaC

corpus (Croatian top-level domain), the srWaC corpus (Serbian top-level domain), the bsWaC corpus (Bosnian top-level domain), the cnrWaC corpus (Montenegrin top-level domain), and the Riznica corpus (Croatian literary works and newspapers). Additionally, new web crawls for Bosnian, Croatian, and Serbian domains were performed and labelled as CLASSLA web corpora, which were deduplicated to remove identical sentences. It also incorporates the cc100 corpora from CommonCrawl data, further deduplicated to remove overlaps with the WaC and CLASSLA corpora. The final dataset comprises 8.3 billion words. Text preprocessing preserved all Unicode characters. Common tasks executed on BERTić are part-of-speech tagging, NER, geolocation prediction, commonsense causal reasoning and hate speech detection.

Given that our dataset is rather small, we want to enhance it and also explore whether incorporating different styles of text can improve the performance of the language model. Therefore, we use the Croatian Film Review Dataset (Cro-FiReDa) (Thakkar et al., 2023)[2]. It is the first sentiment movie review dataset for the Croatian language. It includes reviews from a Croatian movie review website, covering genres such as adventure, series, and sci-fi. Each review consists of the review text, a summary, an overall assessment score, and the review date. The dataset contains 216 adventure-related reviews, 114 sci-fi reviews, and 76 series reviews. The sentiment annotation task was conducted at the sentence level, with each sentence labelled as "negative", "neutral", "positive", "mixed", or "other/sarcasm". Reviews were segmented into sentences and annotated by undergraduate students in linguistics and informatics. The annotation was facilitated by a deep-learning sentiment classification model trained on the SentiNews dataset, which includes Croatian and Slovenian news articles. The final dataset comprises 10,464 sentences, with 59 percent labelled as neutral. The dataset, annotation guidelines, trained models, and associated code are to be made publicly available. As MaChAmp requires the same format for all datasets used, we add a column called "sentence_i" to Cro-FiReDa. Additionally, we remove all sentences labelled as "other/sarcasm" because detecting sarcasm can be particularly challenging and may introduce significant noise into the data, complicating the sentiment analysis and potentially affecting the model's accuracy.

Moreover, we split our annotated dataset, hence after abbreviated as "LA" into a train (LA-train), a dev (LA-dev) and test (LA-test) dataset by using the ratio 70-10-20. By that, we get the division of the datasets, shown in Table 2.

[2] https://shorturl.at/Pvnr3

**Table 2: Dataset train-test distribution.** The table outlines the splitting of Cro-FiReDa and our annotated dataset LA, into training, development, and test sets.

| | Cro-FiReDa | LA |
|---|---|---|
| Train | 7480 sentences | 2056 sentences |
| Dev | 832 sentences | 291 sentences |
| Test | 2079 sentences | 591 sentences |

# 5 Experimental Setup

In total, we run 12 experiments, using four systems and three language models. Our reference system (REF) uses the complete Cro-FiReDa (Thakkar et al., 2023) dataset. By using it, we want to create reference values for the best case scenario of how well our dataset and language model can perform. Then, we have two baseline systems, called BASE-1 and BASE-2. In contrast to the reference system, we use the baseline system to see, how well the language model performs when trained on Cro-FiReDa and tested on our annotated dataset. While BASE-1 is trained and further developed with the Cro-Fi-ReDA datasets, the BASE-2 system uses the LA-dataset for further development. Our final system is the COM system, which is trained, developed and tested on our LA-dataset. Each system is then used once with mBERT, cseBERT and BERTić.

# 6 Results

Table 3 presents the results of our experiments. Here we can see that for REF, the best performance is achieved with the BERTić model, achieving the highest accuracy (0.80) and F1-score (0.77) among the three models. This suggests that BERTić is particularly effective at capturing the nuances required for the REF task, likely due to its specialised training on Croatian social media data. However, the relatively high F1-score indicates that while accuracy is high, there remains some challenge in correctly identifying all relevant instances, possibly due to nuances in sentiment and expression typical of newspaper comments compared to formal language.

The BASE-1 system performs better with cseBERT (0.57) and BERTić (0.56), showing significantly lower performance with mBERT (0.44). This variation highlights the sensitivity of the system's performance to the choice of pre-trained model, where domain-specific training appears to offer advantages over more general models like mBERT. The differences in source sentences between the Cro-FireDa dataset used by BASE-1 and the LA dataset may contribute to these disparities, as the linguistic and sentiment characteristics vary widely between these sources. Moreover, the substantial dataset imbalance in the LA dataset, with a predominance of negative and neutral labelled sentences compared to the movie review dataset, may skew performance metrics, particularly affecting the F1-score.

The BASE-2 system shows a similar trend to BASE-1, with an accuracy of 0.59 with cseBERT and 0.56 with BERTić. This consistency across different systems further underscores the influence of dataset origin on performance outcomes. The specific challenges posed by social media text in capturing sentiment accurately are evident here, affecting both accuracy and F1-score. The dataset imbalance also plays a role, as models trained on datasets with uneven class distributions may struggle to generalise well to new data, particularly in identifying less represented classes like mixed sentiment.

The COM system shows strong performance across all models, with the best accuracy using BERTić (0.71). However, its F1-score is highest with cseBERT (0.59) and notably lower with BERTić (0.41). This discrepancy between accuracy and F1-score suggests that while the model predicts many labels correctly, it may struggle with the more nuanced task of correctly identifying all instances of sentiment, particularly mixed sentiments, which can be expressed subtly in newspaper comments.

In summary, these results highlight several factors influencing performance, including the domain of training data (movie reviews vs. newspaper comments), the specificity of the pre-trained models used, and the inherent challenges of sentiment analysis in informal, user-generated content. The differences in annotated datasets and the level of agreement between annotators are likely another reason why model performance is not the same across all systems and metrics. Additionally, the dataset imbalance in the annotated dataset poses a challenge, impacting model training and evaluation, particularly in accurately predicting less represented sentiment classes.

# 7 Conclusion

In conclusion, our study utilises the MetaLangNEWS-COMMENTS-Hr corpus to explore sentiment analysis on newspaper comments, focussing on annotating and analysing 2938 sentences across 26 articles. Through annotation and cleaning processes facilitated by our annotation environment, we achieve a moderate inter-annotator agreement of 0.47 using Fleiss' Kappa. Employing MaChAmp and various transformer models, including mBERT, cseBERT, and BERTić, our experiments revealed nuanced performance variations influenced by dataset origin, model specificity, and the challenge of sentiment analysis in informal textual contexts. Despite these complexities, our findings underscore the viability of leveraging specialised models like BERTić for achieving robust sentiment analysis results in Croatian newspaper comments. The results show that automatic annotation works best when the language model is trained on a dataset using the same style of language.

**Table 3: Results.** The table presents the performance metrics (accuracy and F1-score) of different models (mBERT, cseBERT, BERTic) across various systems. REF represents the reference system, which has been exclusively trained and tested on the Cro-FiReDa dataset. BASE-1 represents our first baseline system, trained and developed on Cro-FiReDa before tested on our annotated LA dataset. In contrast, our second baseline system BASE-2 has used Cro-FiReDa only for training, while development and test have been conducted with the LA dataset. Finally, COM represents our experimental system which is only trained, developed and tested on the LA dataset. For each model, the accuracy (ACC) and F1-score (F1) are reported, highlighting the comparative effectiveness of each approach. The highest achieved values are shown in bold

| | mBERT | | cseBERT | | BERTić | |
|---|---|---|---|---|---|---|
| | ACC | F1 | ACC | F1 | ACC | F1 |
| REF | 0.75 | 0.70 | 0.79 | 0.75 | **0.80** | **0.77** |
| BASE-1 | 0.44 | 0.44 | **0.57** | **0.54** | 0.56 | 0.53 |
| BASE-2 | 0.47 | 0.45 | **0.59** | **0.54** | 0.56 | 0.53 |
| COM | 0.68 | 0.54 | 0.68 | **0.59** | **0.71** | 0.41 |

# 8 Future Work

Further improvement to the approach can be achieved by increasing the size of the dataset. We can either further annotate the dataset or combine the existing dataset with the newspaper dataset from Bošnjak and Karan (2019). Not only would this provide more data using the same style of language, but also increase the variety of newspaper types, as the majority of our sentences are derived from articles from the Croatian newspaper "Večernji list", while Bošnjak and Karan's dataset consists of comments from the newspaper "24 sata". Another improvement of results can be achieved by balancing the dataset and adding more sentences, labelled as positive or mixed. Additionally, further improvement of training for the annotators as well as annotation guidelines can increase the inter-annotator agreement. By doing so, the annotations for the training dataset would increase in quality, leading to better training of the language model and increasing its accuracy. Also, we use the default parameters defined by MaChAmp. When we fine-tune these parameters, increased results are to be expected. Finally, to gain insights from the dataset, the next step is to identify the targets of the sentiment within the sentences. For that, we can either use the approach from Babić et al. (2021) of clustering topics from COVID-19-related tweets. An attractive alternative is also enhancing our experimental setup by training aspect detection and automatically defining a target within the annotation process. Here, we can use the targets, identified by the annotators, as a training base.

## Limitations

The performance of pre-trained models in highly specialised or domain-specific tasks may be limited due to the broad coverage of topics in their training data. The pre-trained models learn from the data they are trained on, which can result in the introduction of any inherent biases in the training data. This bias can affect model outputs, particularly when the data do not represent all demographic, cultural, or social groups.

## References


Babić, K., Petrović, M., Beliga, S., Martinčić-Ipšić, S., Matešić, M., and Meštrović, A. (2021). Characterisation of covid-19-related tweets in the croatian language: framework based on the cro-cov-csebert model. *Applied Sciences*, 11(21):10442.

Bogetić, K. and Batanović, V. (2020). Annotated corpus of croatian language-related news comments MetaLangNEWS-COMMENTS-hr. Slovenian language resource repository CLARIN.SI.

Bošnjak, M. and Karan, M. (2019). Data set for stance and sentiment analysis from user comments on croatian news. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 50–55.

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.

Durham, M. (2016). Changing attitudes towards the welsh english accent: A view from twitter. *Sociolinguistics in wales*, pages 181–205.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Garrett, P. (2006). Language attitudes. In *The Routledge Companion to Sociolinguistics*, pages 136–141. Routledge.

Kalogjera, D. (2001). On attitudes toward Croatian dialects and on their changing status. *International Journal of the Sociology of Language*, (147):91–100.

Kralj Novak, P., Smailović, J., Sluban, B., and Mozetič, I. (2015). Sentiment of emojis. *PloS one*, 10(12):e0144296.

Ljubešić, N. and Lauc, D. (2021). Bertić-the transformer language model for Bosnian, Croatian, Montenegrin and Serbian. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 37–42.

Mohammad, S. (2016). A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 174–179.

Mønnesland, S. (1997). Emerging literary standards and nationalism. The disintegration of Serbo-Croatian. In *Actas do I Simposio Internacional sobre o Bilingüismo*, pages 1103–1113.

Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

Thakkar, G., Mikelić Preradović, N., and Tadić, M. (2021). Multi-task learning for cross-lingual sentiment analysis. In *2nd International Workshop on Cross-lingual Event-centric Open Analytics (CLEOPATRA 2021)*, pages 76–84.

Thakkar, G., Preradović, N. M., and Tadić, M. (2023). Croatian film review dataset (cro-fireda): A sentiment annotated dataset of film reviews. In *Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023)*, pages 25–31.

Ulčar, M. and 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, Brno, Czech Republic, September 8–11, 2020, Proceedings 23*, pages 104–111. Springer.

Van Der Goot, R., Üstün, A., Ramponi, A., Sharaf, I., and Plank, B. (2021). Massive choice, ample tasks (machamp): A toolkit for multi-task learning in nlp. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 176–197.

Vlada Republike Hrvatske/Hina (2024). Sjednica vlade: Zakonom o hrvatskom jeziku osigurava se sustavna i stručna skrb o njemu. https://vlada.gov.hr/vijesti Last Accessed: 2024-06-27.

# Building a Large Language Model for Moderately Resourced Language: A Case of Croatian

**Gaurish Thakkar, Vanja Štefanec, Daša Farkaš, Marko Tadić**
University of Zagreb, Faculty of Humanities and Social Sciences
Institute of Linguistics
Ivana Lučića 3, Zagreb, Croatia
{gthakkar, vstefane, dfarkas, marko.tadic}@ffzg.unizg.hr

**Abstract.** *Large Language Models (LLMs) have recently demonstrated significant advancements in processing natural language tasks. Consequently, the tide in the natural language processing (NLP) and language technologies (LT) turned towards the development of LLMs for different languages. As a significant number of monolingual and multilingual LLMs are being developed and deployed, it has become crucial to assess their capacities and performance. This document presents a summary of the current activities to develop a monolingual LLM for Croatian (HR-LLM) as a moderately resourced language with less than ten million speakers. The paper presents the relevant previous work, explains why the model is needed, which methodology was used, and how the HR-LLM was trained and how it will be evaluated for selected downstream tasks.*

**Keywords.** Large Language Models, LLMs, GPT, Evaluation, Croatian Language, Natural Language Processing, Neural Networks, Deep Learning.

## 1 Introduction

In recent years, in the natural language processing (NLP) and language technologies (LT) there has been significant progress in the development of language models (LMs), particularly large language models (LLMs). Today, many LLMs are trained on massive datasets of natural language text and they have the capability to perform a wide range of NLP/LT tasks at a state-of-the-art level. One of the recent key advances in LLMs was the development of transformer architecture (Vaswani et al., 2017). The transformer is a neural network architecture that is well-suited for modelling long-range dependencies in text, and this approach enabled the training of LLMs that can represent complex linguistic relationships and thus generate more fluent text. Transformer-based models (Kenton and Toutanova, 2019; Liu et al., 2019; Brown et al., 2020) have influenced several industries and fields by enabling task automation, improving efficiency, and expanding the possibilities for human-AI collaboration. Their ability to generate human-like text and take into account the co-text has made them a transforming force in the world of AI and technologies that are being built upon. Most of the LLMs created so far are multilingual, i.e., they were trained using texts in different languages, and the number of languages varies from 2 to 200+.

One inherent problem with multilingual LLMs is the unbalanced dataset used for training (Lai et al., 2023). The languages with large amounts of training data are usually well represented in the model, whereas languages with limited resources are constrained by the paucity of training data (Kahana et al., 2024). Studies (Martin et al., 2020; Papadimitriou et al., 2023; Virtanen et al., 2019), have shown that models trained in only a few languages (bilingual/trilingual LLMs) or in a single language (monolingual LLMs) perform better in a number of NLP tasks than models trained in many languages (multilingual LLMs). There are also recent attempts to distil the monolingual model from a larger multilingual LLM in order to come up with a smaller and more efficient monolingual LLM (Singh et al., 2023). That is the main reason for instigating this project.

We strongly believe that this case study can serve as an example for the building of a LLM for a moderately resourced language with less than 10 million speakers, or for a lower-resourced language with more speakers. The reason for this is that the overall production of e-text, digitalisation and global usage of mobile devices open the doors to the collecting and harvesting of sufficient language data even for languages that were until recently considered to stay beyond the digital divide, and all this even without the need to artificially generate additional training data, i.e., to use the data augmentation approaches. It seems like many of the initial pioneering multilingual LLMs will soon be in the position to be either replaced by monolingual LLMs or, more probably, fine-tuned with additional monolingual data for the majority of monolingual tasks. In this respect and from the perspective of digital language equality (see e.g., the project funded by the EU Parliament with a similar name[1]), the production of a LLM for a lan-

[1] https://european-language-equality.eu/

guage could be considered as one of the building blocks of the contemporary BLARK (Krauwer, 1998) since its existence ensures the level of development of LT for a language, which is expected today. We believe that the real digital language equality should incorporate LLM as a highly valuable type of language resource.

The main objective of this particular project, Croatian Extended Reality Extensions (HR-XR-XTEND)[2], a FSTP subproject of a Horizon Europe-funded project UTTER[3], is to build a large monolingual generative pre-trained transformer model for Croatian (HR-LLM). As Croatian has already been included in multilingual LLMs, it remains unclear whether the presence of other languages within the LLM influences their performance with Croatian, particularly across different tasks. Since at this point in the development of LLMs we still lack the standardised and balanced evaluation tools for measurement of multilingual LLMs' performance, one of the ways to establish their performance baseline is to construct a monolingual LLM. To establish that baseline performance, it is necessary to construct and evaluate a monolingual LLM. The secondary objective is to develop a set of evaluation benchmarks and methods that are specifically tailored to measure the performance of any LLM when it is applied to tasks involving Croatian language. The third objective of this project is to successfully implement the HR-LLM in the production process of the industrial partner.

We propose the creation of a LLM tailored to address the needs of lower-resourced languages, focusing on those with less than 6 million speakers (Language Report Croatian, 2022). The success of this project could serve as proof of a concept that a similar building process could be applicable to other languages with less than 10 million speakers. The research primarily focuses on training an LLM for a language like Croatian using text collections already gathered by the project beneficiary and other publicly available datasets that are mostly crawled. This could serve as an example for a number of languages that don't have a lot of language resources at disposal. Primarily, it could contribute to the research theme of deep natural language understanding (NLU) for lower-resourced languages. The paper is structured as follows: after the introduction section, the related work is described, objectives and methods are presented in Section 3, while experiments and evaluation are presented in Sections 4 and 5, respectively. The final remarks and HR-LLM availability are mentioned in the concluding section 6.

## 2 Related Work

The recently introduced transformer architecture was first introduced in the paper by Vaswani et al. (2017). The key advance has been the use of pre-training. In pre-training, an LLM is trained on a large and diverse dataset of text without any specific task in mind, thus becoming a multipurpose resource.

The state-of-the-art LLMs that are based on the transformer architecture are: BERT (Kenton and Toutanova, 2019) developed by Google AI. BERT model had 110 million parameters. BERT is pre-trained on a massive amount of text data, which enables it to come close to general language understanding. During pre-training, BERT is trained to predict missing words, also known as masked language modelling (MLM), within sentences. This unsupervised pre-training process allows BERT to learn rich contextual representations. The bidirectional aspect of BERT is crucial because it allows the model to "understand" the co-text of a word by considering both its left and right neighbours in a sentence, enabling it to capture more complex language patterns. RoBERTa (Liu et al., 2019) is a LLM architecture that is based on the BERT framework. RoBERTa was pretrained on a much larger corpus of text data compared to the original BERT model. While BERT used a static masking pattern during pretraining, RoBERTa introduced a dynamic masking approach and a sentence-boundary objective during pretraining, which helps the model recognise sentence boundaries more effectively. Unlike BERT, RoBERTa does not use the Next Sentence Prediction task during pretraining. RoBERTa-Base and RoBERTa-Large have 137 and 355 billion parameters, respectively. GPT-3 (Brown et al., 2020) is a Large Language Model (LLM) that utilises the transformer architecture. GPT-3 has 175 billion parameters, which is significantly more than any other LLM that was available at the time of its release. This large size allows GPT-3 to learn more complex linguistic relationships and generate more fluent and informative text. GPT-NeoX (Black et al., 2022) is a 20-billion-parameter autoregressive LLM developed by EleutherAI. It uses the transformer architecture and is pre-trained on a massive dataset of text and code. GPT-NeoX is available to the public under an open-source licence. Hungarian GPT-3 (Yang et al., 2023) was developed by the Hungarian Research Centre for Linguistics in collaboration with the University of Pécs. It was the first LLM trained on a Hungarian only dataset with a 6.7 billion-parameter GPT language model. The model was trained using the GPT-NeoX implementation on a dataset of 36.3 billion tokens. Currently, one of the largest models in the world is the GPT-4 (OpenAI et al., 2024), with more than 1 trillion parameters. It is a multimodal LLM. It is the fourth in its series of GPT foundation models. Another model created by OpenAI is ChatGPT, which is a multilingual LLM with code and creative text generation capabilities. Meta AI has also published the LLaMA Models (Touvron et al., 2023a,b) (Llama 1 and 2), which is a collection of foundational language models ranging from 7 billion to 70 billion parameters. Bloom (Workshop et al., 2022) is a massive, multilingual language model (MLM) with 176 billion parameters, trained on a massive dataset of text and code. It can generate text, translate languages, write

[2]https://hr-xr-xtend.ffzg.unizg.hr
[3]https://he-utter.eu

various kinds of creative content, and answer your questions in an informative way. The 7-billion parameter model Mixtral 7B (Jiang et al., 2023) employs an innovative mixture-of-experts (MoE) framework to facilitate training that is both scalable and efficient. By employing this methodology, Mixtral 7B is capable of attaining competitive performance while utilising a considerably reduced number of parameters in contrast to conventional dense models. As a result, it serves as a valuable standard against which other studies in the field are measured. GEITje 7B (Rijgersberg and Lucassen, 2023) is a large open Dutch LLM developed by further training Mistral 7B (Jiang et al., 2023) on no less than 10 billion tokens of Dutch text from the Dutch Gigacorpus and the MADLAD-400 web crawling corpus. Marinova et al. (2023) created BERT-WEB-BG and GPT-WEB-BG LLMs, which are Bulgarian language counterparts of BERT and GPT-2. The web-scraped dataset underwent a specialised method for source filtering, subject selection, and lexicon-based elimination of improper language during the pre-training phase.

**Table 1:** List of the most widely used multilingual LLMs and several similar monolingual LLMs

| Name | Languages | Parameters |
|---|---|---|
| (m)BERT | En/multilingual | 110M |
| RoBERTa | En | 137B-355B |
| GPT-3 | multilingual | 175B |
| GPT-NeoX | En | 20B |
| Hungarian GPT-3 | Hu | 6.7B |
| GPT-4 | multilingual | 1T |
| LLaMA 1-2 | multilingual | 7-70B |
| Bloom | multilingual | 176B |
| Mixtral 7B | multilingual | 7B |
| GEITje 7B | Nl | 7B |
| GPT-WEB-BG | Bg | 1.5B |

In the reminder of this section we also list several LLMs that have been trained in Croatian. All of them are multilingual, and there is no monolingual model built exclusively on Croatian data. Apart of already presented widely used multilingual LLMs, here we mention other less used LLMs that were still trained with Croatian data. Language-agnostic BERT Sentence Encoder (LaBSE) (Feng et al., 2022) is a BERT-based model trained for sentence embedding in 109 languages. X-MOD (Pfeiffer et al., 2022) is a MLM trained on filtered Common Crawl data containing 81 languages. This model reuses the tokenizer of XLM-R. The model has been pre-trained with language-specific modular components (language adapters). CroSloEngual BERT (Ulčar and Robnik-Šikonja, 2020) is a trilingual model (110 million parameters) using BERT-base architecture, trained on Croatian, Slovenian, and English corpora. Focusing on three languages, the model performs better than multilingual BERT while still offering an option for cross-lingual knowledge transfer. BERTić (Ljubešić and Lauc, 2021) is a joint multilingual LLM of four distinct languages: Bosnian, Croatian, Montenegrin and Serbian. Cro-CoV-cseBERT (Babić et al., 2021) is a LLM based on the CroSloEngual BERT and has been further trained on a large corpus of texts related to COVID-19 in the Croatian language (Cro-CoV-Texts) which contains 186,738 news articles, 500,504 user comments related to COVID-19 published on Croatian online news portals, and 28,208 COVID-19 tweets in the Croatian language. Several additional language models (LLMs) have been developed under the InfoCoV project, such as Cro-CoV-BERTić, Senti-CoV-cseBERT, and Multi-Cro-CoV-cseBERT. These models are specifically trained on text data connected to COVID-19. The multilingual parliamentary model (XLM-R-parla) (Mochtak et al., 2023) is the XLM-R-large model, additionally pre-trained on texts of parliamentary proceedings. Texts for the additional pre-training, 1.7 billion words in size, come from the ParlaMint corpus and the EuroParl corpus. EUBERT (European Parliament EUBERT Hugging Face, 2023) is a pretrained BERT uncased model that has been trained on a vast corpus of documents in all 24 EU official languages and published by the European Publications Office. These documents span the last 30 years, providing a comprehensive dataset that encompasses a wide range of topics and domains. HR-RoBERTa (HuggingFace macedonizer hr-roberta-base, 2021) model is a LLM-pretrained model on Macedonian using a MLM objective and has been further trained on Croatian data. The HR-GPT2 (HuggingFace macedonizer hr gpt2, 2021) is a LLM pretrained on a large corpus of Croatian data in a self-supervised fashion to provide text generation capabilities. TwHIN-BERT (Zhang et al., 2022) is a multilingual Tweet language model (250 and 550 million parameters) that is trained on 7 billion Tweets from over 100 distinct languages. It not only outperforms similar models in tasks such as text classification, but also in social recommendation tasks such as predicting user-to-Tweet engagement.

# 3 Objectives and Methods

The project goals were set to:

1. collect at least six billion tokens of Croatian texts and prepare that data for HR-LLM training;
2. train the LLM for the Croatian language using monolingual data only;
3. evaluate the HR-LLM for downstream tasks.

Our objective was to develop a monolingual model specifically designed for the Croatian language and based on monolingual data only. Croatian as a moderately resourced language has so far appeared only in multilingual LLMs (e.g., mBERT, BERTić, CroSloEngual) and it is not clear whether other languages contribute to improvement or degrading of LLMs' performance when using on Croatian texts alone. With the existence of a monolingual model, this baseline is set, and any increase or decrease in performance could be

**Table 2:** List of LLMs that include Croatian

| Name | Language |
|---|---|
| (m)BERT | En/multilingual |
| RoBERTa | En |
| GPT-3 | multilingual |
| GPT-4 | multilingual |
| LLaMA 1-2 | multilingual |
| LaBSE | multilingual |
| X-MOD | multilingual |
| CroSloEngual BERT | Hr, Sl, En |
| BERTić | Bo, Hr, Me, Sr |
| Cro-CoV-cseBERT | Hr, Sl, En |
| Cro-CoV-BERTić | Bo, Hr, Me, Sr |
| XLM-R-parla | multilingual |
| HR-RoBERTa | Mk, Hr |
| HR-GPT2 | Mk, Hr |
| EUBERT | multilingual |
| TwHIN-BERT | multilingual |

measured in comparison to that baseline. To build a monolingual LLM, a large number of tokens had to be collected from various sources: corpora and text collections from the existing repositories (although often under-performing for Croatian, e.g. results from the Oscar project[4]) and newly collected data (online and offline) that were not available in previous data collection campaigns. Table 3 lists the large repositories or data and collection campaigns that we used as sources of data.

The LLM type that was targeted stemmed from the European AI initiative OpenGPT-X[5] and recent achievements in building the first GPT-3 model for Hungarian (PULI GPT-3SX[6]). The OpenGPT-X at this moment features only five major European languages (English, French, German, Italian, and Spanish), with over 50 million speakers, while other European languages are not covered. On the other hand, the Hungarian GPT-3 LLM demonstrates that such an endeavour could be achieved for a language with approximately 13 million speakers (Language Report Hungarian, 2022).

We used the datatrove[7] library to perform the near deduplication with MinHashLSH and a threshold of 0.72. This was done following the advice that LLMs trained on deduplicated data are better and memorise less of their data (Lee et al., 2022). After the deduplication, where repeating paragraphs were removed from the training data, the total number of tokens used for training was downsized to 7.72 billion, compared to the original dataset which contains 8.9 billion tokens.

Our training arrangement was divided into three distinct steps:

1. We initialise the model and train it using the training configurations provided by the gpt-neox library. The following configurations were chosen as the starting candidates for training our model. This setup is designed to measure the performance of the model when trained exclusively on Croatian data from the beginning. We trained models with 160M, 350M, 410M, and 1.4B parameters respectively.

2. We conduct constant pre-training (CPT) on the gpt-2 model using the Croatian data outlined in the preceding section. This model validates its performance by leveraging existing knowledge of the English language.

3. We engage in ongoing pre-training (OPT) on the "unsloth/gemma-2-9b-bnb-4bit" model, which is a compressed version of Gemma 2. This is a multilingual language model that aids in the evaluation of Croatian CPT performance within a multilingual framework.

**Table 3:** Non-exhaustive list of largest data sources used for training the HR-LLM (beta version) with approximate size in tokens. *Croatian texts only

| Name | Approx Size |
|---|---|
| CLASSLA Hr Web corpus 1.0 | 2.5 billion |
| CC100-Hr Dataset | 2.27 billion |
| Corpus of Croatian News Feeds | 2.25 billion |
| Parallel data for En-Hr on OPUS Resources*, | 1.48 billion |
| Hr-news from XLM-R-BERTić dataset | 1.4 billion |
| news/legal corpus | 175 million |
| Corpus of Croatian Academic Theses | 312 million |
| ParaCrawl* | 69.96 million |
| Riznica from XLM-R-BERTić dataset | 69.51 million |
| MARCELL Croatian legislative subcorpus | 56 million |
| CURLICAT Croatian corpus | 49 million |
| MARCELL Croatian-English Parallel Corpus of Legislative Texts* | 14.3 million |
| Romance-Croatian Parallel Corpus (literary works) | 2.5 million |
| Total | 8.9 billion |

# 4 Experiments

The experimental phase was focused on developing and evaluating the HR-LLM architecture and training process. This involved:

[4]https://oscar-project.org/
[5]https://www.aleph-alpha.com
[6]https://nytud.github.io
[7]https://github.com/huggingface/datatrove

1. Collecting and preprocessing of the training dataset (see section 3 above): This data will also be partially used within newly established initiatives such as Language Data Space[8] and ALT-EDIC[9].
2. Training the model: For training the model, we used the publicly available open source library at gpt-neox[10].
3. Local GPU infrastructure was utilised for experimentation, while EuroHPC and University of Zagreb Computing Centre supercomputers with GPU nodes were utilised for advanced experimenting and final version training.
4. Configuring the training hyperparameters: hyperparameters such as the learning rate, batch size, and number of epochs were tuned to achieve optimal performance.
5. Evaluating the model: the model was evaluated using language model evaluation Harness[11].

# 5 Evaluation

To evaluate the performance of the trained HR-LLM, we will conduct benchmarking tasks including Named Entity Recognition (NER), Sentiment Analysis (SA), and Choice Of Plausible Alternatives (COPA) on the Croatian language using the BENCHić (Rupnik et al., 2023) benchmarking dataset [12]. Furthermore, we intend to assess the LLM's proficiency in knowledge completion in Croatian by utilising established datasets such as MMLU (Hendrycks et al., 2021b,a), ARC (Clark et al., 2018), HellaSwag (Zellers et al., 2019), Belebele (Bandarkar et al., 2023), and TruthfulQA (Dac Lai et al., 2023). Here we present our initial zero-shot evaluation results on benchmarking datasets in Table 4. Since the full evaluation of the HR-LLM performance is still a work-in-progress, its final results are beyond the scope of this paper.

# 6 Conclusion

Recent advancements in NLP/LT have demonstrated that LLMs have become essential for reliable and efficient language processing capabilities. Imbalanced data distributions among different languages during multilingual LLM pre-training demonstrably weakens the monolingual proficiency of a multilingual LLM applied to the specific monolingual tasks, particularly in a lower-resourced languages.

As an attempt to overcome this situation, we have developed the first monolingual Croatian LLM: HR-LLM (beta version). This paper presents the results of the project HR-XR-XTEND so far (trained HR-LLM for Croatian) and the remaining work to be completed (its full evaluation for selected downstream tasks). While the primary objective of this study is to develop a HR-LLM using the Croatian language data only, the key findings and achievements may have broader implications for lower-resourced languages in Europe and around the globe in subsequent developments.

The direct outputs of the project, particularly the current state of the HR-LLM (beta version), will be available through HR-CLARIN[13] repository with permissive licenses. The final version of the HR-LLM will also be availabe through the same repository after the completion of the project.

# 7 Acknowledgements

This project has received funding from the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No 101070631 and from the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (Grant No 10039436).

[8] https://language-data-space.ec.europa.eu/

[9] https://language-data-space.ec.europa.eu/related-initiatives/alt-edic

[10] https://github.com/EleutherAI/gpt-neox

[11] https://github.com/EleutherAI/lm-evaluation-harness

[12] https://github.com/clarinsi/benchich

[13] https://www.clarin.hr

| benchmark | metric | 160M | 350M | 160M | 1.4B | gpt2-cpt |
|---|---|---|---|---|---|---|
| arc_hr | acc | 18.91 | **20.96** | 20.36 | 20.44 | 19.5 |
| | acc_norm | 23.44 | **25.49** | 25.06 | 24.89 | 23.7 |
| belebele_hrv_Latn | acc | 22.78 | 23 | **23.11** | 22.67 | 22.89 |
| | acc_norm | 22.78 | 23 | **23.11** | 22.67 | **22.89** |
| hellaswag_hr | acc | 28.43 | 29.87 | 30.08 | **31.36** | 26.67 |
| | acc_norm | 30.07 | 32.74 | 33.38 | **35.52** | 27.95 |
| m_mmlu_hr | acc | 22.65 | **25.21** | 22.8 | 22.54 | 23.05 |
| truthfulqa_hr_mc1 | acc | 25.88 | 24.58 | 25.75 | 26.27 | **27.05** |
| truthfulqa_hr_mc2 | acc | 43.82 | 42.21 | 42.34 | 42.52 | **46.88** |

**Table 4:** Results of zero-shot evaluation on benchmarking datasets ARC, Belebele, MMLU, HellaSwag and TruthfulQA.

# References


Babić, K., Petrović, M., Beliga, S., Martinčić-Ipšić, S., Matešić, M., and Meštrović, A. (2021). Characterisation of covid-19-related tweets in the croatian language: framework based on the cro-cov-csebert model. *Applied Sciences*, 11(21):10442.

Bandarkar, L., Liang, D., Muller, B., Artetxe, M., Shukla, S. N., Husa, D., Goyal, N., Krishnan, A., Zettlemoyer, L., and Khabsa, M. (2023). The belebele benchmark: a parallel reading comprehension dataset in 122 language variants. *arXiv preprint arXiv:2308.16884*.

Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., He, H., Leahy, C., McDonell, K., Phang, J., Pieler, M., Prashanth, U. S., Purohit, S., Reynolds, L., Tow, J., Wang, B., and Weinbach, S. (2022). GPT-NeoX-20B: An open-source autoregressive language model. In Fan, A., Ilic, S., Wolf, T., and Gallé, M., editors, *Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models*, pages 95–136, virtual+Dublin. Association for Computational Linguistics.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Clark, P., Cowhey, I., Etzioni, O., Khot, T., Sabharwal, A., Schoenick, C., and Tafjord, O. (2018). Think you have solved question answering? try arc, the ai2 reasoning challenge. *arXiv:1803.05457v1*.

Dac Lai, V., Van Nguyen, C., Ngo, N. T., Nguyen, T., Dernoncourt, F., Rossi, R. A., and Nguyen, T. H. (2023). Okapi: Instruction-tuned large language models in multiple languages with reinforcement learning from human feedback. *arXiv e-prints*, pages arXiv–2307.

European Parliament EUBERT Hugging Face (2023). Huggingface/europeanparliament/eubert. Accessed: 19-02-2024.

Feng, F., Yang, Y., Cer, D., Arivazhagan, N., and Wang, W. (2022). Language-agnostic bert sentence embedding.

Hendrycks, D., Burns, C., Basart, S., Critch, A., Li, J., Song, D., and Steinhardt, J. (2021a). Aligning ai with shared human values. *Proceedings of the International Conference on Learning Representations (ICLR)*.

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., and Steinhardt, J. (2021b). Measuring massive multitask language understanding. *Proceedings of the International Conference on Learning Representations (ICLR)*.

HuggingFace macedonizer hr gpt2 (2021). Huggingface/macedonizer/hr-gpt2. Accessed: 19-02-2024.

HuggingFace macedonizer hr-roberta-base (2021). Huggingface/macedonizer/hr-roberta-base. Accessed: 19-02-2024.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., and Sayed, W. E. (2023). Mistral 7b.

Kahana, A., Mathew, J. S., Bleik, S., Reynolds, J., and Elisha, O. (2024). Evaluation methodology for large language models for multilingual document question and answer. *arXiv preprint arXiv:2402.01065*.

Kenton, J. D. M.-W. C. and Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Krauwer, S. (1998). Elsnet and elra: A common past and a common future. *ELRA Newsletter*, 3(2):4–5.

Lai, V., Ngo, N., Pouran Ben Veyseh, A., Man, H., Dernoncourt, F., Bui, T., and Nguyen, T. (2023). ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. In Bouamor, H., Pino, J., and Bali, K., editors, *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13171–13189, Singapore. Association for Computational Linguistics.

Language Report Croatian (2022). European language equality. Accessed: 19-02-2024.

Language Report Hungarian (2022). European language equality. Accessed: 19-02-2024.

Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., and Carlini, N. (2022). Deduplicating training data makes language models better. In Muresan, S., Nakov, P., and Villavicencio, A., editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8424–8445, Dublin, Ireland. Association for Computational Linguistics.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach.

Ljubešić, N. and Lauc, D. (2021). BERTić - the transformer language model for Bosnian, Croatian, Montenegrin and Serbian. In Babych, B., Kanishcheva, O., Nakov, P., Piskorski, J., Pivovarova, L., Starko, V., Steinberger, J., Yangarber, R., Marcińczuk, M., Pollak, S., Přibáň, P., and Robnik-Šikonja, M., editors, *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 37–42, Kiyv, Ukraine. ACL.

Marinova, I., Simov, K., and Osenova, P. (2023). Transformer-based language models for Bulgarian. In Mitkov, R. and Angelova, G., editors, *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 712–720, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., and Sagot, B. (2020). CamemBERT: a tasty French language model. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.

Mochtak, M., Rupnik, P., and Ljubešić, N. (2023). The parlasent multilingual training dataset for sentiment identification in parliamentary proceedings. *arXiv preprint arXiv:2309.09783*.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.-L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R., Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H. W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus, L., Felix, N., Fishman, S. P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafstein, M., Gray, S., Greene, R., Gross, J., Gu, S. S., Guo, Y., Hallacy, C., Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S., Jain, S., Jang, J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B., Jun, H., Kaftan, T., Kaiser, Ł., Kamali, A., Kanitscheider, I., Keskar, N. S., Khan, T., Kilpatrick, L., Kim, J. W., Kim, C., Kim, Y., Kirchner, J. H., Kiros, J., Knight, M., Kokotajlo, D., Kondraciuk, Ł., Kondrich, A., Konstantinidis, A., Kosic, K., Krueger, G., Kuo, V., Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C. M., Lim, R., Lin, M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A., Malfacini, K., Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne, A., McGrew, B., McKinney, S. M., McLeavey, C., McMillan, P., McNeil, J., Medina, D., Mehta, A., Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa, E., Mossing, D., Mu, T., Murati, M., Murk, O., MÃ©ly, D., Nair, A., Nakano, R., Nayak, R., Neelakantan, A., Ngo, R., Noh, H., Ouyang, L., O'Keefe, C., Pachocki, J., Paino, A., Palermo, J., Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos, A., Pavlov, M., Peng, A., Perelman, A., de Avila Belbute Peres, F., Petrov, M., de Oliveira Pinto, H. P., Michael, Pokorny, Pokrass, M., Pong, V. H., Powell, T., Power, A., Power, B., Proehl, E., Puri, R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach, K., Ross, C., Rotsted, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar, S., Sastry, G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K., Sherbakov, T., Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M., Sitkin, J., Slama, K., Sohl, I., Sokolowsky, B., Song, Y., Staudacher, N., Such, F. P., Summers, N., Sutskever, I., Tang, J., Tezak, N., Thompson, M. B., Tillet, P., Tootoonchian, A., Tseng,

E., Tuggle, P., Turley, N., Tworek, J., Uribe, J. F. C., Vallone, A., Vijayvergiya, A., Voss, C., Wainwright, C., Wang, J. J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann, C., Welihinda, A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter, C., Wolrich, S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T., Yoo, S., Yu, K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S., Zheng, T., Zhuang, J., Zhuk, W., and Zoph, B. (2024). Gpt-4 technical report.

Papadimitriou, I., Lopez, K., and Jurafsky, D. (2023). Multilingual BERT has an accent: Evaluating English influences on fluency in multilingual models. In Beinborn, L., Goswami, K., Muradoğlu, S., Sorokin, A., Kumar, R., Shcherbakov, A., Ponti, E. M., Cotterell, R., and Vylomova, E., editors, *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 143–146, Dubrovnik, Croatia. Association for Computational Linguistics.

Pfeiffer, J., Goyal, N., Lin, X. V., Li, X., Cross, J., Riedel, S., and Artetxe, M. (2022). Lifting the curse of multilinguality by pre-training modular transformers. *arXiv preprint arXiv:2205.06266*.

Rijgersberg, E. and Lucassen, B. (2023). Geitje: een groot open nederlands taalmodel.

Rupnik, P., Kuzman, T., and Ljubešić, N. (2023). BENCHić-lang: A benchmark for discriminating between Bosnian, Croatian, Montenegrin and Serbian. In Scherrer, Y., Jauhiainen, T., Ljubešić, N., Nakov, P., Tiedemann, J., and Zampieri, M., editors, *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2023)*, pages 113–120, Dubrovnik, Croatia. Association for Computational Linguistics.

Singh, P., Maladry, A., and Lefever, E. (2023). Too many cooks spoil the model: Are bilingual models for Slovene better than a large multilingual model? In Piskorski, J., Marcińczuk, M., Nakov, P., Ogrodniczuk, M., Pollak, S., Přibáň, P., Rybak, P., Steinberger, J., and Yangarber, R., editors, *Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023)*, pages 32–39, Dubrovnik, Croatia. Association for Computational Linguistics.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023a). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al. (2023b). Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Ulčar, M. and 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, Brno, Czech Republic, September 8–11, 2020, Proceedings 23*, pages 104–111. Springer.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Virtanen, A., Kanerva, J., Ilo, R., Luoma, J., Luotolahti, J., Salakoski, T., Ginter, F., and Pyysalo, S. (2019). Multilingual is not enough: Bert for finnish.

Workshop, B., Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., et al. (2022). Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.

Yang, Z. G., Laki, L. J., Váradi, T., and Prószéky, G. (2023). Mono-and multilingual gpt-3 models for hungarian. In *International Conference on Text, Speech, and Dialogue*, pages 94–104. Springer.

Zellers, R., Holtzman, A., Bisk, Y., Farhadi, A., and Choi, Y. (2019). Hellaswag: Can a machine really finish your sentence? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Zhang, X., Malkov, Y., Florez, O., Park, S., McWilliams, B., Han, J., and El-Kishky, A. (2022). Twhin-bert: a socially-enriched pre-trained language model for multilingual tweet representations. *arXiv preprint arXiv:2209.07562*.

0em 1.5em

# Exploring usability in combining data sources for detecting toxic behavior in social media

**Oliver Lohaj, Ján Paralič**
Technical university of Košice
Faculty of Electrical Engineering and Informatics
Department of Cybernetics and Artificial Intelligence
Vysokoškolská 4, 040 01 Košice, Slovakia
{oliver.lohaj, jan.paralic}@tuke.sk

**Ruslana Buinytska**
Technical university of Košice
Faculty of Electrical Engineering and Informatics
Department of Cybernetics and Artificial Intelligence
Vysokoškolská 4, 040 01 Košice, Slovakia
ruslana.buinytska@student.tuke.sk

***Abstract.*** *This article deals with the issue of detecting toxic behavior on social networks. The main goal of the article is to find an effective approach to detecting toxic behavior using a selected machine learning method with a combination of multiple data sources. The article is divided into a theoretical and a practical part. The theoretical part provides an overview of toxic behavior in social media, existing research and various machine learning models such as CNN (Convolutional Neural Network) and BERT (Bidirectional Representation Coding from Transformers). The practical part follows the CRISP-DM methodology and focuses on data preparation and modeling. Several experiments were conducted to evaluate the performance of different models and configurations. The results showed that combining multiple data sources and advanced machine learning models can improve the accuracy of toxic behavior detection. The article concludes that the proposed approach is effective and stresses the importance of proper data preparation and comprehensive evaluation metrics.*

**Keywords.** BERT, CNN, social media, toxic behavior, usability

## 1 Introduction

The use of social media has increased a lot in the last decade (Perrin, 2021; Chou, 2009), which is directly related to the increase in the amount of time people spend online. This trend is most pronounced among young people, who spent up to twice as much time online in 2020 compared to the previous decade, averaging two to three and a half hours a day (Lei, 2024). This trend is significant because it shows that young people are increasingly oriented towards their online lives. They use social media to communicate with friends and family, get information (Mertz, 2024), but they are also exposed to negative influences such as bullying or misinformation (Parent, 2019), which can affect their view of the world. Toxic behaviour in the online environment can take many forms and often manifests itself in the form of insults, hateful comments, the spread of disinformation or cyberbullying (Beknazar-Yuzbashev, 2022). These manifestations can have a negative impact on the psychological health of individuals, can cause emotional stress and even lead to long-term psychological problems (Zsila, 2022).

In addition, toxic behavior can also lead to the division of the community, exacerbating online discussions, and creating a hostile environment on the Internet. And with each passing year, the situation in the online environment may deteriorate (Avalle, 2024). The aim of this research is to investigate how the combination of multiple data sources affects the accuracy of detection of toxic behavior in social media and to create a model using this data to increase detection efficiency.

To solve this task, we proceeded using the CRISP-DM methodology (Wirth and Hipp, 2000) and worked in Google Colab with the Python programming language. We first modified the data and then created models, which we evaluated at the end by monitoring several criteria.

## 2 Analysis of the current state and motivation

Before starting our own experiments, we analyzed what experiments had been done before, what models and datasets had been used, and what results had been achieved.

In an interesting study (Fan, 2021), the authors analyzed the detection of toxic behavior on social networks using deep learning. For this purpose, they chose the BERT model, specifically the BERT-base version, which they trained on the labelled dataset Kaggle Toxic Comment Classification Challenge[1]). The authors of this paper compared the performance of the BERT model with three other models, namely

[1] https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge

Multilingual BERT, RoBERTa and DistilBERT. To assess the performance of the classification model, they used the AUC-ROC as a performance metric. This metric is the standard ranking method used in competitions on the Kaggle platform.

Before the model was created, preprocessing operations were used, such as removing punctuation, references, and non-English words. A pre-trained tokenizer for "bert_base_uncased" was used for tokenization. The evaluation showed that the BERT-based model performs very well (0.98561 (AUC-ROC)) as a public score and 0.98603 (AUC-ROC) as a private score) in detecting toxic content. The multilingual BERT model achieved the second-best position, followed by the DistilBERT and RoBERTa models.

In other article (Anand and Enswari, 2019) simple neural networks, CNNs and LSTMs are used with or without pre-trained GloVe models. The Jigsaw dataset was used to train the model (Sorensen, 2017). The metrics used to evaluate the performance of each model are classification success (accuracy), which describes the percentage of samples correctly placed in their actual class, and the value of the loss. The results show that ANN achieved a high classification success rate (98%) on trained data with minimal loss during training.

However, it achieved a high loss value on the test data and was the third worst in terms of classification success among all models. CNN achieved a relatively high success rate of classification in training (97.8%) and a low loss value (5.42%). It has a lower loss on test data than ANN, but also a lower classification success rate. LSTM achieved a lower classification success rate and a higher loss value during training compared to ANN and CNN. However, it achieved a better classification success rate and a lower loss value than previous models on test data.

The GloVe & CNN model had a lower classification success rate during training, so its loss value was higher than previous models. It performed better than other models on test data, although its loss value was similar to that of the LSTM. GloVe & LSTM had a lower classification success rate during training and its loss value was higher than that of ANN and CNN. On test data, it achieved significant results in terms of classification success and loss rate. GloVe & LSTM & CNN achieved the lowest classification success rate and the highest loss value during both training and testing. This model achieved the worst results of all.

These analyses of existing research show that the CNN and BERT models are among the best approaches for detecting toxic comments on social media (Lee, 2020). Deep-learning-based models such as CNN can capture complex patterns and structures in data, allowing for successful classification of toxic comments (Georgakopoulos, 2018).

Transformer-based models, such as BERT, on the other hand, can better understand the context of text and capture its meaning, which also leads to excellent results in detecting toxic content (Ashwin Geet d'Sa, 2020).

Based on these findings, we decided to train CNN and BERT models to detect toxic comments on social media and find out the usability of these models. These models offer a combination of advanced word processing techniques and the ability to capture complex patterns in data, which should lead to a high success rate in detecting toxic content. The goal of our work from a business perspective is to find a portable model that will be able to detect toxic behavior in text data. In the future, it could be used to create a safe and positive environment for users on social media. In terms of data mining, we expect the creation of a classification model into two classes. Data analysis makes it possible to identify relevant patterns and trends in user behavior and provides the basis for creating effective and accurate models for detecting toxic behavior on social media.

# 3 Datasets

As the title of the article suggests, we chose to combine multiple data sources for training our models. Our approach involved training the model on one dataset and evaluating it on a second dataset. This method allows us to assess whether a model trained on one dataset can generalize knowledge sufficiently to identify toxic content across different social networks.

Our research was based on pre-existing datasets. The first dataset is the Twitter dataset by Davidson. Tweets are collected based on keywords from the hatebase.org lexicon. The dataset contains 24783 tweets and annotations made by CrowdFlower. It is important to mention here, that the app named Twitter has been rebranded by new owner to X in the April of 2023 (Hayes, 2024).

Each tweet has been annotated by at least 3 annotators, and the consensus of annotators is 92%. The labels correspond to three classes: hate speech, offensive language and none, with the percentages of each class being 6%, 77% and 17% respectively.

Before reviewing the dataset, we reduced the number of classes to two: toxic marked as 1 and non-toxic as 0. When examining the dataset, we found that the average length of toxic comments is smaller than the average length of non-toxic comments, see. Fig. 1 and the following words are most often used:

b**ch – 11480, h*e – 4352, like – 2873, f**k – 2269, p**si – 2267, n**ga – 2019, get – 1785, a*s – 1599, s**t – 1312, u – 1307.

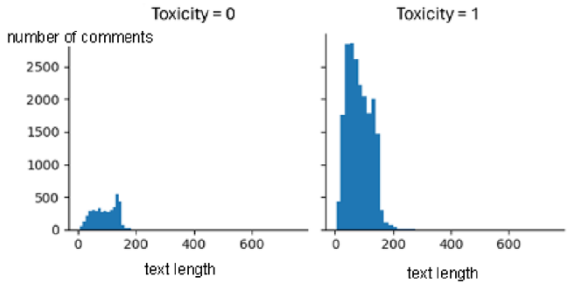



**Figure 1.** Comments length of different classes on Twitter dataset by Davidson

The second dataset is the Toxic Tweets Dataset, which contains tweet data from different sources that showed a significant imbalance in the number of examples between classes. This dataset is the result of a combination of various existing datasets that have been obtained from publicly available sources. The goal of combining these datasets was to achieve a balance between classes, which could lead to better results when training classification models.

The dataset contains 56745 tweets. The labels correspond to two classes: toxic behavior marked as 1 and non-toxic as 0, with the percentages of each class being 57% and 43% respectively. Regarding the length of comments of different classes, the result is the same as in the previous dataset: the average length of toxic comments is less than the average length of non-toxic comments, see Fig. 2. It also turns out that the words that are most often found in toxic messages are similar to the first set of data: b**ch – 11553, h*e – 4370, like – 4122, love – 4020, day – 3381, get – 3046, u – 2501, f*ck – 2480, p**si – 2318.

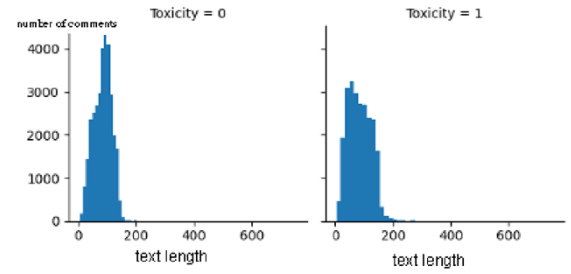



**Figure 2.** Comments length of different classes on Toxic Tweets Dataset

The third dataset is a dataset from Wikipedia's Talk edit pages. Wikipedia is the largest and most popular reference resource on the internet, with about 500 million unique visitors per month. These editing pages are a crucial forum for community interaction, where the contributing community discusses, debates, and communicates about changes related to a specific topic.

Overall, this dataset provides a data source for research and development of NLP models and community management tools on the online platform Wikipedia. The dataset contains 5000 messages. The labels correspond to two classes: toxic and non-toxic behavior, with the percentages of each class being 9% and 91% respectively. The graph (see Fig. 3) shows that the average length of toxic comments is less than the average length of non-toxic comments. We can see that, unlike datasets from the social network X (Twitter), the keywords in toxic comments are different: articl – 2232, page – 1874, wikipedia – 1455, edit – 1378, talk – 1322, f**ck – 1198, use - 1131, please – 1046, a*s – 990, would – 966.

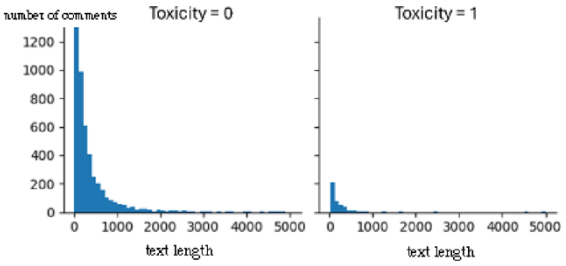



**Figure 3.** Comments length of different classes of Wikipedia toxicity dataset

# 4 Modelling

In the modelling phase, we focused on several types of models: BERT and CNN. We used two metrics to evaluate the performance of the models: accuracy and false negative rate. Accuracy provides us with an overall level of classification correctness, while false negative rates help us identify the risk of misclassification of toxic comments.

The sequence of our experiments is designed in such a way that almost every experiment depends on the results of the previous experiment. Based on each experiment we decided how to design further procedures to consider various aspects that can affect the success of detection.

The 1st experiment was designed to provide a basic view of the CNN model's performance in detecting toxic comments without the use of under sampling and using specific model parameters. As part of the experiment, we used the CNN model. We divided the Twitter dataset by Davidson into a training and test set in an 80:20 ratio without the use of under sampling. Before training, we performed tokenization and vectorization of the text data. We trained the model over 10 epochs. After training, we evaluated the performance of the model on a test set. We achieved a classification success rate: 0.934 false negative rate: 0.026.

For the 2nd experiment, we used the same parameters as for the first model, except that we used Twitter dataset by Davidson data with under sampling for training set. Despite the fact that we modified the training set, the results of the second model were worse than those of the first model. We achieved a lower classification success rate of 0.929 and a higher false negative rate of 0.068.

We have several explanations as to why this procedure may have negatively affected the results. First, CNN works well with local patterns and dependencies in data, but they can miss global patterns, especially if certain data classes are underrepresented. When the number of training examples is limited, this can have a negative impact on the CNN's ability to identify and generalize these patterns to a wider range of data.

Furthermore, in the case of significant under sampling, the model may be overtrained for limited patterns and specific interference elements in the training data. This phenomenon leads to a reduced ability of the model to generalize to new ones, as the model may tend to reproduce errors learned from an incomplete training set. In addition, CNNs require a relatively large amount of data to function effectively due to their depth and large number of parameters. Data reduction causes a reduction in the variability and complexity of information available for learning, which can limit the model's ability to learn and adapt to different situations or diverse data.

For the 3rd experiment, we used again the same parameters as for the first model. However, unlike the first two experiments, we used the Twitter dataset by Davidson, which was preprocessed using a tweet preprocessing script written in the Ruby programming language, which was interpreted in Python (Davidson, 2017). Nevertheless, the results of the third model were worse compared to the first model. We achieved a lower classification success rate of 0.929 and a higher false negative rate of 0.028.

In the 4th experiment, we decided to test how well the BERT model can cope with the classification of toxic comments. We took advantage of BERT's built-in tokenization feature, which is optimized for this model. For training and testing, we still use the Twitter dataset by Davidson, split 80:20 in favor of the training set. Initially, we initialized the model and tokenizer, using the "*bert-base-uncased*" model, which is one of the most common BERT models. We have specified that the number of output classes is 2, which corresponds to our binary classification task. Next, we tokenized the texts and prepared the data for training and testing. For training, we used *DataLoader* to load data in the form of batches, which we then used to update the model weights during training. To minimize losses, we used the *AdamW* optimizer with a learning coefficient $5x10^{-5}$.

The training cycle itself was implemented for 3 epochs. In each epoch, we iterated over the training data and performed back-propagation of the error and update the model weights according to the gradient. After completing the training, we tested and evaluated the model. The success rate of the model classification reached a value of 0.955, indicating that more than 95% of the cases were classified correctly. The false negative rate was 0.021, indicating that only about 2% of toxic comments were misclassified as harmless.

In the following experiments, we focused on testing the robustness of models on other datasets, which was the main goal. We wanted to see, if model trained on one kind of data is able to classify toxic comments from another platform. If yes, that would mean that the trained models are universal for wider range of use, than just using it on the same data as it was trained on.

During 5th experiment, we decided to use the model we successfully tested in the first experiment (CNN) and applied it to other Toxic Tweets Dataset. The results of this experiment gave us an important insight into how our model behaves on different data. Although we achieved a relatively good classification success rate on the different dataset (0.778), it was important to note that the false negative rate was slightly higher (0.103) compared to previous experiments. During the *"data preparation"* phase, we found that, there is a small difference, between the frequently used words in toxic comments, which may have prevented the comments from being classified correctly.

In the 6th experiment, we used the BERT model that we trained in the 4th experiment. As in the previous experiment with the CNN model, the aim of the experiment was to test the model on the Toxic Tweets Dataset, i.e. different from the one used during the training. In the previous experiment, we described how the difference between the data affected the success of the classification, now let's look at the differences at the model level. By comparing these results with the previous CNN experiment, we found that the BERT model achieved a higher classification success rate (0.882 vs. 0.778). This clearly shows that the BERT model is able to identify toxic texts better. However, the BERT model also achieved a slightly higher false negative rate (0.106 versus 0.103), which means that more toxic texts were misclassified as harmless. Although we are aware of this imperfection, the results clearly show that the BERT model surpassed the CNN model in the success rate of classifying toxic texts. This result shows that the BERT model is a better choice for toxicity detection compared to the CNN model. In this case, the difference in the false negative rate between the two models is relatively small. In case of a larger difference, we would have to consider other factors and perhaps we would prefer the CNN model, although its classification success rate is not as high as that of the BERT model.

In the following 7th experiment, we decided to use a BERT-based model based on the results of previous experiments. Since the Toxic Tweets Dataset contains more toxic comments, we decided to use it as a training set. We chose the Wikipedia toxicity dataset for testing. We left the parameters of the model unchanged, as we observed good results with these parameters in previous experiments. Our results show that we have achieved a classification success rate of 0.869. We also recorded a false negative rate of 0.478. Since the training dataset contained more than 55,000 comments, there is a chance that the BERT model may

have been overtrained. Another reason could be that the toxic comments in Wikipedia contain different keywords than in the Toxic Tweets Dataset, which is why the model had trouble identifying new toxic patterns. In addition, differences in data annotations may have affected the results. These differences may be due to subjective annotator evaluations, inconsistent labeling standards and variability in the training process, but it showed us that training on one set and testing on other achieves pretty good results and the model can be deemed usable.

In the 8th experiment, we decided to verify our assumptions about overlearning the model from the 7th experiment. To do this, we used a Twitter dataset by Davidson containing fewer messages, with the assumption that the model would be able to better generalize dependencies across a smaller dataset. We achieved a high classification success rate of 0.947 and a lower false negative rate of 0.355. In this way, our assumptions were confirmed, and this experiment gave us valuable insight into the behavior of the model depending on the size of the training data.

We decided to create the last experiment mirroring the 8th experiment, where we used a dataset from Wikipedia for training and a Twitter dataset by Davidson for testing. This strategy is useful because in real conditions we can encounter text from different sources and different writing styles. Using multiple source datasets and testing the model on texts from different domains can provide a comprehensive view of its overall performance and ability to identify toxicity, regardless of the specific source of the text.

The results of this experiment showed that the model achieved a classification success rate of 0.822 and a false negative rate of 0.16. These results, although lower compared to previous experiments, still show the model's ability to work on new and diverse data and can be deemed usable. We have to consider that the training dataset was relatively small compared to the test dataset (5000 comments out of only 437 toxic), but at the same time it identified 17316 toxic comments, which indicates that the dataset contains a selection of high-quality and representative samples. However, the high false negative rate suggests that the model still has trouble identifying some of the toxic texts in the Twitter dataset by Davidson. This result highlights the importance of using multiple source datasets and continuously improving the model to improve its ability to identify toxicity in different contexts and text sources.

# 6 Evaluation and comparison

Based on the experiments carried out and the data analyzed, we can conclude that modelling toxic behavior in social media is a complex task that requires fine-tuning and understanding of the details that affect the performance of the models. In this research, we used CNN and the BERT model, which are cutting-edge methods in the field of NLP, to detect toxic behavior in datasets originating from Twitter and Wikipedia.

The results of the experiments, provided in Table 1 show that: The CNN model achieved a high level of classification success without the need for subsampling, indicating that it is capable of processing unevenly distributed data. The BERT model has shown a better classification success rate in detecting toxic comments, which confirms its robustness and adaptability to different linguistic contexts in the data.

However, in one case, it showed a higher rate of false negativity, indicating the need for further tuning. Data preprocessing was critical to model performance, with modifications such as removing special characters and transforming text data impacting the models' ability to identify toxic comments. Overfitting models is a risk to consider, especially when the model shows high accuracy on training data, but subsequently has problems with generalization on test data.

Compared to other previously published papers, this article advances the field of detecting toxic behavior on social media in several distinctive ways. Previous studies have primarily focused on evaluating the effectiveness of individual machine learning models like BERT or CNN using single data sources. For example, studies by Fan (2021) and Anand and Enswari (2019) assessed models like BERT, RoBERTa, DistilBERT, CNNs, and LSTMs individually on specific datasets, often achieving high performance metrics but within limited contexts.

This article, however, stands out by emphasizing the combination of multiple data sources for training and testing models, which significantly enhances the generalizability and robustness of the detection systems. By training models on one dataset and evaluating them on another, the research demonstrates a more comprehensive approach to understanding model performance across different social media platforms, thus addressing a key limitation in earlier studies.

Moreover, while earlier works often used standard datasets like the Kaggle Toxic Comment Classification Challenge and Jigsaw datasets, this article introduces and experiments with diverse datasets including Twitter datasets by Davidson, Toxic Tweets Dataset, and Wikipedia's Talk edit pages. This diversity in data sources provides a broader perspective on the models' capabilities and limitations, offering more generalizable and practical insights into their effectiveness and moreover, their usability.

Furthermore, the methodological rigor of following the CRISP-DM methodology ensures a structured and replicable approach to data preparation, modeling, and evaluation. This contrasts with some previously mentioned studies that might not have provided as detailed a methodological framework. The article's use of comprehensive evaluation metrics, including accuracy and false negative rate, adds depth to the analysis, offering a more nuanced understanding of

model performance compared to simpler metrics like accuracy alone used in some earlier research.

This article not only confirms the effectiveness of advanced models like BERT and CNN but also highlights the critical importance of combining multiple data sources and rigorous data preparation. It provides a more robust framework for detecting toxic behavior on social media, pushing the boundaries of current research and offering practical solutions for real-world applications.

**Table 1.** Comparison of experiment results

| Experiment number | Model | Dataset for training | Dataset for testing | Classification success rate | False negative rate |
|---|---|---|---|---|---|
| *1* | CNN | Twitter by Davidson (unsampled, self-processed) | Twitter by Davidson | 0.934 | 0.026 |
| *2* | CNN | Twitter by Davidson (under sampling, self-processed) | Twitter by Davidson | 0.929 | 0.068 |
| *3* | CNN | Twitter by Davidson (no subsampling, preprocessing: Ruby script) | Twitter by Davidson | 0.929 | 0.028 |
| *4* | BERT | Twitter by Davidson | Twitter by Davidson | 0.955 | 0.021 |
| *5* | CNN | Twitter by Davidson (unsampled, self-processed) | Toxic Tweets Dataset | 0.778 | 0.103 |
| *6* | BERT | Twitter by Davidson | Toxic Tweets Dataset | 0.882 | 0.106 |
| *7* | BERT | Toxic Tweets Dataset | Wikipedia dataset | 0.869 | 0.478 |
| *8* | BERT | Twitter by Davidson (100%) | Wikipedia dataset | 0.947 | 0.355 |
| *9* | BERT | Wikipedia dataset | Twitter by Davidson (100%) | 0.822 | 0.16 |

# 7 Conclusion

Based on our experiments, we can conclude that combining multiple sources can be very effective to find a model that would work best, as it includes aspects such as model generalization, which shows how the model has learned to recognize toxic behavior in different contexts and situations, and whether it can work effectively with new, unknown data. Also, the combination of training and testing on different data points to the robustness of the model, whether the model can work with different communication styles, language variations and forms of toxic behavior, which makes it a more reliable tool for detection in different contexts. The use of a combination of multiple data sources achieves better efficiency of the model in detecting toxic behavior in different contexts and situations, increasing its practical applicability and reliability.

But before using testing datasets different from the training dataset, a sufficient level of classification success and a false negative rate on the underlying data must be achieved. We have found that this is affected by data preprocessing, not every preprocessing technique may be suitable for the selected dataset, so it is better to test several variants. Other important thing to consider is to not always considering emojis, hyperlinks and similar details. This leads to the lower efficiency of preprocessing, and sometimes it can only create unnecessary noise.

In addition, we realized that a balanced dataset is not always the best solution, during subsampling we can miss valuable data based on which the model could learn to better classify comments. It is also very important to choose the right model that is suitable for the specific task, in our experiments we used CNN and BERT and found that BERT captures context and global dependencies better. An equally important factor is what metrics will be used to evaluate the effectiveness of the model. For toxic behavior detection models, it is important to use other metrics such as false negative rates in addition to classification success.

Despite its strengths, this research has several limitations that highlight areas for improvement. The small training dataset compared to the test dataset suggests the potential for enhanced performance with larger data. While the model's high false negative rate points to areas needing refinement, it also shows capability in identifying much toxic content. The risk

of overfitting calls for improved training methods to boost generalization. Data preprocessing impacts performance, indicating the need for varied techniques. Annotation subjectivity and labeling inconsistencies suggest a need for standardized approaches. Using CNN and BERT models offers insights into their strengths, with BERT excelling in context and dependency capture. Variations in model effectiveness across datasets underscore the importance of considering linguistic and contextual diversity to improve robustness and applicability.

This work provides a contribution to understanding the dynamics of toxic behavior detection in social media using multiple data sources and modelling. The combination of different approaches to data pre-processing and the application of different models can lead to improved detection of antisocial behavior. The experiments showed the importance of the adaptability of models to new datasets and their ability to generalize learned patterns of behavior. We also found out that BERT and CNN models enhance usability in detecting toxic behavior in social media. BERT effectively captures language context, while CNN identifies patterns efficiently. Together, they improve accuracy and adaptability across various datasets and can be deemed of high usability in future research.

In the context of social media, where it is important to quickly and accurately identify toxic comments, our findings show that finding the optimal model and proper data pre-processing is crucial. For practical applications, it is important to balance detection accuracy and the model's ability to adapt to constantly changing behavior patterns in the online space.

## Acknowledgments

This research was funded by the Slovak Research and Development Agency under the contracts No. APVV-22-0414 and APVV-20-0232 and by the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic and the Slovak Academy of Sciences under grants No. 1/0685/21 and 1/0259/24.

## References

Anand, Mukul, and R. Eswari. "Classification of abusive comments in social media using deep learning." *2019 3rd international conference on computing methodologies and communication (ICCMC)*. IEEE, 2019.

Avalle, Michele, et al. "Persistent interaction patterns across social media platforms and over time." *Nature* 628.8008 (2024): 582-589.

Beknazar-Yuzbashev, George, et al. "Toxic content and user engagement on social media: Evidence from a field experiment." *Available at SSRN 4307346* (2022).

Chou, Wen-ying Sylvia, et al. "Social media use in the United States: implications for health communication." *Journal of medical Internet research* 11.4 (2009): e1249.

d'Sa, Ashwin Geet, Irina Illina, and Dominique Fohr. "Bert and fasttext embeddings for automatic detection of toxic speech." *2020 International Multi-Conference on:"Organization of Knowledge and Advanced Technologies"(OCTA)*. IEEE, 2020.

Hayes, David. C. 2024. "X." The editors of *Encyclopaedia Britannica. Available* online: https://www.britannica.com/money/Twitter

Davidson, T., Warmsley, D., Macy, M., & Weber, I. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. In *Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM '17)* (pp. 512-515). Montreal, Canada. GitHub repository, https://github.com/t-davidson/hate-speech-and-offensive-language

Fan, Hong, et al. "Social media toxicity classification using deep learning: real-world application UK brexit." *Electronics* 10.11 (2021): 1332.

Georgakopoulos, Spiros V., et al. "Convolutional neural networks for toxic comment classification." *Proceedings of the 10th hellenic conference on artificial intelligence*. 2018.

Lee, Jieh-Sheng, and Jieh Hsiang. "Patent classification by fine-tuning BERT language model." World Patent Information 61 (2020): 101965.

Lei, Xiaojing, et al. "The relationship between social media use and psychosocial outcomes in older adults: a systematic review." *International Psychogeriatrics* (2024): 1-33.

Mertz, Breanne A., et al. "# SocialMediaWellness: Exploring a research agenda and conceptualization for healthy social media consumption." *Journal of Consumer Behaviour* 23.2 (2024): 321-335.

Parent, Mike C., Teresa D. Gobble, and Aaron Rochlen. "Social media behavior, toxic masculinity, and depression." *Psychology of Men & Masculinities* 20.3 (2019): 277.

Perrin, Andrew. "Social media usage." *Pew research center* 125 (2015): 52-68.

Sorensen, C. J., et al. *Toxic Comment Classification Challenge*. Kaggle, 2017. Available online: https://kaggle.com/competitions/jigsaw-toxic-comment-classification-challenge

Wirth, Rüdiger, and Jochen Hipp. "CRISP-DM: Towards a standard process model for data

mining." *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*. Vol. 1. 2000.

Zsila, Ágnes, et al. "Toxic behaviors in online multiplayer games: Prevalence, perception, risk factors of victimization, and psychological consequences." *Aggressive Behavior* 48.3 (2022): 356-364.

# A Review of Methods for Analyzing Online Media Texts Using Large Language Models

# Pregled metoda za analizu *online* medijskih tekstova primjenom velikih jezičnih modela

**Ana Meštrović**
University of Rijeka,
Faculty of Informatics and Digital Technologies
R. Matejčić 2, 51000 Rijeka
amestrovic@uniri.hr

***Abstract***. *This paper explores the potential applications of generative artificial intelligence in tasks related to media text analysis. It highlights the capabilities of large language models, which, with their advanced pre-training and context awareness, support numerous NLP tasks essential for media text analysis, such as keyword extraction, topic modeling, named entity recognition, sentiment analysis, and disinformation detection. The study also delves into techniques such as prompt engineering, retrieval-augmented generation, and fine-tuning, which further enhance the performance of large language models. In addition, this paper provides an overview of a set of NLP-based methods for media text analysis, enhanced with generative AI techniques. By integrating these advanced technologies, it is possible to develop a powerful system for the comprehensive media texts analysis.*

**Keywords.** Media Texts Analysis, Natual Language Processing, Generative Artificial Intelligence

***Sažetak***. *Ovaj rad istražuje potencijalne primjene generativne umjetne inteligencije u zadacima vezanim uz analizu medijskih tekstova. Ističe mogućnosti velikih jezičnih modela koji, zahvaljujući svom naprednom pred-treniranju i znanju o kontekstu, podržavaju brojne NLP zadatke bitne za analizu medijskih tekstova, kao što su ekstrakcija ključnih riječi, modeliranje tema, prepoznavanje imenovanih entiteta, analiza sentimenta i otkrivanje dezinformacija. Studija se također bavi tehnikama kao što su inženjering upita, RAG tehnike i fino podešavanje, koje dodatno poboljšavaju performanse velikih jezičnih modela. Pored toga, rad pruža pregled skupa metoda za analizu medijskih tekstova temeljenih na NLP-u, poboljšanih generativnim AI tehnikama. Integracijom naprednih tehnologija moguće je razviti moćan sustav za sveobuhvatnu analizu medijskih tekstova.*

**Ključne riječi.** Analiza medijskih tekstova, obrada prirodnog jezika, generativna umjetna inteligencija

## 1 Introduction

Online media text analysis is important aspect of media monitoring that involves the systematic collection, analysis, and interpretation of media content from online platforms. This process plays a crucial role in understanding public discourse, tracking the dissemination of information, and identifying emerging trends. Media text analysis allows stakeholders to extract relevant insights from the vast amounts of texts published on online platforms daily. The ability to efficiently and accurately analyze media texts has become increasingly important in today's digital age, where information is abundant and constantly evolving. The importance of media text analysis and media

## 1 Uvod

Analiza medijskih tekstova važan je aspekt praćenja medija koji uključuje sustavno prikupljanje, analizu i tumačenje medijskog sadržaja s online platformi. Ovaj proces ima važnu ulogu u razumijevanju javnog diskursa, praćenju širenja informacija i prepoznavanju novih trendova. Analiza medijskih tekstova omogućava dionicima da izvuku relevantne uvide iz velike količine tekstova objavljenih na online platformama svakodnevno. Sposobnost učinkovitog analiziranja medijskih tekstova postaje sve važnija u današnje digitalno doba, gdje su informacije obilne i neprestano se mijenjaju. Važnost analize medijskih tekstova i

monitoring becomes even more pronounced in the context of crises such as war, global warming, pandemics, etc. In such scenarios, timely and accurate information is critical for decision-making, public safety, and strategic communication. Media texts analyis during crises helps authorities and organizations track the spread of information, identify misinformation or disinformation, and gauge public sentiment. For instance, during the COVID-19 pandemic, monitoring media texts enabled health authorities to understand public concerns, combat misinformation, and effectively disseminate health guidelines (Beliga et al., 2022.). Similarly, in times of conflict, media analysis helps in understanding the narrative, tracking propaganda, and assessing the impact of information campaigns on public opinion (Sufi, 2023.).

Natural Language Processing (NLP) has revolutionized the way media texts are analyzed, offering advanced techniques to process and understand large volumes of text efficiently. Analyzing media content provides insights into relevant topics and entities (such as individuals, institutions, locations, and events) present in mass media, allowing users to quickly and accurately identify information pertinent to their interests. Methods such as keyword extraction, topic modeling, named entity recognition (NER), sentiment analysis, text classification, and disinformation detection are employed. These techniques enable information extraction and the prediction of information spread, all driven by deep learning and large language models.

Large Language Models (LLMs), such as BERT (Devlin et al., 2018) and GPT (Radford et al., 2019, Achiam et al., 2023), and their successors, have demonstrated remarkable capabilities in various NLP tasks, including sentiment analysis, NER, and topic modeling. These models leverage extensive pre-training on diverse text corpora, enabling them to capture intricate linguistic patterns and contextual nuances. By applying LLMs to media text analysis, researchers and practitioners can achieve more accurate and insightful interpretations of media content, ultimately enhancing the effectiveness of media monitoring. This review aims to explore the various methods for analyzing media texts using LLMs, highlighting their potential and application in improving media texts analysis tasks.

Many research studies have already analyzed the potential of NLP in media texts anaylsis and media monitoring task (Farzindar et al., 2015; Germann et al., 2018). However, there is still a lack of research papers that study specific NLP tasks related to media texts analysis in the context of generative artificial intelligence (GenAI). Thus, the main goal of this study is to provide an overview of the set of GenAI-based methods that can be used for a comprehensive media texts analysis. Specifically, the main aim of this study is to show how LLMs can improve tasks of

praćenja medija postaje još izraženija u kontekstu kriza poput ratova, globalnog zagrijavanja, pandemija, itd. U takvim situacijama, pravovremene i točne informacije su ključne za donošenje odluka, javnu sigurnost i stratešku komunikaciju. Analza medijskih tekstova tijekom kriza pomaže vlastima i organizacijama da prate širenje informacija, identificiraju dezinformacije te procijene javno mnijenje. Npr., tijekom pandemije COVID-19, praćenje medijskih tekstova omogućilo je odgovornim osobama da razumiju zabrinutost javnosti, suzbiju dezinformacije i učinkovito prenesu zdravstvene smjernice (Beliga i sur., 2022.). Slično, u razdoblju sukoba, analiza medija pomaže u razumijevanju narativa, praćenju propagande i procjeni utjecaja informacijskih kampanja na javno mnijenje (Sufi, 2023.).

Obrada prirodnog jezika (engl. Natural Language Processing, NLP) unaprijedila je način na koji se analiziraju medijski tekstovi, nudeći napredne tehnike za učinkovitu obradu i razumijevanje velike količine tekstova. Analiza medijskog sadržaja pruža uvid u relevantne teme i entitete (kao što su pojedinci, institucije, lokacije i događaji) prisutne u masovnim medijima, omogućujući korisnicima brzo i točno prepoznavanje važnih informacija. Primjenjuju se metode poput ekstrakcije ključnih riječi, modeliranja tema, prepoznavanja imenovanih entiteta, analize sentimenta, klasifikacije teksta i detekcija dezinformacija. Ove tehnike omogućuju ekstrakciju informacija i predviđanje širenja informacija, a sve to dodatno mogu unaprijediti duboko učenje i veliki jezični modeli.

Veliki jezični modeli (engl. Large Language Models, LLMs), poput BERT-a (Devlin et al., 2018) i GPT-a (Radford et al., 2019, Achiam et al., 2023), te njihovih nasljednika, pokazali su izvanredne sposobnosti u raznim NLP zadacima, uključujući analizu sentimenta, prepoznavanje imenovanih entiteta i modeliranje tema. Ovi modeli koriste opsežan pred-trening na raznovrsnim tekstualnim korpusima, što im omogućuje obuhvaćanje složenih jezičnih obrazaca i kontekstualnih nijansi. Primjenom LLM-ova u analizi medijskih tekstova, istraživači mogu postići točnije i dublje interpretacije medijskog sadržaja, čime se u poboljšava učinkovitost praćenja medija. Vilj rada je istražiti različite metode za analizu medijskih tekstova korištenjem LLM-ova, ističući njihov potencijal za analizu medijskih tekstova.

Mnoge istraživačke studije već su analizirale potencijal NLP-a u zadacima analize medijskih tekstova i praćenja medija općenito (Farzindar et al., 2015; Germann et al., 2018). Međutim, još uvijek nedostaje radova koji proučavaju specifične NLP zadatke vezane uz analizu medijskih tekstova u kontekstu generativne umjetne inteligencije. Stoga je glavni cilj ove studije pružiti pregled skupa metoda temeljenih na GenAI koje se mogu koristiti

media texts analysis. This paper discusses the potential of transformer-based models, such as BERT and GPT, for media texts analysis. Furthermore, the study analyzes techniques that can improve these models, such as prompt engineering, retrieval-augmented generation (RAG), and fine-tuning.

The deep understanding of language allows LLMs to perform more accurately and efficiently across various media texts anaylsis tasks. For instance, in keyword extraction and topic modeling, LLMs can identify and cluster relevant terms with greater precision, providing a more comprehensive understanding of the topics and trends present in media content. NER and sentiment classification also benefit from LLMs' ability to understand context and disambiguate entities and sentiments within diverse and unstructured media texts. The incorporation of retrieval-augmented generation further enhances the capabilities of LLMs for media texts analysis. RAG combines the strengths of retrieval-based and generation-based models, allowing the system to access external knowledge sources and generate contextually accurate and up-to-date analyses. This approach is particularly beneficial for tasks like disinformation detection, where accessing current and domain-specific information is crucial for identifying false or misleading content. By retrieving relevant information from external databases and knowledge graphs, RAG can provide enriched and context-aware insights that classical NLP methods might miss. Fine-tuning LLMs on specific datasets developed for media texts analysis can significantly improve their performance by adapting the models to the particularities of the media content being analyzed. Fine-tuning allows the models to learn the specific linguistic features, terminologies, and contextual nuances relevant to media texts, thereby increasing the accuracy and relevance of the analyses. Techniques such as prompt engineering can also optimize the interaction with LLMs, ensuring that the models generate more precise and relevant outputs tailored to the specific needs of media texts analysis.

The rest of the paper is organized as follows. The next section covers the motivation and background. The third section details methods based on large language models that can be utilized for efficient media text analysis. The final section presents the discussion and concluding remarks.

# 2 Motivation and Background

The methods and techniques integrated into media texts analysis and media monitoring systems should enable the automatic extraction of relevant information from diverse media sources, offering a comprehensive overview of content, relationships, public opinion or sentiment, and trends related to the topics and entities of interest. Developing methods to extract relevant information from a large volume of

za sveobuhvatnu analizu medijskih tekstova. Konkretno, glavni cilj ove studije je pokazati kako LLM-ovi mogu unaprijediti analizu medijskih tekstova. Ovaj rad raspravlja o mogućnostima modela temeljenih na transformerima, kao što su BERT i GPT za analizu medijskih tekstova. Nadalje, studija analizira tehnike koje mogu poboljšati te modele, kao što su inženjering upita, generiranje s poboljšanjem pretrage (engl. Retrieval-Augmented Generation RAG) i fino podešavanje.

Duboko razumijevanje jezika omogućuje LLM-ovima da preciznije i učinkovitije obavljaju različite zadatke analize medijskih tekstova. Na primjer, u ekstrakciji ključnih riječi i modeliranju tema, LLM-ovi mogu identificirati i grupirati relevantne pojmove s većom preciznošću, pružajući sveobuhvatnije razumijevanje tema i trendova prisutnih u medijskom sadržaju. Prepoznavanje imenovanih entiteta i klasifikacija sentimenta također imaju koristi od sposobnosti LLM-ova da razumiju kontekst i razjasne entitete i sentimente unutar raznolikih i nestrukturiranih medijskih tekstova.

RAG dodatno poboljšava sposobnosti LLM-ova u analizi medijskih tekstova. RAG kombinira prednosti modela temeljenih na pretraživanju i modela temeljenih na generiranju, omogućujući sustavu pristup vanjskim izvorima znanja i generiranje kontekstualno točnih i ažuriranih analiza. Ovaj pristup posebno je koristan za zadatke poput otkrivanja dezinformacija, gdje je pristup trenutnim i specifičnim informacijama ključan za prepoznavanje lažnog sadržaja. Prikupljanjem relevantnih informacija iz vanjskih baza podataka i grafova znanja, RAG može pružiti obogaćene i kontekstualno svjesne uvide koje klasične NLP metode možda propuštaju.

Fino podešavanje LLM-ova na skupovima podataka (korpusima) generiranim specifično za zadatke analize medijskih tekstova može poboljšati njihovu izvedbu prilagođavanjem modela specifičnostima analiziranog medijskog sadržaja. Fino podešavanje omogućuje modelima da nauče specifične jezične značajke, terminologiju i kontekstualne nijanse relevantne za medijske tekstove, čime se povećava točnost analiza. Tehnike poput inženjeringa upita također mogu optimizirati interakciju s LLM-ovima, osiguravajući da modeli generiraju preciznije rezultate prilagođene specifičnim potrebama za analizu medijskih tekstova.

Ostatak rada je organiziran na sljedeći način. Iduća sekcija pokriva motivaciju i pozadinu istraživanja. Treće dio detaljno opisuje metode temeljene na LLM-ovima koje se mogu koristiti za učinkovitu analizu medijskih tekstova. Završni dio predstavlja raspravu i zaključne primjedbe.

unstructured data from multiple sources in various formats or free text is a challenging task. Additional challenges in exploring media data include constantly changing trends, the vast number of different topics and entities discussed, sensationalism, biased texts, misinformation, and more.

Investigation into neural network architectures has resulted in the emergence of a wide range of neural text representation models (Babić et al., 2020). Particularly, transformer-based models like BERT or GPT have brought about significant transformations in the realm of natural language processing. These models, which are pre-trained on extensive text datasets and recognized as large language models, capture linguistic structures and semantic relationships, enabling them to surpass conventional language models in various NLP assignments. The influence of LLMs goes beyond enhanced accuracy; they facilitate the creation of more sophisticated and user-friendly language-oriented applications. LLMs employ deep learning architectures with billions of parameters, enabling them to comprehend intricate patterns and contextual subtleties in text. Consequently, LLMs have demonstrated improved performance in tasks such as text categorization (Sun et al., 2019; Balkus and Yan, 2022), sentiment analysis (Babić et al., 2021; Beliga et al., 2021), identification of paraphrases (Vrbanec & Meštrović, 2023), machine translation (Zhu et al., 2020; Yang et al., 2020), question-answering (Wang et al., 2019), prediction of information dissemination (Meštrović et al., 2022), and so forth.

In the realm of media text analysis and media monitoring tasks, numerous studies have applied methods and techniques based on deep learning. For example, Bhoi et al. (2020) propose a deep learning-based social media text analysis framework for disaster resource management. There have been research directions aimed at enhancing deep learning models with additional knowledge-based techniques. For instance, Swati et al. (2023) propose a framework for predicting political bias in news headlines. Their framework employs a neural knowledge model and knowledge graphs to simplify, interpret, and explain events that are not explicitly stated in the headlines.

Furthermore, there have been numerous attempts to analyze themes in the media (Korenčić et al., 2018). An important direction in this domain is the analysis of trends and dynamics of themes. To identify changes in themes and content of posts on social media, Zhong & Schweidel (2020) constructed a thematic model with multiple latent change points. This allows the identification of temporal patterns in themes, including the prevalence of themes that remain constant over time, temporary changes in theme prevalence, and themes that continue to change over time. This is particularly important in the context of social media content analysis, as content such as discussions changes rapidly, necessitating further monitoring of relevant information (Srikanth et al.,

# 2 Motivacija i pozadina

Metode i tehnike integrirane u sustave za analizu medijskih tekstova i praćenje medija trebale bi omogućiti automatsku ekstrakciju relevantnih informacija iz različitih medijskih izvora, nudeći sveobuhvatan pregled sadržaja, odnosa, javnog mišljenja ili sentimenta te trendova povezanih s temama i entitetima od interesa. Razvijanje metoda za ekstrakciju relevantnih informacija iz velike količine nestrukturiranih podataka iz više izvora u različitim formatima ili slobodnom tekstu je izazovan zadatak. Dodatni izazovi u istraživanju medijskih podataka uključuju stalno mijenjanje trendova, veliki broj različitih tema i entiteta koji se raspravljaju, senzacionalizam, pristrane tekstove, dezinformacije i slično.

Istraživanje arhitektura neuronskih mreža rezultiralo je pojavom širokog spektra modela za reprezentaciju teksta temeljenih na neuronskim mrežama (Babić i sur., 2020). Posebno transformeri, poput BERT-a ili GPT-a, donijeli su značajne promjene u području obrade prirodnog jezika (engl. Natural Language Processing, NLP). Ovi modeli, koji su unaprijed trenirani na opsežnim tekstualnim skupovima podataka obuhvaćaju jezične strukture i semantičke odnose, omogućujući im da budu bolji od konvencionalnih pristupima u raznim NLP zadacima. Utjecaj LLM-ova nadilazi postojeće pristupe; oni olakšavaju stvaranje sofisticiranijih aplikacija usmjerenih na jezik. LLM-ovi koriste duboke neuronske arhitekture s milijardama parametara, omogućujući im razumijevanje složenih obrazaca i kontekstualnih suptilnosti u tekstu. Posljedično, LLM-ovi su pokazali poboljšane performanse u zadacima poput klasifiacije teksta (Sun et al., 2019; Balkus i Yan, 2022), analize sentimenta (Babić i sur., 2021; Beliga i sur., 2021), identifikacije parafraziranja (Vrbanec i Meštrović, 2023), strojnog prevođenja (Zhu i sur., 2020; Yang i sur., 2020), odgovaranja na pitanja (Wang i sur., 2019), predviđanja širenja informacija (Meštrović i sur., 2022) i tako dalje.

U području analize medijskih tekstova i zadataka praćenja medija, brojne studije primijenile su metode i tehnike temeljene na dubokom učenju. Na primjer, Bhoi i sur. (2020) predlažu okvir za analizu tekstova sa društvenih medija temeljen na dubokom učenju za upravljanje resursima u slučaju katastrofa. Postoje istraživački pravci usmjereni na poboljšanje modela dubokog učenja dodatnim tehnikama temeljenim na znanju. Primjerice, Swati i sur. (2023) predlažu okvir za predviđanje političke pristranosti u naslovima vijesti. Njihov okvir koristi neuronski model znanja i grafove znanja za pojednostavljenje, interpretaciju i objašnjavanje događaja koji nisu eksplicitno navedeni u naslovima.

2021). Besides new approaches, which include the application of deep neural networks, a new direction in NLP research involves embedding background knowledge into learning models. For example, Koloski et al. (2022) recently proposed a new approach for detecting fake news by combining contextual representation with knowledge graphs as a heterogeneous representation of text (Koloski et al., 2022).

Morover, it is important to develop methods and techniques for media txts analysis in low-resourced and medium-resourced languages. In the case of the Croatian language, there have been several studies dedicated to media text analysis and media monitoring. The recent successful application of NLP-based methods for media monitoring has been described in many studies. For example, in Barić et al. (2023), the authors present a sentiment and tone analysis of Croatian news headlines. Sentiment analysis can be combined with topic modeling, as demonstrated in the research analyzing COVID-19-related news on Croatian internet portals (Buhin Pandur et al., 2021). Furthermore, using transfer learning techniques, where models trained on high-resource languages are adapted to low-resource languages, can be a viable strategy. For instance, cross-lingual embeddings and multilingual models such as FinEst BERT and CroSloEngual BERT (Ulčar & Robnik-Šikonja, 2020) can facilitate the transfer of knowledge from well-resourced languages to less-resourced ones, improving the performance of media texts analysis. Additionally, the use of generative AI techniques can be particularly beneficial for low-resourced languages. By incorporating external knowledge sources and fine-tuning on available datasets, generative AI can produce high-quality analyses even in languages with limited resources. The adaptability of these models makes them ideal for handling the diverse linguistic structures and vocabularies of less-resourced languages, thereby enhancing the accuracy and effectiveness of media texts analysis across different linguistic contexts.

Most existing research studies focus on specific domains and specific tasks related to media text analysis (e.g., sentiment analysis, disinformation detection, or topic modeling). Therefore, there is a need for studies on systems that integrate all possible NLP solutions for media text analysis into one comprehensive system. This study aims to explore cutting-edge NLP approaches and techniques based on LLMs and to propose a set of methods for media text analysis.

Osim toga, bilo je brojnih pokušaja analize tema u medijima (Korenčić i sur., 2018). Važan smjer u toj domeni je analiza trendova i dinamike tema. Kako bi identificirali promjene u temama i sadržaju objava na društvenim medijima, Zhong i Schweidel (2020) izgradili su tematski model s više latentnih točaka promjene. Ovo omogućava identifikaciju vremenskih obrazaca u temama, uključujući zastupljenost tema koje ostaju konstantne tijekom vremena, privremene promjene u zastupljenosti tema te zastupljenost tema koje se nastavljaju mijenjati tijekom vremena. Ovo je posebno važno u kontekstu analize sadržaja društvenih medija, budući da se sadržaji poput rasprava brzo mijenjaju te je potrebno dodatno praćenje relevantnih informacija (Srikanth i sur., 2021). Osim novih pristupa, koji uključuju primjenu dubokih neuronskih mreža, novi smjer istraživanja u domeni NLP-a uključuje ugrađivanje pozadinskog znanja u modele učenja. Tako, primjerice, Koloski i sur. (2022) nedavno predlažu novi pristup za otkrivanje lažnih vijesti kombiniranjem kontekstualne reprezentacije s grafovima znanja kao heterogene reprezentacije teksta.

Također, važno je razvijati metode i tehnike za analizu medijskih tekstova u jezicima s malo resursa i jezicima sa srednje dostupnim resursima. U slučaju hrvatskog jezika, postoji više studija posvećenih analizi medijskih tekstova i praćenju medija. Nedavna uspješna primjena metoda temeljenih na NLP-u za praćenje medija opisana je u mnogim studijama. Na primjer, Barić i sur. (2023) predstavljaju analizu sentimenta i tona naslova vijesti na hrvatskom jeziku. Analiza sentimenta može se kombinirati s modeliranjem tema, kao što je prikazano u istraživanju analize vijesti vezanih za COVID-19 na hrvatskim internet portalima (Buhin Pandur i sur., 2021). Nadalje, korištenje tehnika prijenosnog učenja, gdje se modeli trenirani na jezicima s puno resursa prilagođavaju jezicima s malo resursa, može biti održiva strategija. Na primjer, višejezični modeli kao što su FinEst BERT i CroSloEngual BERT (Ulčar i Robnik-Šikonja, 2020) mogu olakšati prijenos znanja s jezika s puno resursa na jezike s malo resursa, poboljšavajući performanse sustava za analizu medijskih tekstova. Dodatno, uporaba generativnih AI tehnika može biti posebno korisna za jezike s malo resursa. Integriranjem vanjskih izvora znanja i finim podešavanjem dostupnih skupova podataka, generativna AI može proizvesti visokokvalitetne analize čak i u jezicima s ograničenim resursima. Prilagodljivost ovih modela čini ih idealnima za obradu raznolikih jezičnih struktura i vokabulara jezika s malo resursa, čime se povećava točnost i učinkovitost analize medijskih tekstova u različitim jezičnim kontekstima.

Većina postojećih istraživačkih studija fokusira se na specifične domene i specifične zadatke vezane uz analizu medijskih tekstova (npr. analiza

# 3 Methods and Techniques for Media Texts Analysis

## 3.1 LLMs and Media Texts Analysis

Large language models, such as BERT, GPT-3, and their successors, have already demonstrated potential in the field of NLP (Balkus & Yan, 2022). These models leverage extensive pre-training on vast text corpora, allowing them to capture intricate linguistic patterns and semantic nuances that are essential for understanding media texts. One of the primary ways LLMs are employed in media texts analysis tasks is through their capability to perform various NLP functions with high accuracy and efficiency.

For keyword extraction, LLMs can identify and extract relevant keywords from a body of text, facilitating the quick identification of key themes and topics within large datasets. This capability is particularly useful in monitoring media coverage on specific issues or events, enabling stakeholders to stay informed about the most pertinent aspects without sifting through vast amounts of information. In topic modeling, LLMs can analyze large volumes of text to uncover underlying themes and trends. By clustering related terms and phrases, these models can generate comprehensive summaries of the main topics being discussed in the media. This is invaluable for organizations seeking to understand public discourse and sentiment around particular subjects or to track the evolution of media narratives over time. NER is another task where LLMs can be useful. By accurately identifying and categorizing entities such as people, organizations, locations, and events mentioned in media texts, LLMs facilitate the organization and retrieval of information. This enables users to quickly pinpoint specific entities of interest and understand their relevance within the context of the media coverage.

Large language models (LLMs) are highly effective for media text classification due to their ability to process and analyze large volumes of text data with remarkable accuracy and efficiency. Within the media texts analysis task, text classification involves categorizing articles, posts, and other media content into relevant categories such as topics, sentiment, or relevance to specific events or entities. LLMs, like BERT and GPT-3, excel in this task by leveraging their deep understanding of language semantics and context, which they acquire through extensive pre-training on diverse text corpora. This enables them to discern subtle differences in language use, such as distinguishing between positive and negative sentiments or identifying nuanced topics within complex text. Furthermore, LLMs can handle the variability and richness of media content, including slang, idioms, and varying writing styles, ensuring accurate classification across different sentimenta, otkrivanje dezinformacija ili modeliranje tema). Stoga postoji potreba za studijama koje istražuju sustave koji integriraju sva moguća NLP rješenja za analizu medijskih tekstova u jedan sveobuhvatan sustav. Ova studija ima za cilj istražiti najnovije NLP pristupe i tehnike temeljene na LLM-ovima te predložiti skup metoda za analizu medijskih tekstova.

# 3 Metode i tehnike za analizu medijskih tekstova

## 3.1 LLM-ovi i analiza medijskih tekstova

Veliki jezični modeli, kao što su BERT, GPT-3 i njihovi nasljednici, već su pokazali potencijal u području obrade prirodnog jezika (Balkus i Yan, 2022). Ovi modeli koriste opsežno pre-treniranje na ogromnim tekstualnim korpusima, što im omogućuje obuhvaćanje složenih jezičnih obrazaca i semantičkih nijansi koje su ključne za razumijevanje medijskih tekstova. Jedan od glavnih načina na koji se LLM-ovi koriste u zadacima analize medijskih tekstova je njihova sposobnost obavljanja raznih NLP zadataka s visokim stupnjem točnosti i učinkovitosti.

Za ekstrakciju ključnih riječi, LLM-ovi mogu identificirati i izdvojiti relevantne ključne riječi iz teksta, olakšavajući brzo prepoznavanje ključnih tema i pitanja unutar velikih skupova podataka. Ova sposobnost posebno je korisna u praćenju na koje načine mediji pokrivaju određena pitanja ili događaje, omogućujući dionicima da budu informirani o najvažnijim aspektima bez potrebe za pretraživanjem ogromnih količina informacija. U modeliranju tema, LLM-ovi mogu analizirati velike količine teksta kako bi otkrili temeljne teme i trendove. Grupiranjem povezanih pojmova i fraza, ovi modeli mogu generirati sveobuhvatne sažetke glavnih tema koje se raspravljaju u medijima. To je vrlo važno za organizacije koje žele razumjeti javni diskurs i sentiment o određenim temama ili pratiti promjene medijskih narativa tijekom vremena.

Prepoznavanje imenovanih entiteta još je jedan zadatak u kojem LLM-ovi mogu biti korisni. Točnim prepoznavanjem i kategorizacijom entiteta kao što su ljudi, organizacije, lokacije i događaji spomenuti u medijskim tekstovima, LLM-ovi olakšavaju organizaciju i dohvat informacija. To korisnicima omogućuje brzo prepoznavanje specifičnih entiteta od interesa i razumijevanje njihove relevantnosti u kontekstu medija.

LLM-ovi vrlo su učinkoviti za klasifikaciju medijskih tekstova zbog svoje sposobnosti obrade i analize velikih količina tekstualnih podataka s visiokom točnošću i učinkovitošću. U okviru analize medijskih tekstova, klasifikacija teksta

media sources. Their capability to learn and adapt to new data through fine-tuning makes them particularly adept at evolving with the dynamic nature of media narratives. By integrating LLMs into media monitoring systems that enables automatic text analysis, organizations can achieve more precise and insightful classifications, facilitating better tracking of trends, public opinion, and emerging issues, ultimately enabling more informed decision-making and strategic communication.

Sentiment classification is enhanced by the nuanced understanding that LLMs have of language. These models can determine the sentiment expressed in media texts—whether positive, negative, or neutral—with high precision. This allows for a more granular analysis of public opinion and the emotional tone of media coverage, which is crucial for reputation management and crisis communication. Disinformation detection benefits greatly from the advanced pattern recognition capabilities of LLMs. By analyzing text for inconsistencies, factual inaccuracies, and other indicators of false information, LLMs can help identify and flag potential disinformation in media content. This is particularly important in the current digital landscape, where the spread of false information can have significant consequences.

Additionally, the flexibility of LLMs allows for the incorporation of advanced techniques such as prompt engineering, retrieval-augmented generation, and fine-tuning. These techniques can further enhance the performance of LLMs in media texts analysis tasks by adapting them to specific contexts and improving their ability to generate accurate and contextually relevant outputs.

## 3.2 Enhancing Media Texts Analysis with Retrieval-Augmented Generation Techniques

RAG is a versatile fine-tuning technique that combines pre-trained parametric and non-parametric memory for language generation (Lewis et al., 2020). This method has demonstrated significant promise by integrating knowledge from external databases, leading to enhanced accuracy and credibility of models, especially for tasks that require substantial domain knowledge (Gao et al., 2023). It also facilitates continuous knowledge updates and the incorporation of domain-specific information.

In the context of media txts analysis, the RAG technique can provide a powerful tool for improving various tasks. By merging the strengths of retrieval-based and generation-based models, RAG enhances the accuracy and contextual relevance of media analysis. This hybrid approach leverages external knowledge sources such as knowledge graphs, background knowledge databases, and real-time information feeds to provide contextually rich and precise analyses. RAG operates by first retrieving

uključuje kategorizaciju članaka, objava i drugih medijskih sadržaja u relevantne kategorije kao što su teme, sentiment ili relevantnost za određene događaje ili entitete. LLM-ovi, poput BERT-a i GPT-3, pokazali su se dobri u ovom zadatku iskorištavanjem svog dubokog razumijevanja semantike i konteksta, koje stječu kroz opsežan pred-trening na raznolikim tekstualnim korpusima. To im omogućuje prepoznavanje razlika u korištenju jezika, pozitivnog, negativnog i neutralnog sentimenata, prepoznavanje nijansiranih tema unutar složenih tekstova i sl. Nadalje, LLM-ovi mogu rukovati varijabilnošću i bogatstvom medijskog sadržaja, uključujući žargon, idiome i različite stilove pisanja, osiguravajući točnu klasifikaciju kroz različite medijske izvore. Njihova sposobnost učenja i prilagodbe novim podacima kroz fino podešavanje čini ih posebno vještima u evoluciji s dinamičnom prirodom medijskih narativa. Integracijom LLM-ova u sustave praćenja medija, organizacije mogu postići preciznije i dublje klasifikacije, olakšavajući bolje praćenje trendova, javnog mišljenja i novih pitanja, što u konačnici omogućuje informiranije donošenje odluka i stratešku komunikaciju.

Klasifikacija sentimenta poboljšana je nijansiranim razumijevanjem jezika koje posjeduju LLM-ovi. Ovi modeli mogu s velikom preciznošću odrediti sentiment izražen u medijskim tekstovima - bilo pozitivan, negativan ili neutralan. To omogućuje detaljniju analizu javnog mišljenja i emocionalnog tona medijskog teksta, što je ključno za upravljanje reputacijom i komunikaciju u kriznim situacijama. Otkrivanje dezinformacija također ima velike koristi od naprednih sposobnosti prepoznavanja obrazaca LLM-ova. Analizom teksta za nedosljednosti, činjenične netočnosti i druge indikatore lažnih informacija, LLM-ovi mogu pomoći u identificiranju i označavanju potencijalnih dezinformacija u medijskom sadržaju. To je posebno važno u današnjem digitalnom okruženju, gdje širenje lažnih informacija može imati negativne posljedice.

Osim toga, fleksibilnost LLM-ova omogućuje primjenu naprednih tehnika poput inženjeringa upita, generiranja s poboljšanjem pretrage i finog podešavanja. Ove tehnike mogu dodatno poboljšati performanse LLM-ova u zadacima analize medijskih tekstova prilagođavanjem specifičnim kontekstima i poboljšanjem njihove sposobnosti generiranja točnih i kontekstualno relevantnih rezultata.

## 3.2 Unaprijeđivanje analize medijskih tekstova primjenom RAG tehnika

RAG je tehnika za optimizaciju LLM-ova koja kombinira unaprijed treniranu parametrijsku i neparametrijsku memoriju za generiranje jezika (Lewis et al., 2020). Ova tehnika je pokazala

relevant information from external sources before generating an analysis or summary of the media content. This retrieval step is crucial because it allows the model to access and incorporate up-to-date and domain-specific knowledge, enhancing the quality and accuracy of the output. By integrating this external knowledge, RAG can produce insights that are not only linguistically accurate but also contextually appropriate and informative.

One of the key benefits of RAG in media texts analysis tasks is its use of knowledge graphs. Knowledge graphs, such as Google Knowledge Graph, WordNet, or DBPedia, provide structured information about entities and their relationships. By leveraging these graphs, RAG can better understand the context and meaning of complex terms or concepts within media texts. For example, when monitoring news articles, the model can retrieve relevant background information and explanations from a knowledge graph, ensuring that the analysis remains accurate and informative. Background knowledge databases play a similar role in enhancing media texts analysis. These databases contain extensive information on various topics, which the model can use to generate more comprehensive and context-aware analyses. For instance, when monitoring reports on geopolitical events, the model can access background information about specific regions or historical contexts, allowing it to provide analyses that maintain the richness and depth of the original content.

Dictionaries and thesauri are also valuable resources in the RAG framework. They offer synonyms, definitions, and usage examples that the model can use to clarify complex terms and ensure that the analysis is both accurate and accessible. This capability is particularly useful for maintaining the readability of the content while ensuring that the insights are still precise and contextually appropriate. For example, technical jargon in economic media texts can be clarified with simpler terms, making the analysis more understandable to a broader audience.

## 3.3 Fine-Tuning Techniques of LLMs for Media Texts Analysis Tasks

Large language models can be fine-tuned for various tasks related to media texts analysis. Fine-tuning these models allows them to adapt specifically to the intricacies of media content, improving their performance on tasks such as sentiment analysis, named entity recognition, and disinformation detection. By leveraging their pre-trained knowledge, LLMs can be tailored to recognize and classify specific topics, identify key entities like people and organizations, and even detect changes in public sentiment over time. For instance, in sentiment analysis, a fine-tuned LLM can accurately gauge the public's reaction to a news event, providing valuable

značajan potencijal integriranjem znanja iz vanjskih baza podataka, što dovodi do poboljšane točnosti i vjerodostojnosti modela, posebno za zadatke koji zahtijevaju značajno domensko znanje (Gao et al., 2023). Također olakšava kontinuirano ažuriranje znanja i uključivanje informacija specifičnih za određenu domenu.

U kontekstu analize medijskih tekstova, tehnika RAG može pružiti moćan alat za poboljšanje raznih zadataka. Spajanjem modela temeljenih na pretraživanju i generiranju teksta, RAG povećava točnost i kontekstualnu relevantnost analize medija. Ovaj hibridni pristup koristi vanjske izvore znanja kao što su grafovi znanja, baze podataka s pozadinskim znanjem i izvori informacija u stvarnom vremenu kako bi pružio kontekstualno bogate i precizne analize. RAG funkcionira tako da prvo pretražuje relevantne informacije iz vanjskih izvora prije nego što provede analizu ili generira sažetak medijskog sadržaja. Ovaj korak pretraživanja je ključan jer omogućuje modelu pristup i uključivanje ažuriranih i specifičnih znanja, čime se poboljšava kvaliteta i točnost rezultata. Integriranjem vanjskog znanja, RAG može proizvesti uvide koji su ne samo jezično točni već i kontekstualno prikladni i informativni.

Jedna od ključnih prednosti RAG-a u analizi medijskih tekstova je njegova uporaba grafova znanja. Grafovi znanja, poput Google Knowledge grafa, WordNet-a ili DBPedije, pružaju strukturirane informacije o entitetima i njihovim odnosima. Korištenjem ovih grafova, RAG može bolje razumjeti kontekst i značenje složenih pojmova ili koncepata unutar medijskih tekstova. Na primjer, prilikom praćenja novinskih članaka, model može preuzeti relevantne pozadinske informacije i objašnjenja iz grafova znanja, osiguravajući da analiza ostane točna i informativna. Baze podataka s pozadinskim znanjem igraju sličnu ulogu u poboljšanju praćenja medija. Ove baze podataka sadrže opsežne informacije o raznim temama koje model može koristiti za generiranje sveobuhvatnijih i kontekstualno svjesnijih analiza. Na primjer, prilikom praćenja izvještaja o geopolitičkim događajima, model može pristupiti pozadinskim informacijama o specifičnim regijama ili povijesnom kontekstu, omogućujući analize koje zadržavaju bogatstvo i dubinu izvornog sadržaja.

Rječnici i tezaurusi također su vrijedni resursi u okviru RAG-a. Oni nude sinonime, definicije i primjere upotrebe koje model može koristiti za pojašnjavanje složenih pojmova i osiguravanje da analiza bude i točna i dostupna. Ova sposobnost je posebno korisna za održavanje čitljivosti sadržaja, dok se osigurava da su uvjeti još uvijek precizni i kontekstualno prikladni. Na primjer, tehnički žargon u ekonomskim medijskim tekstovima može se pojasniti jednostavnijim pojmovima, čineći analizu razumljivijom širem krugu publike.

insights for strategic communication. Similarly, in named entity recognition, these models can consistently identify and categorize entities, even in diverse and unstructured media sources.

Furthermore, fine-tuned LLMs can enhance disinformation detection by identifying subtle cues and inconsistencies that signal false information. This tailored approach not only increases the accuracy and relevance of media texts analysis but also enables organizations to respond swiftly and effectively to the evolving media landscape.

## 3.4. An Overview of the NLP Methods for Media Texts Analysis

Integrating NLP methods based on GenAI into a comprehensive system for media text analysis offers a robust framework for processing and understanding vast amounts of media content. By combining various NLP techniques, such as keyword extraction, NER, sentiment analysis, topic modeling, hate speech detection, and disinformation detection, a holistic approach to media texts anaylsis can be achieved. Each method plays a critical role as it is shown in Table 1. These technologies are particularly powerful when applied to the analysis of trends and changes over time, allowing stakeholders to monitor the evolution of public discourse and sentiment across different media platforms.

**Table 1.** Description of NLP tasks and methods for media texts analysis

| **NLP task** | **Media texts analysis - application description** |
|---|---|
| Keyword Extraction | Extraction of key terms that highlight the central topics in media content. **Example:** Fine-tuning LLMs for KE in online media texts |
| Topic Modelling | Detection of main topics that are in the focus of media. **Example:** Fine-tuning LLMs for topic modeling in online media texts. This method can be combined with sentiment analysis to assess sentiment related to specific topics and can include the analysis of trends and changes over time. |
| NER | Identification of specific entities such as people, organizations, and |

## 3.3 Tehnike finog podešavanja velikih jezičnih modela za analizu medijskih tekstova

Veliki jezični modeli mogu se fino podešavati za razne zadatke povezane s analizom medijskih tekstova. Fino podešavanje LLM-ova omogućuje prilagodbu specifičnostima medijskog sadržaja, poboljšavajući njihovu izvedbu u zadacima kao što su analiza sentimenta, prepoznavanje imenovanih entiteta i otkrivanje dezinformacija. Korištenjem svog pred-treninga, LLM-ovi mogu se prilagoditi za prepoznavanje i klasifikaciju specifičnih tema, identifikaciju ključnih entiteta kao što su ljudi i organizacije, pa čak i detekciju promjena u javnom sentimentu tijekom vremena. Na primjer, u analizi sentimenta, fino podešavan LLM može točno procijeniti reakciju javnosti na neki događaj u vijestima, pružajući vrijedne uvide za stratešku komunikaciju. Slično, u prepoznavanju imenovanih entiteta, ovi modeli mogu dosljedno identificirati i kategorizirati entitete, čak i u raznolikim i nestrukturiranim medijskim izvorima.

Nadalje, fino podešavani LLM-ovi mogu poboljšati otkrivanje dezinformacija identificiranjem suptilnih znakova i nedosljednosti koje ukazuju na lažne informacije. Ovaj prilagođeni pristup ne samo da povećava točnost i relevantnost analize medijskih tekstova, već također omogućuje organizacijama da brzo i učinkovito reagiraju na promjenjivost medija.

## 3.4. Tehnike finog podešavanja velikih jezičnih modela za analizu medijskih tekstova

Integracija NLP metoda temeljenih na GenAI u sveobuhvatan sustav za analizu medijskih tekstova nudi snažan okvir za obradu i razumijevanje velike količine medijskog sadržaja. Kombiniranjem različitih NLP tehnika, poput ekstrakcije ključnih riječi, NER, analize sentimenta, modeliranja tema, otkrivanja govora mržnje i otkrivanja dezinformacija, može se postići cjelovit pristup analizi medijskih tekstova. Svaka metoda ima važnu ulogu, što je prikazano u Tablici 1. Ove tehnologije su posebno moćne kada se primjenjuju na analizu trendova i promjena tijekom vremena, omogućujući dionicima praćenje evolucije javnog diskursa i sentimenta na različitim medijskim platformama.

**Tablica 1.** Opis NLP zadataka i metoda za analizu medijskih tektova

| **NLP zadatak** | **Analiza medijskih tekstova – opis primjene** |
|---|---|

| | |
|---|---|
| | locations mentioned in media content. **Example:** Applying LLMs for the identification of crucial persons, locations, institutions, and other important entities appearing in online media. This method can be combined with sentiment analysis to assess sentiment related to specific entities and can include the analysis of trends and changes over time. |
| Sentiment Analysis | Evaluation of the emotional tone expressed in media content, determining whether it is positive, negative, or neutral. **Example:** Prompt engineering with LLMs in combination with RAG, incorporating external knowledge about sentiment. |
| Hate Speech Recognition | Identification of offensive or harmful language in media content that targets individuals or groups based on attributes like race, religion, or gender. **Example:** Fine-tuning LLMs for the task of media text classification into two or more classes, including one that contains hate speech. |
| Disinformation Detection | Detection of false or misleading information in media content that is intentionally spread to deceive or misinform the audience. **Example:** Fine-tunning LLMs in combination with RAG, incorporating external knowledge about the domain of interest. |

By fine-tuning LLMs for these specific tasks, the system can efficiently handle diverse and unstructured media data, providing deeper insights and more accurate classifications. Moreover, integrating external knowledge sources through RAG

| | |
|---|---|
| Ekstrakcija ključnih riječi | Ekstrakcija ključnih pojmova koji ističu središnje teme u medijskom sadržaju. **Primjer:** Fino podešavanje LLM-ova za ekstrakciju ključnih pojmova u online medijskim tekstovima. |
| Modeliranje tema | Otkrivanje glavnih tema koje su u fokusu medija. **Primjer:** Fino podešavanje LLM-ova za modeliranje tema u online medijskim tekstovima. Ova metoda se može kombinirati s analizom sentimenta kako bi se odredio sentiment povezan s određenim temama te može uključivati analizu trendova i promjena tijekom vremena. |
| NER | Identifikacija specifičnih entiteta kao što su ljudi, organizacije i lokacije spomenuti u medijskom sadržaju. **Primjer:** Primjena LLM-ova za identifikaciju ključnih osoba, lokacija, institucija i drugih važnih entiteta koji se pojavljuju u online medijima. Ova metoda se može kombinirati s analizom sentimenta kako bi se procijenio sentiment povezan s određenim entitetima te može uključivati analizu trendova i promjena tijekom vremena. |
| Analiza sentimenta | Analiza sentimenta izraženog u medijskom sadržaju, određivanje je li pozitivan, negativan ili neutralan. **Primjer:** Prompt inženjering upita s LLM-ovima u kombinaciji s RAG-om, uz uključivanje vanjskog znanja o sentimentu. |
| Prepoznavanje govora mržnje | Identifikacija govora mržnje u medijskom sadržaju koji cilja pojedince ili skupine na |

and employing techniques like prompt engineering can further enhance the system's capability, ensuring that it remains up-to-date and contextually relevant. This comprehensive approach not only improves the accuracy and efficiency of media analysis but also enables stakeholders to respond quickly and effectively to emerging trends and issues in the media landscape.

# 4 Conclusion

This paper has focused on the methods for media texts analysis, particularly those based on GenAI. GenAI-based methods offer significant improvements in accuracy, efficiency, and contextual understanding over traditional NLP approaches, enabling more effective analysis and interpretation of media texts.

The main contribution of this paper is an overview of NLP methods based on GenAI and their potential in the field of media text analysis. This is the first step toward the definition and implementation of a comprehensive media monitoring system.

The integration of large language models into systems for media texts analysis and media monitoring marks a significant leap forward in the ability to analyze and understand media content. These models excel at performing complex NLP tasks with high precision and adaptability, making them indispensable tools for modern media monitoring and media texts analysis. Their ability to handle diverse and unstructured media texts, capture intricate linguistic patterns, and provide nuanced insights sets them apart from classical methods.

Furthermore, the application of RAG to media texts analysis represents a substantial advancement, enabling more accurate, informed, and contextually relevant analyses. By combining the strengths of retrieval and generation capabilities with extensive external knowledge, RAG enhances the overall quality and utility of media txts analysis efforts. This hybrid approach ensures that the models are not only current with real-time information but also enriched with domain-specific knowledge, thereby improving their effectiveness in detecting misinformation, conducting sentiment analysis, and extracting key entities and topics.

Future research in this direction will focus on developing and refining GenAI-based systems for media texts analysis. This includes enhancing the integration of external knowledge sources, improving fine-tuning techniques, and optimizing model interactions through advanced methods such as prompt engineering. By continuing to innovate and apply GenAI technologies, we can further elevate the capabilities of media texts analysis systems, ensuring they remain robust, accurate, and contextually aware in an ever-evolving digital landscape.

| | |
|---|---|
| | temelju atributa poput rase, religije ili spola. **Primjer:** Fino podešavanje LLM-ova za zadatak klasifikacije medijskih tekstova u dvije ili više kategorija, uključujući jednu koja sadrži govor mržnje. |
| Detekcija dezinformacija | Otkrivanje lažnih ili obmanjujućih informacija u medijskom sadržaju koje se namjerno šire kako bi se prevarilo ili dezinformiralo publiku. **Primjer:** Fino podešavanje LLM-ova u kombinaciji s RAG-om, uz uključivanje vanjskog znanja o domeni interesa. |

Finim podešavanjem LLM-ova za ove specifične zadatke, sustav može učinkovito upravljati raznolikim i nestrukturiranim medijskim podacima, pružajući dublje uvide i točnije klasifikacije. Nadalje, integracija vanjskih izvora znanja putem RAG-a i primjena tehnika poput inženjeringa upita mogu dodatno poboljšati sposobnosti sustava, osiguravajući da ostane ažuran i kontekstualno relevantan. Ovaj sveobuhvatan pristup ne samo da poboljšava točnost i učinkovitost analize medija, već također omogućuje dionicima da brzo i učinkovito reagiraju na nove trendove i pitanja u medijskom okruženju.

# 4 Zaključak

Ovaj istražuje metode analize medijskih tekstova, posebno one temeljene na gnerativnoj umejtnoj inteligenciji. Metode temeljene na GenAI nude značajna poboljšanja u točnosti, učinkovitosti i kontekstualnom razumijevanju u usporedbi s tradicionalnim pristupima NLP-a, omogućujući učinkovitiju analizu i interpretaciju medijskih tekstova.

Glavni doprinos ovog rada je pregled NLP metoda temeljenih na generativnoj umjetnoj inteligenciji i njihovog potencijala u području analize medijskih tekstova. Ovo je prvi korak prema definiranju i implementaciji sveobuhvatnog sustava za praćenje medija.

Integracija LLM-ova u sustave za analizu medijskih tekstova i praćenje medija predstavlja značajan iskorak u sposobnosti analize i razumijevanja medijskog sadržaja. Veliki jezični modeli izvrsno obavljaju složene NLP zadatke s visokom preciznošću i prilagodljivošću, čineći ih

## Acknowledgments

This work has been supported by the UNIRI project uniri-iskusni-drustv-23-95.

## References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Babić, K., Martinčić-Ipšić, S., & Meštrović, A. (2020). Survey of neural text representation models. Information, 11(11), 511.

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. Applied Sciences, 11(21), 10442.

Balkus, S. V., & Yan, D. (2022). Improving short text classification with augmented data using GPT-3. Natural Language Engineering, 1-30.

Barić, A., Majer, L., Dukić, D., Grbeša-Zenzerović, M., & Šnajder, J. (2023, May). Target Two Birds With One SToNe: Entity-Level Sentiment and Tone Analysis in Croatian News Headlines. In Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023) (pp. 78-85).

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.

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. In The Covid-19 pandemic as a challenge for media and communication studies (pp. 78-92). Routledge.

Bhoi, Ashutosh, Sthita Pragyan Pujari, and Rakesh Chandra Balabantaray. "A deep learning-based social media text analysis framework for disaster resource management." Social Network Analysis and Mining 10 (2020): 1-14.

Buhin Pandur, M., Dobša, J., Beliga, S., & Meštrović, A. (2021). Topic modelling and sentiment analysis of COVID-19 related news on Croatian Internet portal. In SiKDD Conference on Data Mining and Data Warehouses. Slovenian

neophodnim alatima za moderno praćenje i analizu medija. Njihova sposobnost obrade raznolikih i nestrukturiranih medijskih tekstova, hvatanje složenih jezičnih obrazaca i pružanje nijansiranih uvida izdvaja ih od klasičnih metoda.

Nadalje, primjena RAG tehnika na sustave za analizu medijskih tekstova predstavlja značajan napredak, omogućujući točnije, informiranije i kontekstualno relevantne analize. Kombinirajući sposobnosti pretraživanja i generiranja s opsežnim vanjskim znanjem, RAG poboljšava ukupnu kvalitetu i korisnost postupaka analize medijskih tekstova. Ovaj hibridni pristup osigurava da su modeli ne samo u skladu s informacijama u stvarnom vremenu, već i obogaćeni znanjem specifičnim za određenu domenu, čime se poboljšava njihova učinkovitost u otkrivanju dezinformacija, provođenju analize sentimenta i izdvajanja ključnih entiteta i tema.

Buduća istraživanja u ovom smjeru fokusirat će se na razvoj i usavršavanje sustava za analizu medijskih tekstova temeljenih na generativnoj umjetnoj inteligenciji. To uključuje poboljšanje integracije vanjskih izvora znanja, usavršavanje tehnika finog podešavanja i optimizaciju interakcija modela kroz napredne metode poput inženjeringa upita. Nastavljajući inovirati i primjenjivati tehnologije GenAI, možemo dodatno unaprijediti sposobnosti sustava za analizu medijskoh tekstova, osiguravajući da ostanu robusni, točni i kontekstualno svjesni u stalno promjenjivoj digitalnoj okolini.

## Zahvale

Ovo istraživanje i rad podržan je UNIRI projektom uniri-iskusni-drustv-23-95.

## Reference

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Babić, K., Martinčić-Ipšić, S., & Meštrović, A. (2020). Survey of neural text representation models. Information, 11(11), 511.

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. Applied Sciences, 11(21), 10442.

Balkus, S. V., & Yan, D. (2022). Improving short text classification with augmented data using GPT-3. Natural Language Engineering, 1-30.

KDD Conference on Data Mining and Data Warehouses.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Farzindar, A., Inkpen, D., & Hirst, G. (2015). Natural language processing for social media. San Rafael: Morgan & Claypool.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.

Germann, U., Liepins, R., Gosko, D., & Barzdins, G. (2018, July). SUMMA: Integrating multiple NLP technologies into an open-source platform for multilingual media monitoring. In ACL 2018 Workshop for Natural Language Processing Open Source Software (pp. 47-51). Association for Computational Linguistics.

Koloski, B, Stepišnik Perdih, T., Robnik-Šikonja, M., Pollak, S., and Škrlj, B. 2022. "Knowledge graph informed fake news classification via heterogeneous representation ensembles." Neurocomputing.

Korenčić, D., Ristov, S., & Šnajder, J. (2018). Document-based topic coherence measures for news media text. Expert systems with Applications, 114, 357-373.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.

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.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

Srikanth M., Liu A., Adams-Cohen N., Cao J., Alvarez R. M., and Anandkumar A.. 2021. Dynamic Social Media Monitoring for Fast-Evolving Online Discussions. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21). Association for Computing Machinery, New York, NY, USA, 3576–3584. https://doi.org/10.1145/3447548.3467171

Sufi, F. (2023). Social media analytics on Russia–Ukraine cyber war with natural language

Barić, A., Majer, L., Dukić, D., Grbeša-Zenzerović, M., & Šnajder, J. (2023, May). Target Two Birds With One SToNe: Entity-Level Sentiment and Tone Analysis in Croatian News Headlines. In Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023) (pp. 78-85).

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.

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. In The Covid-19 pandemic as a challenge for media and communication studies (pp. 78-92). Routledge.

Bhoi, Ashutosh, Sthita Pragyan Pujari, and Rakesh Chandra Balabantaray. "A deep learning-based social media text analysis framework for disaster resource management." Social Network Analysis and Mining 10 (2020): 1-14.

Buhin Pandur, M., Dobša, J., Beliga, S., & Meštrović, A. (2021). Topic modelling and sentiment analysis of COVID-19 related news on Croatian Internet portal. In SiKDD Conference on Data Mining and Data Warehouses. Slovenian KDD Conference on Data Mining and Data Warehouses.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Farzindar, A., Inkpen, D., & Hirst, G. (2015). Natural language processing for social media. San Rafael: Morgan & Claypool.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.

Germann, U., Liepins, R., Gosko, D., & Barzdins, G. (2018, July). SUMMA: Integrating multiple NLP technologies into an open-source platform for multilingual media monitoring. In ACL 2018 Workshop for Natural Language Processing Open Source Software (pp. 47-51). Association for Computational Linguistics.

Koloski, B, Stepišnik Perdih, T., Robnik-Šikonja, M., Pollak, S., and Škrlj, B. 2022. "Knowledge graph informed fake news classification via

processing: Perspectives and challenges. Information, 14(9), 485.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification?. In Chinese computational linguistics: 18th China national conference, CCL 2019, Kunming, China, October 18–20, 2019, proceedings 18 (pp. 194-206). Springer International Publishing.

Swati, S., Mladenić, D., & Grobelnik, M. (2023). An Inferential Commonsense-Driven Framework for Predicting Political Bias in News Headlines. IEEE Access.

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, Brno, Czech Republic, September 8–11, 2020, Proceedings 23 (pp. 104-111). Springer International Publishing.

Vrbanec, T., & Meštrović, A. (2023). Comparison study of unsupervised paraphrase detection: Deep learning—The key for semantic similarity detection. Expert systems, 40(9), e13386.

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.

Yang, J., Wang, M., Zhou, H., Zhao, C., Zhang, W., Yu, Y., & Li, L. (2020, April). Towards making the most of BERT in neural machine translation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 05, pp. 9378-9385).

Zhong N., Schweidel D. A. (2020) Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model. Marketing Science 39(4):827-846.

Zhu, J., Xia, Y., Wu, L., He, D., Qin, T., Zhou, W., ... & Liu, T. Y. (2020). Incorporating BERT into neural machine translation. arXiv preprint arXiv:2002.06823.

heterogeneous representation ensembles." Neurocomputing.

Korenčić, D., Ristov, S., & Šnajder, J. (2018). Document-based topic coherence measures for news media text. Expert systems with Applications, 114, 357-373.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.

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.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

Srikanth M., Liu A., Adams-Cohen N., Cao J., Alvarez R. M., and Anandkumar A.. 2021. Dynamic Social Media Monitoring for Fast-Evolving Online Discussions. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21). Association for Computing Machinery, New York, NY, USA, 3576–3584. https://doi.org/10.1145/3447548.3467171

Sufi, F. (2023). Social media analytics on Russia–Ukraine cyber war with natural language processing: Perspectives and challenges. Information, 14(9), 485.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification?. In Chinese computational linguistics: 18th China national conference, CCL 2019, Kunming, China, October 18–20, 2019, proceedings 18 (pp. 194-206). Springer International Publishing.

Swati, S., Mladenić, D., & Grobelnik, M. (2023). An Inferential Commonsense-Driven Framework for Predicting Political Bias in News Headlines. IEEE Access.

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, Brno, Czech Republic, September 8–11, 2020, Proceedings 23 (pp. 104-111). Springer International Publishing.

Vrbanec, T., & Meštrović, A. (2023). Comparison study of unsupervised paraphrase detection: Deep learning—The key for semantic

similarity detection. Expert systems, 40(9), e13386.

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.

Yang, J., Wang, M., Zhou, H., Zhao, C., Zhang, W., Yu, Y., & Li, L. (2020, April). Towards making the most of BERT in neural machine translation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 05, pp. 9378-9385).

Zhong N., Schweidel D. A. (2020) Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model. Marketing Science 39(4):827-846.

Zhu, J., Xia, Y., Wu, L., He, D., Qin, T., Zhou, W., ... & Liu, T. Y. (2020). Incorporating BERT into neural machine translation. arXiv preprint arXiv:2002.06823

# Text Analysis Methods for Identifying and Recognizing Gender Bias, Inequality, and Violence in Language: An Overview

# Metode analize teksta za identifikaciju i prepoznavanje rodne pristranosti, nejednakosti i nasilja u jeziku: Pregled

**Lucia Načinović Prskalo**
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, 51000 Rijeka
lnacinovic@uniri.hr

**Abstract**. *This paper provides an overview of text analysis methods used to identify and uncover gender bias, inequality and violence in language. By examining text analysis methods and techniques such as corpus linguistics, sentiment analysis, content analysis, discourse analysis, network analysis and machine translation, the study shows how these methods can uncover patterns and tendencies in texts that perpetuate gender bias, stereotypes and violence. The findings highlight the importance of using text analysis tools to promote gender equality and develop more inclusive language practices.*

**Keywords.** Gender bias, gender inequality, gender violence, text analysis methods

***Sažetak***. *U ovom radu dan je pregled metoda analize teksta koje se koriste za identificiranje i otkrivanje rodne pristranosti, nejednakosti i nasilja u jeziku. Korištenje metoda i tehnika analize teksta kao što su korpusna lingvistika, analiza sentimenta, analiza sadržaja, analiza diskursa, analiza mreža i strojno prevođenje, može otkriti obrasce i tendencije u tekstovima koje perpetuiraju rodnu pristranost, stereotipe i nasilje. Naglašava se važnost korištenja alata za analizu teksta za promicanje rodne jednakosti i razvoj inkluzivnijih jezičnih praksi.*

**Ključne riječi.** Rodna pristranost, rodna nejednakost, rodno nasilje, metode analize teksta

## 1 Introduction

Gender equality refers to the idea that people, regardless of their gender, should have the same rights, opportunities and access to resources (United Nations, 2020). This means that men and women, as well as nonbinary individuals, should have equal opportunities for education, employment, participation in the political process, healthcare, access to resources and other aspects of society (World Economic Forum, 2023). The concept of gender equality assumes that gender differences should not affect the rights and opportunities of individuals in society (European Commission, 2021). Gender equality is important to create a fairer and more inclusive society that enables everyone to reach their full potential (UN Women, 2019). There are various ways in which individuals and society as a whole can contribute to gender equality, including supporting gender equality in policy, eliminating

## 1 Uvod

Rodna jednakost odnosi se na ideju da bi ljudi, bez obzira na njihov spol, trebali imati ista prava, mogućnosti i pristup resursima (United Nations, 2020). To znači da bi muškarci i žene, kao i nebinarne osobe, trebali imati jednake mogućnosti za obrazovanje, zapošljavanje, sudjelovanje u političkom procesu, zdravstvenu skrb, pristup resursima i drugim aspektima društva (World Economic Forum, 2023). Koncept rodne jednakosti pretpostavlja da rodne razlike ne bi trebale utjecati na prava i mogućnosti pojedinaca u društvu (European Commission, 2021). Rodna jednakost je važna za stvaranje pravednijeg i inkluzivnijeg društva koje omogućuje svima da ostvare svoj puni potencijal (UN Women, 2019). Postoje različiti načini na koje pojedinci i društvo u cjelini mogu doprinijeti rodnoj jednakosti, uključujući podržavanje rodne jednakosti u politici, eliminiranje rodno uvjetovanih stereotipa

gender-based stereotypes and promoting and supporting women in leadership and decision-making positions (Kabeer, 2015).

Language is an important cultural factor that shapes our social norms and values (Lakoff, 2004). Gendered language can contribute to inequality and discrimination based on gender (Cameron, 2005). Analysing language from a gender equality perspective can help combat gender discrimination and inequality (Holmes, 2006). In many languages, including Croatian, there are gender-specific word forms that often reflect traditional gender roles and stereotypes. In Croatian, for example, the masculine gender is often used to describe jobs and professions such as "doctor" and "driver"," while the feminine gender is used to describe care roles, such as "nurse" and "housewife".

Language analysis from an equality perspective is about recognizing and weakening such gender-specific word forms in order to achieve gender equality in language. In addition, language analysis can help to identify and eliminate gender-based stereotypes in the language (Mills, 1995). In the Croatian language, for example, the word "weak" is often used as a synonym for femininity, while the word "strong" is used as a synonym for masculinity. Such word usage can perpetuate and support traditional gender roles and stereotypes. Language analysis can help to identify these stereotypes and replace them with more neutral expressions that do not support gender-specific roles and stereotypes.

Analysing violence in texts from a gender equality perspective can also be useful in the fight against gender-based violence. Violence in language can manifest itself in various ways, including the use of words that insult and discriminate against individuals based on their gender (Kimmel, 2000). This type of analysis can help identify forms of violence that often occur silently, such as emotional and verbal abuse. For example, derogatory terms towards women are often used in certain text categories, and their use contributes to a culture of tolerance towards violence against women (European Parliament, 2018).

The main goal of this paper is to provide an overview of text analysis methods used to identify and recognize gender bias, inequality, and violence in language. By exploring different text analysis methods and their applications, this paper aims to highlight the importance of addressing gender issues in language and offer insights into the ways in which these methods can be used to promote gender equality and mitigate discrimination and violence.

The next section discusses the socio-economic context of gender equality in Croatia and Europe, highlighting relevant policy documents and initiatives to promote gender equality. The literature review section provides an overview of the main literature on the relationship between language and gender, including foundational works and recent

te promicanje i podržavanje žena na vodećim pozicijama (Kabeer, 2015).

Jezik je važan kulturni faktor koji oblikuje naše društvene norme i vrijednosti (Lakoff, 2004). Rodno obilježen jezik može pridonijeti nejednakosti i diskriminaciji na temelju spola (Cameron, 2005). Analiziranje jezika iz perspektive rodne jednakosti može pomoći u borbi protiv rodne diskriminacije i nejednakosti (Holmes, 2006). U mnogim jezicima, uključujući hrvatski, postoje rodno specifični oblici riječi koji često odražavaju tradicionalne rodne uloge i stereotipe. U hrvatskom jeziku, na primjer, muški rod se često koristi za opisivanje poslova i profesija kao što su "doktor" i "vozač", dok se ženski rod koristi za opisivanje uloga koje uključuju pružanje njege, kao što su "medicinska sestra" i "domaćica".

Analiza jezika iz perspektive jednakosti odnosi se na prepoznavanje i slabljenje takvih rodno specifičnih oblika riječi kako bi se postigla rodna jednakost u jeziku. Osim toga, analiza jezika može pomoći u identificiranju i eliminiranju rodno uvjetovanih stereotipa u jeziku (Mills, 1995). U hrvatskom jeziku, na primjer, riječ "slab" često se koristi kao sinonim za ženstvenost, dok se riječ "jak" koristi kao sinonim za muškost. Takva upotreba riječi može perpetuirati i održavati tradicionalne rodne uloge i stereotipe. Analiza jezika može pomoći u prepoznavanju tih stereotipa i njihovoj zamjeni neutralnijim izrazima koji ne podržavaju rodno specifične uloge i stereotipe.

Analiziranje nasilja u tekstovima iz perspektive rodne jednakosti može također biti korisno u borbi protiv rodno uvjetovanog nasilja. Nasilje u jeziku može se manifestirati na različite načine, uključujući upotrebu riječi koje vrijeđaju i diskriminiraju pojedince na temelju njihovog spola (Kimmel, 2000). Ova vrsta analize može pomoći u prepoznavanju oblika nasilja koji se često događaju neprimjetno, poput emocionalnog i verbalnog zlostavljanja. Na primjer, pogrdni izrazi prema ženama često se koriste u određenim kategorijama tekstova, a njihova upotreba doprinosi kulturi tolerancije prema nasilju nad ženama (European Parliament, 2018).

Glavni cilj ovog rada je pružiti pregled metoda analize teksta koje se koriste za identificiranje i prepoznavanje rodne pristranosti, nejednakosti i nasilja u jeziku. Proučavanjem različitih metoda analize teksta i njihove primjene, ovaj rad ima za cilj istaknuti važnost rješavanja rodnih pitanja u jeziku te ponuditi uvide u načine na koje se te metode mogu koristiti za promicanje rodne jednakosti i smanjenje diskriminacije i nasilja.

U sljedećoj se sekciji raspravlja o društveno-ekonomskom kontekstu rodne jednakosti u Hrvatskoj i Europi, naglašavajući relevantne političke dokumente i inicijative za promicanje rodne jednakosti. Sekcija s pregledom literature pruža pregled glavne literature o odnosu između jezika i roda, uključujući temeljna djela i nedavne studije

studies that examine gender stereotypes, prejudice and violence in language. The next section introduces and explains different text analysis methods used to examine gender inequality, bias and violence in language. Each method is described with its applications, tools, examples and related research. Finally, the main findings are summarized, the importance of text analysis for the promotion of gender equality is highlighted and suggestions for future research and practical applications are made.

# 2 Socio-economic Framework

In Croatia, gender equality is regulated by law and there are a number of laws promoting gender equality in all areas of life. In practice, however, there are still numerous challenges and problems, such as discrimination against women in the labour market, underrepresentation of women in leadership positions and violence against women. For example, a survey "Women in Business" (Croatian Employers' Association & Deloitte, 2024) showed that women in Croatia still face discrimination in the labour market and are paid less than their male colleagues for the same work.

Although gender equality is recognized as a fundamental human right in Europe, there are significant differences between countries in the realization of this right. According to EU research, women are still paid less than men for the same work and are less represented in leadership positions. For example, according to the 2022 EU Gender Equality Report (European Commission, 2022), women in the EU earn on average 14.1% less than men; in EU national parliaments, women hold an average of 34% of seats, meaning they are underrepresented in politics; women make up only 29.6% of board members of large companies in the EU, and the number of women in leadership positions in companies has stagnated for years.

Similar problems exist in other countries around the world, regardless of the level of development and the social climate. Gender inequality is widespread in many areas and women are often neglected or discriminated against.

## 2.1 Relevant European and National Strategic Documents

At EU level, the EU Gender Equality Strategy (2020-2025) (European Commission, 2020) emphasises the importance of reducing gender inequalities and gender discrimination in all areas of society, including language. This strategy places particular emphasis on the need to ensure equal participation of women and men in education and research. In addition, The Action Plan for Gender Equality under the European Social Fund for the period 2021-2027

koje ispituju rodne stereotipe, predrasude i nasilje u jeziku. Zatim slijedi sekcija koja uvodi i objašnjava različite metode analize teksta koje se koriste za ispitivanje rodne nejednakosti, pristranosti i nasilja u jeziku. Svaka metoda je opisana sa svojim primjenama, alatima, primjerima i relevantnim povezanim istraživanjima. Na kraju su u zaključku sažeti glavni nalazi, istaknuta je važnost analize teksta za promicanje rodne jednakosti te su dane sugestije za buduća istraživanja i praktične primjene.

# 2 Društveno-ekonomski okvir

U Hrvatskoj je rodna jednakost regulirana zakonom i postoji niz zakona koji promiču rodnu jednakost u svim područjima života. U praksi, međutim, još uvijek postoje brojni izazovi i problemi, poput diskriminacije žena na tržištu rada, nedovoljne zastupljenosti žena na rukovodećim pozicijama i nasilja nad ženama. Na primjer, istraživanje "Women in Business" (Croatian Employers' Association & Deloitte, 2024) pokazalo je da se žene u Hrvatskoj još uvijek suočavaju s diskriminacijom na tržištu rada i da su manje plaćene od svojih muških kolega za isti posao.

Iako je rodna jednakost prepoznata kao temeljno ljudsko pravo u Europi, postoje značajne razlike među zemljama u ostvarivanju ovog prava. Prema istraživanjima EU, žene su i dalje manje plaćene od muškaraca za isti posao i manje su zastupljene na rukovodećim pozicijama. Na primjer, prema istraživanju iz 2022. koje je opisano u EU Gender Equality Report (European Commission, 2022), žene u EU zarađuju u prosjeku 14,1% manje od muškaraca; u EU nacionalnim parlamentima žene drže prosječno 34% mjesta, što znači da su nedovoljno zastupljene u politici; žene čine samo 29,6% članova upravnih odbora velikih tvrtki u EU, a broj žena na rukovodećim pozicijama u tvrtkama stagnira godinama.

Slični problemi postoje i u drugim zemljama svijeta, bez obzira na razinu razvoja i društvenu klimu. Rodna nejednakost je raširena u mnogim područjima i žene su često zanemarene ili diskriminirane.

## 2.1 Relevantni europski i nacionalni strateški dokumenti

Na razini EU, EU Gender Equality Strategy (2020-2025) (European Commission, 2020) naglašava važnost smanjenja rodnih nejednakosti i rodne diskriminacije u svim područjima društva, uključujući jezik. Ova strategija stavlja poseban naglasak na potrebu osiguravanja ravnopravnog sudjelovanja žena i muškaraca u obrazovanju i istraživanju. Zatim, Action Plan for Gender Equality under the European Social Fund za razdoblje 2021-2027 (European Commission, 2021) ima za cilj

(European Commission, 2021), aims to ensure greater gender equality in the labour market and in society in general.

At the national level, Croatia has a National Strategy for Gender Equality for the period 2023-2027 (Government of Croatia, 2023), which aims to improve gender equality in all areas of society, including education and research. Analysing language with the aim of identifying sexist and gender-specific language and creating guidelines for the use of gender-sensitive language in academic and scientific communication can contribute to these strategies. It also supports the National Curriculum Reform Strategy for Education for the period 2022-2027 (Government of Croatia, 2022), which focuses on creating an inclusive and gender-sensitive education system.

The University of Rijeka has adopted a Gender Equality Plan 2021-2025 (University of Rijeka, 2021), which aims to promote gender equality at the university. The plan consists of a series of activities and measures aimed at various areas, including the introduction of gender-sensitive language in everyday language use at the university, the promotion of research in the field of gender equality and the integration of a gender perspective into teaching programmes, as well as the promotion of gender equality and the prevention of gender-based violence among students and university staff.

# 3 Literature Review

Text analysis is used in numerous applications, often in combination with machine learning and deep learning techniques. These combinations have led to significant advances in areas such as emotion recognition and mental health, as recent studies show. For example, the work (Jonnalagadda et al., 2023) shows how machine learning can be used effectively to assess emotional states from text. Similarly, the study (Rustagi et al., 2022) explores the use of deep learning methods to improve the detection of sarcasm in online reviews by incorporating contextual cues such as emojis and ratings.

When it comes to analysing texts and language in the context of gender equality, there are numerous studies that show how language can be a powerful tool to perpetuate existing gender stereotypes and inequalities.

Key publications in the field of language and gender equality research include: "Language and Woman's Place" by Robin Lakoff (1973). This book is considered one of the most influential publications on the relationship between language and gender. Lakoff argues that there are certain language patterns that contribute to the unequal position of women in society. "Language and Gender: An Introduction" by Mary Talbot (1998) is a book that provides a detailed

osigurati veću rodnu ravnopravnost na tržištu rada i u društvu općenito.

Na nacionalnoj razini, Hrvatska ima Nacionalnu strategiju za ravnopravnost spolova za razdoblje 2023-2027 (Government of Croatia, 2023), koja ima za cilj poboljšati rodnu ravnopravnost u svim područjima društva, uključujući obrazovanje i istraživanje. Analiza jezika s ciljem identificiranja seksističkog i rodno specifičnog jezika te stvaranja smjernica za upotrebu rodno osjetljivog jezika u akademskoj i znanstvenoj komunikaciji može doprinijeti ovim strategijama. Također podržava Strategiju kurikularne reforme za obrazovanje za razdoblje 2022-2027 (Government of Croatia, 2022), koja se fokusira na stvaranje uključivog i rodno osjetljivog obrazovnog sustava.

Sveučilište u Rijeci usvojilo je Plan rodne ravnopravnosti 2021-2025 (University of Rijeka, 2021), koji ima za cilj promicanje rodne ravnopravnosti na sveučilištu. Plan se sastoji od niza aktivnosti i mjera usmjerenih na različita područja, uključujući uvođenje rodno osjetljivog jezika u svakodnevnu upotrebu na sveučilištu, promicanje istraživanja u području rodne ravnopravnosti i integraciju rodne perspektive u nastavne programe, kao i promicanje rodne ravnopravnosti i sprječavanje rodno uvjetovanog nasilja među studentima i osobljem sveučilišta.

# 3 Pregled literature

Analiza teksta koristi se u brojnim aplikacijama, često u kombinaciji sa strojnim učenjem i tehnikama dubokog učenja. Ove su kombinacije dovele do značajnog napretka u područjima kao što su prepoznavanje emocija i mentalnog zdravlja, što pokazuju nedavne studije. Na primjer, rad (Jonnalagadda et al., 2023.) pokazuje kako se strojno učenje može učinkovito koristiti za procjenu emocionalnih stanja iz teksta. Slično tome, studija (Rustagi et al., 2022.) istražuje upotrebu metoda dubinskog učenja za poboljšanje otkrivanja sarkazma u online recenzijama uključivanjem kontekstualnih znakova kao što su emoji i ocjene.

Kada je riječ o analizi tekstova i jezika u kontekstu rodne jednakosti, postoji mnoštvo studija koje pokazuju da jezik može biti moćan alat za perpetuiranje postojećih rodnih stereotipa i nejednakosti.

Ključne publikacije u području istraživanja jezika i rodne jednakosti uključuju knjigu "Language and Woman's Place" Robin Lakoff (1973). Ova knjiga smatra se jednom od najutjecajnijih publikacija o odnosu između jezika i roda. Lakoff tvrdi da postoje određeni jezični obrasci koji doprinose nejednakom položaju žena u društvu. "Language and Gender: An Introduction" Mary Talbot (1998) je knjiga koja pruža detaljan pregled odnosa između jezika i roda, uključujući pregled različitih teoretskih perspektiva i

overview of the relationship between language and gender, including an overview of different theoretical perspectives and examples from research. In the book "Language and Sexism" by Sara Mills (2008), the author examines various examples of sexism in language, including an analysis of the differences in the use of feminine and masculine words and the stereotypical roles associated with gender. "The Gendering of Language: A Comparison of Gender Equality in Countries with Gendered, Natural Gender, and Genderless Languages" by Jennifer L. et. al (2012) is a study that compares the level of gender equality in countries with gendered and genderless languages. The collection of essays "The Handbook of Language, Gender, and Sexuality" by Susan Ehrlich and colleagues (2014) provides an interdisciplinary overview of language and gender studies with a focus on sexuality and queer theory. The authors cover topics such as linguistic identities, gender, sexuality and language policies. The book "Language and Gender: An Advanced Resource Book" by Jane Sunderland (2006) provides an overview of the most important topics in language and gender studies, including the analysis of gender stereotypes, linguistic variation and gender identities. The author emphasizes the importance of a critical understanding of language and gender in today's society. "Doing Gender in Media, Art and Culture" by Rosemarie Buikema, Liedeke Plate, Kathrin Thiele (2018) is a book that provides an overview of the role of language in the construction of gender in media, art and culture, with a particular focus on critically analyzing and reflecting on the role of women and LGBTQ (lesbian, gay, bisexual, transgender, queer) individuals. The paper "Introduction to Gender in Language Education" by Widodo and Elyas (2020) introduces approaches and strategies for incorporating gender issues into language education. The authors emphasize the importance of including gender equality in language education, arguing that this contributes to the development of various critical language skills and improves social justice and equality.

In addition, there are numerous studies that point to the presence of gendered language stereotypes and stereotypes in relation to job roles and job advertisements. Some of these studies are: "An Exploratory Study on Linguistic Gender Stereotypes and their Effects on Perception" by Mattias Lindvall-Östling, et. al (2020), which examines linguistic gender stereotypes and their effects on people's perceptions. The results point to the existence of certain linguistic stereotypes that affect how people perceive social situations and the potential negative impact of these stereotypes on gender equality. Milica Mihaljević (2021) analyses the use of male and female gender in the Croatian language and highlights stereotypes and prejudices that are conveyed and perpetuated through language. The author highlights the problem of discrimination and

primjere istraživanja. U knjizi "Language and Sexism" Sare Mills (2008), autorica istražuje različite primjere seksizma u jeziku, uključujući analizu razlika u upotrebi ženskih i muških riječi i stereotipne uloge povezane s rodom. "The Gendering of Language: A Comparison of Gender Equality in Countries with Gendered, Natural Gender, and Genderless Languages" Jennifer L. i sur. (2012) je studija koja uspoređuje razinu rodne jednakosti u zemljama s rodnim i bezrodnim jezicima. Zbirka eseja "The Handbook of Language, Gender, and Sexuality" Susan Ehrlich i sur. (2014) pruža interdisciplinarni pregled studija jezika i roda s fokusom na seksualnost i „teoriju neobičnosti“ (eng. queer theory). Autori pokrivaju teme poput jezičnih identiteta, roda, seksualnosti i jezičnih politika. Knjiga "Language and Gender: An Advanced Resource Book" Jane Sunderland (2006) pruža pregled najvažnijih tema u studijama jezika i roda, uključujući analizu rodnih stereotipa, jezičnu varijaciju i rodne identitete. Autorica naglašava važnost kritičkog razumijevanja jezika i roda u današnjem društvu. "Doing Gender in Media, Art and Culture" Rosemarie Buikema, Liedeke Plate i Kathrin Thiele (2018) je knjiga koja pruža pregled uloge jezika u konstrukciji roda u medijima, umjetnosti i kulturi, s posebnim fokusom na kritičku analizu i refleksiju uloge žena i LGBTQ (lezbijskih, gej, biseksualnih, transrodnih, queer) osoba. Rad "Introduction to Gender in Language Education" Widodo i Elyas (2020) pruža uvod u pristupe i strategije za uključivanje rodnih pitanja u obrazovanje jezika. Autori naglašavaju važnost uključivanja rodne jednakosti u učenje jezika, tvrdeći da to doprinosi razvoju različitih kritičnih jezičnih vještina i poboljšava socijalnu pravdu i jednakost.

Osim toga, postoje brojna istraživanja koja upućuju na prisutnost rodnih jezičnih stereotipa i stereotipa u vezi s radnim ulogama i oglasima za posao. Neka od tih istraživanja su: "An Exploratory Study on Linguistic Gender Stereotypes and their Effects on Perception" Mattiasa Lindvall-Östlinga i sur. (2020), koje proučava jezične rodne stereotipe i njihove učinke na percepciju ljudi. Rezultati ukazuju na postojanje određenih jezičnih stereotipa koji utječu na to kako ljudi percipiraju socijalne situacije i potencijalni negativni utjecaj tih stereotipa na rodnu jednakost. "Muško i žensko u hrvatskome jeziku i leksikografiji – stereotipi i jezična diskriminacija" Milice Mihaljević (2021) je djelo koje analizira upotrebu muškog i ženskog roda u hrvatskom jeziku i ističe stereotipe i predrasude koje se prenose i perpetuiraju kroz jezik. Autorica ističe problem diskriminacije i nejednakosti koja se odražava u jeziku i poziva na promjene u jezičnom izrazu kako bi se postigla rodna jednakost. U članku "Gender Stereotypes and Their Impact on Women's Career Progressions from a Managerial Perspective" Tabassuma i Nayaka (2021), autori tvrde da su rodni stereotipi prisutni u današnjem društvu i da utječu na

inequality reflected in language and calls for changes in linguistic expression in order to achieve gender equality. In the article "Gender Stereotypes and Their Impact on Women's Career Progressions from a Managerial Perspective" by Tabassum and Nayak (2021), the authors argue that gender stereotypes are present in today's society and influence women's career progression from a managerial perspective. They propose the development of a gender-neutral and affirmative managerial perspective based on critical gender theory. The study "Gender Stereotypes and Peer Selection in STEM Domains Among Children and Adolescents" by Luke McGuire et al. (2022) examines how gender stereotypes influence peer selection in STEM fields among children and adolescents and how these patterns can be changed.

The study "Gender Stereotyping in Newspaper Advertisements: A Cross-Cultural Study" by Tartaglia and Rollero (2015) examined the presence of gender stereotypes in relation to occupations and sexualization in newspaper advertisements in two European countries that differ in terms of language, cultural values and gender equality - Italy and the Netherlands. The results showed that men were most often portrayed in professional roles, while women were more often depicted as decorative objects. Compared to men, female roles were also more sexualized, and this phenomenon was particularly pronounced in Italy. The paper "Balancing Gender Bias in Job Advertisements with Text-Level Bias Mitigation" by Shenggang Hu et al. (2022) examines the problem of gender bias in job advertisements, which can selectively attract applicants of a particular gender and reinforce gender segregation in the labour market. The authors propose an algorithm that assesses gender bias in job advertisements and provides guidance on how the text should be modified to eliminate gender bias and discrimination. This method has potential applications in the field of human resources, including the development of job advertisements and the screening of job applications using algorithms.

There are also studies that address the issue of gendered language and bias in translation, such as: “Gender Bias in Machine Translation” by Beatrice Savoldi et. al. (2021). This study provides a critical review of the current literature on gender bias in machine translation, including the conceptualization of bias, previous analyses, and proposed strategies for mitigation. The authors also suggest possible directions for future research in this area.

Due to the increasing popularity of natural language processing (NLP) and machine learning (ML) tools, it is becoming increasingly important to recognise their role in the formation of social biases and stereotypes. Below are examples of research applying artificial intelligence algorithms in natural language processing to improve gender equality and mitigate gender bias in language. “Mitigating Gender napredovanje žena u karijeri iz menadžerske perspektive. Predlažu razvoj rodno neutralne i afirmativne menadžerske perspektive koja se temelji na kritičkoj rodnoj teoriji. Studija "Gender Stereotypes and Peer Selection in STEM Domains Among Children and Adolescents" Lukea McGuirea i sur. (2022) proučava kako rodni stereotipi utječu na odabir vršnjaka u STEM područjima među djecom i adolescentima te kako se ti obrasci mogu promijeniti.

Studija "Gender Stereotyping in Newspaper Advertisements: A Cross-Cultural Study" Tartaglia i Rollero (2015) ispitivala je prisutnost rodnih stereotipa u vezi sa zanimanjima i seksualizacijom u oglasima u novinama u dvjema europskim zemljama koje se razlikuju u pogledu jezika, kulturnih vrijednosti i rodne jednakosti - Italiji i Nizozemskoj. Rezultati su pokazali da su muškarci najčešće prikazivani u profesionalnim ulogama, dok su žene češće prikazivane kao dekorativni objekti. U usporedbi s muškarcima, ženske uloge bile su također više seksualizirane, a ovaj fenomen bio je posebno izražen u Italiji. Rad "Balancing Gender Bias in Job Advertisements with Text-Level Bias Mitigation" Shengganga Hua i sur. (2022) istražuje problem rodne pristranosti u oglasima za posao, koja može selektivno privući kandidate određenog spola i ojačati rodnu segregaciju na tržištu rada. Autori predlažu algoritam koji procjenjuje rodnu pristranost u oglasima za posao i pruža smjernice o tome kako bi tekst trebao biti izmijenjen kako bi se uklonila rodna pristranost i diskriminacija. Ova metoda ima potencijalne primjene u području ljudskih resursa, uključujući razvoj oglasa za posao i pregled prijava za posao koristeći algoritme.

Također postoje studije koje se bave problemom rodno specifičnog jezika i pristranosti u prevođenju, poput: “Gender Bias in Machine Translation” Beatrice Savoldi i sur. (2021). Ova studija pruža kritički pregled trenutne literature o rodnoj pristranosti u strojnom prevođenju, uključujući konceptualizaciju pristranosti, prethodne analize i predložene strategije za ublažavanje. Autori također sugeriraju moguće smjernice za buduća istraživanja u ovom području.

S obzirom na sve veću popularnost alata za obradu prirodnog jezika (NLP) i strojnog učenja (ML), sve je važnije prepoznati njihovu ulogu u oblikovanju društvenih pristranosti i stereotipa. Ispod su primjeri istraživanja koja primjenjuju algoritme umjetne inteligencije u obradi prirodnog jezika kako bi poboljšala rodnu jednakost i ublažila rodnu pristranost u jeziku. “Mitigating Gender Bias in Natural Language Processing: Literature Review" Sun i sur. (2019) je rad koji pregledava trenutne studije koje se bave identificiranjem i ublažavanjem rodne pristranosti u NLP-u, uzimajući u obzir oblike pristranosti zbog rodne nejednakosti i analizirajući metode za identificiranje i ublažavanje rodne pristranosti, kao i moguće smjernice za buduća istraživanja u ovom području. Istraživački rad

Bias in Natural Language Processing: Literature Review" by Sun et al. (2019) is a paper that reviews current studies that address the identification and mitigation of gender bias in NLP, considering forms of bias due to gender inequality and analysing methods for identifying and mitigating gender bias, as well as potential guidelines for future research in this area. The research paper "Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods" by Jieyu Zhao et al. (2018) presents a method for evaluating gender bias in coreference systems and develops techniques for reducing such bias.

In addition to traditional methods, the emergence of generative artificial intelligence is changing the landscape of text and language analysis. Large-scale language models such as GPT-3 and its successors have been trained with human feedback to detect and mitigate biases. For example, Ouyang et al. (2022) describe how models such as InstructGPT utilize user feedback to compensate for gender bias and prevent the prevalence of violence in generated text. Similarly, Dobša (2023) demonstrates how ChatGPT, based on the GPT-4 model, exhibits both human-like reasoning and distinctive AI behavior, highlighting the nuanced capabilities and limitations of these models. These advances represent an important addition to the more traditional text analysis techniques discussed previously and emphasize the importance of incorporating modern AI methods into our understanding of language-related inequalities.

All of the above research and papers demonstrate the importance of working towards gender equality in different areas of society, including text and language analysis. Such research is of great importance because it highlights problems and inadequacies in the way gender identity is discussed in the media and in language in general. Furthermore, this research can contribute to the development of tools and algorithms that help to identify and mitigate gender discrimination, bias and violence in texts.

# 4 Text Analysis Methods

Text analysis has proven to be a powerful tool for examining social issues, including gender inequality and violence in language. By analysing texts, patterns, tendencies and prejudices that are not immediately apparent can be uncovered. It also allows large volumes of texts to be analysed quickly. This section introduces the main methods of text analysis for the study of gender inequality, bias and violence in language and presents their applications, tools, examples and related research.

"Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods" Jieyu Zhao i sur. (2018) predstavlja metodu za evaluaciju rodne pristranosti u sustavima za koreferenciranje i razvija tehnike za smanjenje takve pristranosti.

Uz tradicionalne metode, pojava generativne umjetne inteligencije mijenja područje analize teksta i jezika. Jezični modeli velikih razmjera kao što je GPT-3 i njegovi nasljednici trenirani su s ljudskim povratnim informacijama za otkrivanje i ublažavanje pristranosti. Na primjer, Ouyang i sur. (2022) opisuju kako modeli poput InstructGPT koriste povratne informacije korisnika kako bi kompenzirali rodnu pristranost i spriječili prevalenciju nasilja u generiranom tekstu. Slično tome, Dobša (2023) pokazuje kako ChatGPT, temeljen na GPT-4 modelu, pokazuje i ljudsko razmišljanje i karakteristično ponašanje umjetne inteligencije, ističući nijansirane mogućnosti i ograničenja ovih modela. Ovaj napredak predstavlja važan dodatak tradicionalnijim tehnikama analize teksta o kojima smo prethodno govorili i naglašava važnost uključivanja modernih metoda umjetne inteligencije u naše razumijevanje nejednakosti povezanih s jezikom.

Sva navedena istraživanja i radovi pokazuju važnost rada na rodnoj jednakosti u različitim područjima društva, uključujući analizu teksta i jezika. Takva istraživanja su od velike važnosti jer ističu probleme i neadekvatnosti u načinu na koji se o rodnom identitetu raspravlja u medijima i jeziku općenito. Nadalje, takva istraživanja mogu doprinijeti i razvoju alata i algoritama koji pomažu u identifikaciji i ublažavanju rodne diskriminacije, pristranosti i nasilja u tekstovima.

# 4 Metode analize teksta

Analiza teksta pokazala se kao moćan alat za ispitivanje socijalnih problema, uključujući rodnu nejednakost i nasilje u jeziku. Analizom tekstova mogu se otkriti obrasci, tendencije i predrasude koje nisu očite na prvi pogled. Također, korištenje alata za analizu tekstova omogućava se brzo analiziranje velikih količina teksta. Ova sekcija uvodi glavne metode analize teksta za proučavanje rodne nejednakosti, pristranosti i nasilja u jeziku te prikazuje njihove primjene, alate, primjere i povezana istraživanja.

## 4.1 Korpusna lingvistika

Korpusna lingvistika uključuje proučavanje jezika koristeći velike zbirke tekstova poznate kao korpusi. Korpusi se mogu analizirati kako bi se identificirali obrasci i trendovi u upotrebi jezika (McEnery i Hardie, 2011).

Korpusna lingvistika koristi tehnike kao što su analiza konkordanci i analiza ključnih riječi kako bi

## 4.1 Corpus Linguistics

Corpus linguistics involves the study of language using large collections of texts known as corpora. These corpora can be analysed to identify patterns and trends in language use (McEnery & Hardie, 2011).

Corpus linguistics uses techniques such as concordance analysis and keyword analysis to identify patterns in language use in relation to gender-specific terms and violence. For example, corpus linguistics can be used to study gender bias by analysing the frequency and context of gendered pronouns, terms and titles. Researchers could also investigate shifts in language use from "chairman" to "chairperson" or from "fireman" to "firefighter". In addition, corpus linguistics can be used to examine the occurrence and context of words related to violence such as "abuse"," "assault" and "victim"," particularly in court transcripts and media reports. Researchers could also compile a corpus of newspaper articles from different time periods and analyse the frequency and context of gendered terms to determine changes in societal attitudes toward gender neutrality. Or a corpus of court transcripts can be analysed to examine the language used to describe victims and perpetrators of domestic violence and to identify differences in representation by gender.

For this purpose, AntConc (Anthony, 2024) - a free corpus analysis toolkit for concordance and text analysis or Sketch Engine (Kilgarriff et al., 2014) - a tool for building and analysing corpora - can be used.

Relevant research in this area includes the work of Santoniccolo et al. (2023) and Davis, Mavisakalyan & Weber (2022).

## 4.2 Sentiment Analysis

Sentiment analysis or opinion analysis is about determining the emotional tone behind words. It can reveal the attitudes, opinions and emotions expressed in texts (Liu, 2012).

It uses machine learning algorithms, such as logistic regression or neural networks, to determine the emotional tone of texts. This process can uncover biases in public sentiment towards gender-specific topics. Real-time applications could include analysing social media sentiment during major events, such as protests for gender equality, so that researchers can capture immediate public reactions.

Sentiment analysis can also reveal negative sentiments in texts that could indicate hostility or aggression, especially towards a particular gender. It can also compare the sentiment scores of texts written by or about different genders to reveal potential inequalities. For example, sentiment analysis can be applied to a large dataset of tweets to assess public sentiment towards female political candidates compared to their male counterparts to

se identificirali obrasci u korištenju jezika u odnosu na rodno specifične izraze i nasilje. Na primjer, korpusna lingvistika može se koristiti za proučavanje rodne pristranosti analizom učestalosti i konteksta rodnih zamjenica, pojmova i naslova. Istraživači bi također mogli istražiti promjene u upotrebi izraza "chairman" sa "chairperson" ili "fireman" s "firefighter". Osim toga, korpusna lingvistika može se koristiti za ispitivanje pojavljivanja i konteksta riječi vezanih uz nasilje kao što su "abuse", "assault" i "victim", posebno u zapisima sa suđenja i medijskim izvještajima. Istraživači bi također mogli sastaviti korpus novinskih članaka iz različitih vremenskih razdoblja i analizirati učestalost i kontekst rodnih termina kako bi odredili promjene u društvenim stavovima prema rodnoj neutralnosti. Ili, korpus zapisa sa suđenja koji se može analizirati kako bi se ispitao jezik koji se koristi za opisivanje žrtava i počinitelja obiteljskog nasilja i identificirali razlike u prikazu prema rodu.

Za ovu svrhu mogu se koristiti alati poput AntConc (Anthony, 2024) - besplatni alat za analizu korpusa za konkordanciju i analizu teksta ili Sketch Engine (Kilgarriff i sur. 2014) - alat za izgradnju i analizu korpusa.

Relevantna istraživanja u ovom području uključuju radove Santoniccolo et al. (2023) i Davis, Mavisakalyan i Weber (2022).

## 4.2 Analiza sentimenta

Analiza sentimenta ili analiza mišljenja odnosi se na određivanje emocionalnog tona iza riječi. Može otkriti stavove, mišljenja i emocije izražene u tekstovima (Liu, 2012).

Koristi algoritme strojnog učenja, poput logističke regresije ili neuronskih mreža, za određivanje emocionalnog tona tekstova. Na taj se način mogu otkriti predrasude u stavovima javnosti prema rodno specifičnim temama. Aplikacije u stvarnom vremenu mogu uključivati analizu raspoloženja na društvenim mrežama tijekom velikih događaja, kao što su prosvjedi za ravnopravnost spolova, tako da istraživači mogu prikupiti trenutne reakcije javnosti.

Analizom sentimenta mogu se također otkriti negativni osjećaji u tekstovima koji bi mogli ukazivati na neprijateljstvo ili agresiju, posebno prema određenom spolu. Također, mogu se usporediti rezultati sentimenta tekstova koje su napisali autori različitih spolova ili u kojima se piše o likovima različitih spolova kako bi se otkrile moguće nejednakosti. Na primjer, analiza sentimenta može se primijeniti na veliki skup podataka s tvitovima kako bi se procijenio javni sentiment prema ženama političkim kandidatkinjama u usporedbi s njihovim muškim kolegama i kako bi se na taj način otkrila potencijalna rodna pristranost. Analiza sentimenta se može također koristiti za analizu komentara na mrežnim forumima ili u

uncover potential gender bias. It can also be used to analyse comments in online forums or news articles to identify and quantify hostile or supportive sentiments towards victims of gender-based violence.

Tools that can be used for this purpose include: VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014) - a lexicon and rule-based sentiment analysis tool and TextBlob (Loria, 2018) - a Python library for processing text data.

Research in this area includes the work of Thelwall (2018) and Davidson et al. (2017).

## 4.3 Content Analysis

Content analysis is a method for systematically coding and categorising texts in order to identify themes and patterns (Kleinheksel et al., 2020).

Content analysis can be used for thematic analysis to identify recurring themes related to gender inequality, such as stereotypes, discrimination and power dynamics. It can also analyse the representation and frequency of violent acts in various texts, including media and literature. For example, researchers could conduct a content analysis of television programmes to identify recurring themes of gender stereotypes, such as the portrayal of women primarily in caregiving roles or of men as authority figures. Similarly, novels or films can be examined for themes of violence against women to identify the frequency and context of such violence.

NVivo (Jackson & Bazeley, 2019) – a qualitative data analysis (QDA) software or QDA Miner (Provalis Research, 2021) - a tool for coding and analysing textual data can be used for this purpose.

Relevant research in this area includes works by Dines and Humez (2003) and Organization for Security and Co-operation in Europe [OSCE] (2023).

## 4.4 Discourse Analysis

Discourse analysis focuses on how language is used in texts and contexts and examines the social and cultural contexts that influence language use (Gee, 2014).

Discourse analysis can examine how language reflects and maintains power relations between genders. It can also analyse the narratives and language structures that contribute to gender inequality and the normalisation of violence. Techniques include analyzing the structure and function of political speeches or news articles to reveal biases in reporting on gender and violence. Researchers can analyse political speech to determine how language reflects and perpetuates power dynamics between genders. Similarly, news articles can be examined to understand how the media reports incidents of domestic violence and to

novinskim člancima kako bi se identificirali i kvantificirali neprijateljski ili podržavajući sentimenti prema žrtvama nasilja temeljenog na rodu.

Alati koji se mogu koristiti u tu svrhu uključuju: VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014) - leksikon i alat za analizu sentimenta temeljen na pravilima i TextBlob (Loria, 2018) - Python biblioteka za obradu tekstualnih podataka.

Istraživanja u ovom području uključuju radove Thelwall (2018) i Davidson i sur. (2017).

## 4.3 Analiza sadržaja

Analiza sadržaja je metoda sustavnog kodiranja i kategoriziranja tekstova s ciljem identifikacije tema i obrazaca (Kleinheksel et al., 2020).

Analiza sadržaja može se koristiti za tematsku analizu kako bi se identificirale ponavljajuće teme povezane s rodnom nejednakošću, kao što su stereotipi, diskriminacija i dinamika moći. Također mogu se analizirati prikaz i učestalost nasilnih činova u različitim tekstovima, uključujući medije i književnost. Na primjer, istraživači bi mogli provesti analizu sadržaja televizijskih programa kako bi identificirali ponavljajuće teme rodnih stereotipa, kao što je prikaz žena prvenstveno u ulogama njege ili muškaraca kao autoritativnih figura. Slično tome, romani ili filmovi mogu se ispitati zbog tema nasilja nad ženama kako bi se identificirala učestalost i kontekst takvog nasilja.

NVivo (Jackson & Bazeley, 2019) - softver za kvalitativnu analizu podataka (QDA) ili QDA Miner (Provalis Research, 2021) - alat za kodiranje i analizu tekstualnih podataka mogu se koristiti u te svrhe.

Relevantna istraživanja u ovom području uključuju radove Dines i Humez (2003) i Organization for Security and Co-operation in Europe [OSCE] (2023).

## 4.4 Analiza diskursa

Analiza diskursa usredotočuje se na analizu korištenja jezika u tekstovima i kontekstima te ispituje socijalne i kulturne kontekste koji utječu na upotrebu jezika (Gee, 2014).

Analiza diskursa može ispitati kako jezik odražava i održava odnose moći između spolova. Također može analizirati narative i jezične strukture koje doprinose rodnoj nejednakosti i normalizaciji nasilja. Tehnike uključuju analizu strukture i funkcije političkih govora ili novinskih članaka kako bi se otkrile pristranosti u izvještavanju o rodu i nasilju. Istraživači mogu analizirati političke govore kako bi utvrdili na koji način jezik odražava i perpetuira dinamiku moći između spolova. Slično tome, novinski članci mogu se ispitati kako bi se razumjelo kako mediji izvještavaju o incidentima obiteljskog nasilja i identificirali bilo kakvu

identify any biases in the language used. They can also analyse newspaper reports of rape cases, examining the language used to describe the victim and the accused and identifying patterns that indicate societal attitudes towards gender and violence.

Tools that may be useful in this area include Atlas.ti (ATLAS.ti Scientific Software Development GmbH, 2024) - a qualitative data analysis software, MAXQDA (VERBI Software, 2024) - a qualitative and mixed methods research software or Linguistic Inquiry and Word Count (LIWC) (Boyd, Ashokkumar, Seraj, & Pennebaker, 2022) - a text analysis software that counts words in psychologically meaningful categories.

Relevant research includes work by Cuklanz (2019), Zubrzycka-Czarnecka (2021) and Wodak (2015).

## 4.5 Network Analysis

Network analysis examines the relationships between entities in a text, such as characters in a story or people in social media interactions (Scott, 2017).

Network analysis can examine social networks and interactions depicted in texts to understand gender dynamics and instances of violence. It can also map the influence and power structures in texts to see how gender roles and violence are portrayed. Researchers can conduct a network analysis of novels to map the interactions between male and female characters to uncover the gender dynamics and power structures within the narrative. They can also analyse social media interactions to understand how information about gender violence spreads and identify key influencers.

Relevant tools in this area include Gephi (Bastian, Heymann, & Jacomy, 2009) – an open-source network analysis and visualisation software and NodeXL (Smith & Starke, 2019) - an Excel plugin for network analysis.

Relevant research in this field includes the work of Leone, Lapucci, De Sario and colleagues (2019) and Sánchez-Rodríguez, Moreno-Bella and García-Sánchez (2023).

## 4.6 Machine Translation

In machine translation, software is used to automatically translate text or speech from one language into another. Computational linguistics and artificial intelligence are used to perform these translations (Hutchins & Somers, 1992).

Gender bias in translation includes the study of how gender biases that are present in a language may be maintained or altered in translation. Machine translation systems may inadvertently introduce or reinforce gender biases that are present in the source language or training data. It is also important to recognise where gender-neutral words in the source language may be translated into gender-specific

pristranosti u korištenom jeziku. Također mogu analizirati novinske izvještaje o slučajevima silovanja, ispitujući jezik korišten za opisivanje žrtve i optuženika te identificirati obrasce koji ukazuju na društvene stavove prema rodu i nasilju.

Alati koji mogu biti korisni u ovom području uključuju Atlas.ti (ATLAS.ti Scientific Software Development GmbH, 2024) - softver za kvalitativnu analizu podataka, MAXQDA (VERBI Software, 2024) - softver za kvalitativna i mješovita istraživanja ili Linguistic Inquiry and Word Count (LIWC) (Boyd, Ashokkumar, Seraj, & Pennebaker, 2022) - softver za analizu teksta koji broji riječi unutar pojedinih psihološki značajnih kategorija.

Relevantna istraživanja uključuju radove Cuklanz (2019), Zubrzycka-Czarnecka (2021) i Wodak (2015).

## 4.5 Analiza mreža

Analizom mreža ispituju se odnosi između entiteta u tekstu, kao što su likovi u priči ili ljudi u interakcijama na društvenim mrežama (Scott, 2017).

Analizom mreža mogu se proučavati društvene mreže i interakcije prikazane u tekstovima na društvenim mrežama kako bi se bolje razumjela rodna dinamika i slučajevi nasilja. Također, mogu se mapirati utjecaji i strukture moći u tekstovima kako bi se vidjelo na koji su način prikazane rodne uloge i nasilje. Istraživači mogu, primjerice, provesti i analizu mreža romana kako bi mapirali interakcije između muških i ženskih likova i na taj način otkrili rodne dinamike i strukture moći unutar narativa. Također, mogu analizirati interakcije na društvenim mrežama kako bi razumjeli na koji način informacije o rodnom nasilju cirkuliraju i identificirali ključne influencere.

Relevantni alati u ovom području uključuju Gephi (Bastian, Heymann, & Jacomy, 2009) – softver otvorenog koda za analizu mreža i vizualizaciju i NodeXL (Smith & Starke, 2019) - Excel dodatak za analizu mreža.

Relevantna istraživanja u ovom području uključuju radove Leone, Lapucci, De Sario i sur. (2019) i Sánchez-Rodríguez, Moreno-Bella i García-Sánchez (2023).

## 4.6 Strojno prevođenje

U strojnom prevođenju koristi se softver za automatsko prevođenje teksta ili govora s jednog jezika na drugi. Za izvođenje tih prijevoda koriste se računalna lingvistika i umjetna inteligencija (Hutchins & Somers, 1992).

Rodna pristranost u prevođenju uključuje proučavanje načina na koji rodne pristranosti prisutne u jeziku mogu biti odražene ili izmijenjene u prevođenju. Sustavi za strojno prevođenje mogu nenamjerno uvesti ili pojačati rodne pristranosti prisutne u izvornom jeziku ili podacima za treniranje

words in the target language. New techniques can be developed to reduce gender bias in machine translation.

Another aspect is to compare the translation of violent language in different languages to examine consistency and bias. Differences in translation can shed light on how different cultures and languages deal with or interpret violence.

Additionally, the field includes the analysis of translating gender-neutral sentences and the resulting gender bias of machine translation systems such as Google Translate (Google LLC, 2023) or Microsoft Translator (Microsoft Corporation, 2023).

Research that has analysed machine translation systems in this way includes the works of Prates, Avelar & Lamb (2020) and Savoldi et al. (2021). Research that addresses reducing gender bias in machine translations includes the works of Vanmassenhove, Hardmeier & Way (2018) and Escudé Font and Costa-jussà (2019).

Table 1 presents an overview of the described text analysis methods, their applications, tools that can be used for their implementation, examples of their application, and relevant research in each method.

modela. Također je važno prepoznati gdje se rodno-neutralne riječi u izvornom jeziku mogu prevesti u rodno-specifične riječi u ciljanom jeziku. Moguće je razviti nove tehnike za smanjenje rodne pristranosti u strojnim prijevodima.

Drugi aspekt je uspoređivanje prijevoda nasilnog jezika na različitim jezicima kako bi se ispitala dosljednost i pristranost. Razlike u prevođenju mogu osvijetliti kako različite kulture i jezici tretiraju ili interpretiraju nasilje.

Također, područje obuhvaća i analizu prevođenja rodno-neutralnih rečenica i rezultirajuće rodne pristranosti sustava za strojno prevođenje poput Google prevoditelja (Google LLC, 2023) ili Microsoft prevoditelja (Microsoft Corporation, 2023).

Istraživanja u kojima su na taj način analizirali prijevode sustava za strojno prevođenje uključuju radove Prates, Avelar & Lamb (2020) i Savoldi et al. (2021). Istraživanja koja se bave smanjenjem rodne pristranosti u strojnim prijevodima uključuju radove Vanmassenhove, Hardmeier & Way (2018) i Escudé Font i Costa-jussà (2019).

U Tablici 1 prikazan je pregled opisanih metoda analize teksta, njihova primjena, alati koji se mogu koristiti za njihovu implementaciju, primjeri primjene i relevantna istraživanja u pojedinim metodama.

**Table 1.** Overview of text analysis methods for identifying gender bias, inequality and violence in language
**Tablica 1.** Pregled metoda analize teksta za identifikaciju rodne pristranosti, nejednakosti i nasilja u jeziku

| Method | Applications | Tools | Examples | Related Research |
|---|---|---|---|---|
| **Corpus Linguistics** | Analyzing language patterns and trends in large collections of texts. | AntConc (Anthony, 2024), Sketch Engine (Kilgarriff et al., 2014) | Examining shifts in gendered terms, analyzing violence-related words. | Santoniccolo et al. (2023), Davis, Mavisakalyan & Weber (2022), Mihaljević (2021) |
| **Sentiment Analysis** | Determining the emotional tone and attitudes in texts. | VADER (Hutto & Gilbert, 2014), TextBlob (Loria, 2018) | Assessment of sentiment towards female candidates, analysis of comments on gender violence. | Thelwall (2018), Davidson et al. (2017) |
| **Content Analysis** | Coding and categorizing texts to find themes and patterns. | NVivo (Jackson & Bazeley, 2019), QDA Miner (Provalis Research, 2021) | Identifying gender stereotypes in media, analyzing themes of violence in literature. | Dines and Humez (2003), OSCE (2023), Mills (2008), Talbot (1998), Sunderland (2006) |
| **Discourse Analysis** | Examining the use of language in its social and cultural context. | Atlas.ti (ATLAS.ti Scientific Software Development GmbH, 2024), MAXQDA (VERBI Software, 2024), LIWC (Boyd et al., 2022) | Analyzing political speeches and media reports for bias. | Cuklanz (2019), Zubrzycka-Czarnecka (2021), Lakoff (1973), Buikema et al. (2018) |
| **Network Analysis** | Investigating relationships between entities in texts. | Gephi (Bastian et al., 2009), NodeXL (Smith & Starke, 2019) | Mapping gender dynamics in novels, analyzing social media interactions on violence. | Leone et al. (2019), Sánchez-Rodríguez et al. (2023) |
| **Machine Translation** | Studying gender bias in translation and exploring cultural differences. | Google Translate (Google LLC, 2023), Microsoft Translator (Microsoft Corporation, 2023) | Analyzing gender-neutral and gender-specific translations, comparison of translations of violent language. | Prates et al. (2020), Savoldi et al. (2021), Vanmassenhove et al. (2018), Escudé Font & Costa-jussà (2019) |

# 5 Conclusion

This paper provides an overview of text analysis methods used to identify and recognize gender bias, inequality and violence in language. Gender equality is fundamental to creating a fair and inclusive society, and language plays an important role in shaping social norms and values. Various text analysis methods, including corpus linguistics, sentiment analysis, content analysis, discourse analysis, network analysis and machine translation, provide powerful tools for detecting, eliminating and mitigating gender bias in language.

These methods have been applied in various fields, including media, literature, social media, job advertisements and political speeches, to uncover

# 5 Zaključak

Ovaj rad pruža pregled metoda analize teksta koje se koriste za identificiranje i prepoznavanje rodne pristranosti, nejednakosti i nasilja u jeziku. Rodna ravnopravnost je temelj za stvaranje pravednog i uključivog društva, a jezik igra važnu ulogu u oblikovanju društvenih normi i vrijednosti. Različite metode analize teksta, uključujući korpusnu lingvistiku, analizu sentimenta, analizu sadržaja, analizu diskursa, analizu mreža i strojno prevođenje, predstavljaju snažne alate za otkrivanje, uklanjanje i ublažavanje rodne pristranosti u jeziku.

Te su metode primijenjene u raznim područjima, uključujući medije, književnost, društvene mreže, oglase za posao i političke govore, kako bi se otkrila

and mitigate gender bias. The results of studies highlight the prevalence of gender bias in language and the importance of using text analysis to promote gender equality. In addition, advances in natural language processing (NLP) and machine learning offer new opportunities to develop tools and algorithms that can automatically detect and correct gender bias. By integrating these technologies into everyday applications, such as machine translation systems and sentiment analysis tools, we can promote a gender-neutral and inclusive linguistic environment.

Future research will further refine these methods and explore their applications in different linguistic and cultural contexts. Generative AI and large language models (LLMs) have shown promise in reducing language-based inequalities, demonstrating the need for further research and application of these technologies to further reduce bias. Furthermore, interdisciplinary collaboration is needed to ensure that the results of text analysis are effectively translated into practical strategies to promote gender equality. By harnessing the power of text analysis, we can contribute to a fairer society in which language serves as a tool for inclusion rather than discrimination.

i ublažila rodna pristranost. Rezultati istraživanja ističu prisutnost rodne pristranosti u jeziku i važnost korištenja analize teksta za promicanje rodne ravnopravnosti. Osim toga, napredak u obradi prirodnog jezika (NLP) i strojnog učenja nudi nove mogućnosti za razvoj alata i algoritama koji mogu automatski otkriti i ispraviti rodnu pristranost. Integriranjem ovih tehnologija u svakodnevne aplikacije, kao što su sustavi za strojno prevođenje i alati za analizu sentimenta, možemo promicati rodno neutralno i uključivo jezično okruženje.

Buduća istraživanja će dodatno proučiti ove metode i istražiti njihove primjene u različitim jezičnim i kulturnim kontekstima. Generativna umjetna inteligencija i veliki jezični modeli (LLM) pokazali su obećavajući napredak u smanjenju nejednakosti temeljenih na jeziku, pokazujući potrebu za daljnjim istraživanjem i primjenom ovih tehnologija za daljnje smanjenje pristranosti. Nadalje, potrebna je interdisciplinarna suradnja kako bi se osiguralo da se rezultati analize teksta učinkovito prevedu u praktične strategije za promicanje rodne ravnopravnosti. Korištenjem analize teksta, možemo doprinijeti pravednijem društvu u kojem jezik služi kao alat za uključivanje, a ne diskriminaciju.

## Acknowledgments

This work was fully supported by the University of Rijeka as part of the scientific project “LinguaGender: Primjena računalne obrade jezika u analizi rodne pristranosti i nasilja u jeziku”, uniri-iskusni-drustv-23-271.

## Zahvale

Ovaj rad je u potpunosti podržan od strane Sveučilišta u Rijeci kao dio znanstvenog projekta “LinguaGender: Primjena računalne obrade jezika u analizi rodne pristranosti i nasilja u jeziku”, uniri-iskusni-drustv-23-271.

## References

Anthony, L. (2024). *AntConc (Version 4.3.0)*. Waseda University. https://www.laurenceanthony.net/software

ATLAS.ti Scientific Software Development GmbH. (2024). *ATLAS.ti.* Retrieved from https://atlasti.com

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Weblogs and Social Media*. https://gephi.org/

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. University of Texas at Austin. https://www.liwc.app

Buikema, R., Plate, L., & Thiele, K. (2018). *Doing gender in media, art and culture*. Routledge.

## Reference

Anthony, L. (2024). *AntConc (Version 4.3.0)*. Waseda University. https://www.laurenceanthony.net/software

ATLAS.ti Scientific Software Development GmbH. (2024). *ATLAS.ti.* Retrieved from https://atlasti.com

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Weblogs and Social Media*. https://gephi.org/

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. University of Texas at Austin. https://www.liwc.app

Buikema, R., Plate, L., & Thiele, K. (2018). *Doing gender in media, art and culture*. Routledge.

Cameron, D. (2005). *Language, gender, and sexuality: Current issues and new directions*. Cambridge University Press

Croatian Employers' Association & Deloitte. (2024). *SheXO: Women in business*. Deloitte. Retrieved from https://www.deloitte.com/content/dam/Deloitte/hr/Documents/about-deloitte/hr_SheXO_Brochure_2024_spreads-final.pdf

Cuklanz, L. (2019). Representations of gendered violence in mainstream media. *Questions de communication, 35*, 307-321. https://doi.org/10.4000/questionsdecommunication.19487

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media, 11*(1). https://doi.org/10.1609/icwsm.v11i1.14955

Dines, G., & Humez, J. M. (Eds.). (2003). *Media messages: Gender, race, and class in media*. Sage Publications.

Dobša, J. (2023). *How large language models "think" and can we trust them: A case study of testing ChatGPT on tasks in an introductory statistics course*. Polytechnica: Journal of Technology Education, 7(2). https://doi.org/10.36978/cte.7.2.2

Ehrlich, S., Meyerhoff, M., & Holmes, J. (Eds.). (2014). *The handbook of language, gender, and sexuality*. Wiley-Blackwell.

Escudé Font, J., & Costa-jussà, M. R. (2019). Equalizing gender bias in neural machine translation with word embeddings techniques. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing* (pp. 147–154). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3821

European Commission. (2020). *EU gender equality strategy 2020-2025*. Retrieved from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_hr

European Commission. (2021). *Action plan for gender equality under the European Social Fund 2021-2027*. Retrieved from https://op.europa.eu/webpub/empl/european-pillar-of-socialrights/hr/index.html

European Commission. (2021). *Gender equality strategy 2020-2025*. Retrieved from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-

Cameron, D. (2005). *Language, gender, and sexuality: Current issues and new directions*. Cambridge University Press

Croatian Employers' Association & Deloitte. (2024). *SheXO: Women in business*. Deloitte. Retrieved from https://www.deloitte.com/content/dam/Deloitte/hr/Documents/about-deloitte/hr_SheXO_Brochure_2024_spreads-final.pdf

Cuklanz, L. (2019). Representations of gendered violence in mainstream media. *Questions de communication, 35*, 307-321. https://doi.org/10.4000/questionsdecommunication.19487

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media, 11*(1). https://doi.org/10.1609/icwsm.v11i1.14955

Dines, G., & Humez, J. M. (Eds.). (2003). *Media messages: Gender, race, and class in media*. Sage Publications.

Dobša, J. (2023). *How large language models "think" and can we trust them: A case study of testing ChatGPT on tasks in an introductory statistics course*. Polytechnica: Journal of Technology Education, 7(2). https://doi.org/10.36978/cte.7.2.2

Ehrlich, S., Meyerhoff, M., & Holmes, J. (Eds.). (2014). *The handbook of language, gender, and sexuality*. Wiley-Blackwell.

Escudé Font, J., & Costa-jussà, M. R. (2019). Equalizing gender bias in neural machine translation with word embeddings techniques. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing* (pp. 147–154). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3821

European Commission. (2020). *EU gender equality strategy 2020-2025*. Retrieved from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_hr

European Commission. (2021). *Action plan for gender equality under the European Social Fund 2021-2027*. Retrieved from https://op.europa.eu/webpub/empl/european-pillar-of-socialrights/hr/index.html

European Commission. (2021). *Gender equality strategy 2020-2025*. Retrieved from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-

rights/gender-equality/gender-equality-strategy_en

European Commission. (2022). *Gender equality report*. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/0fb69c07-e79d-11ec-a534-01aa75ed71a1/language-en/format-PDF/source-search

European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, Directorate General for Internal Policies of the Union. (2018). *Cyber violence and hate speech online against women: Women's rights & gender equality* (PE 604.979). Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL_STU(2018)604979_EN.pdf

Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*. Routledge.

Google LLC. (2024). *Google Translate*. Retrieved from https://translate.google.com/

Government of Croatia. (2022). *National curriculum reform strategy for education 2022-2027*.

Government of Croatia. (2023). *National strategy for gender equality 2023-2027*.

Holmes, J. (2006). Gendered Talk at Work: Constructing Gender Identity through Workplace Discourse. Blackwell Publishing.

Hu, S., Al-Ani, J. A., Hughes, K. D., Denier, N., Konnikov, A., Ding, L., Xie, J., Hu, Y., Tarafdar, M., Jiang, B., Kong, L., & Dai, H. (2022). Balancing gender bias in job advertisements with text-level bias mitigation. *Frontiers in Big Data, 5*. https://doi.org/10.3389/fdata.2022.805713.

Hutchins, W. J., & Somers, H. (1992). *An introduction to machine translation*. Academic Press Limited.

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1). https://doi.org/10.1609/icwsm.v8i1.14550

Jackson, K., & Bazeley, P. (2019). *Qualitative data analysis with NVivo* (3rd ed.). Sage Publications.

Jonnalagadda, A., Rajvir, M., Singh, S., Chandramouliswaran, S., George, J., & Kamalov, F. (2022). *An ensemble-based machine learning model for emotion and mental health detection*. Journal of Information & Knowledge Management, 22(2), Article

rights/gender-equality/gender-equality-strategy_en

European Commission. (2022). *Gender equality report*. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/0fb69c07-e79d-11ec-a534-01aa75ed71a1/language-en/format-PDF/source-search

European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, Directorate General for Internal Policies of the Union. (2018). *Cyber violence and hate speech online against women: Women's rights & gender equality* (PE 604.979). Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL_STU(2018)604979_EN.pdf

Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*. Routledge.

Google LLC. (2024). *Google Translate*. Retrieved from https://translate.google.com/

Government of Croatia. (2022). *National curriculum reform strategy for education 2022-2027*.

Government of Croatia. (2023). *National strategy for gender equality 2023-2027*.

Holmes, J. (2006). Gendered Talk at Work: Constructing Gender Identity through Workplace Discourse. Blackwell Publishing.

Hu, S., Al-Ani, J. A., Hughes, K. D., Denier, N., Konnikov, A., Ding, L., Xie, J., Hu, Y., Tarafdar, M., Jiang, B., Kong, L., & Dai, H. (2022). Balancing gender bias in job advertisements with text-level bias mitigation. *Frontiers in Big Data, 5*. https://doi.org/10.3389/fdata.2022.805713.

Hutchins, W. J., & Somers, H. (1992). *An introduction to machine translation*. Academic Press Limited.

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1). https://doi.org/10.1609/icwsm.v8i1.14550

Jackson, K., & Bazeley, P. (2019). *Qualitative data analysis with NVivo* (3rd ed.). Sage Publications.

Jonnalagadda, A., Rajvir, M., Singh, S., Chandramouliswaran, S., George, J., & Kamalov, F. (2022). *An ensemble-based machine learning model for emotion and mental health detection*. Journal of Information & Knowledge Management, 22(2), Article

2250075. https://doi.org/10.1142/S0219649222500757

Kabeer, N. (2015). Gender equality, economic growth, and women's agency: The "endless variety" and "monotonous similarity" of patriarchal constraints. *Feminist Economics, 22*(1), 295–321. https://doi.org/10.1080/13545701.2015.1090009

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*(1), 7-36. https://doi.org/10.1007/s40607-014-0009-9

Kimmel, M. S. (2000). *The gendered society*. Oxford University Press.

Kleinheksel, A. J., Rockich-Winston, N., Tawfik, H., & Wyatt, T. R. (2020). Demystifying content analysis. *American Journal of Pharmaceutical Education, 84*(1), 7113. https://doi.org/10.5688/ajpe7113

Lakoff, R. (1973). *Language and woman's place*. Harper & Row.

Lakoff, R. (2004). *Language and woman's place: Text and commentaries*. Oxford University Press.

Leone, M., Lapucci, E., De Sario, M., & et al. (2019). Social network analysis to characterize women victims of violence. *BMC Public Health, 19*, 494. https://doi.org/10.1186/s12889-019-6797-y

Lindvall-Östling, M., Deutschmann, M., & Steinvall, A. (2020). An exploratory study on linguistic gender stereotypes and their effects on perception. *Open Linguistics, 6*(1), 567-583. https://doi.org/10.1515/opli-2020-0033

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

Loria, S. (2018). *textblob documentation (Release 0.15)*.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McGuire, L., Hoffman, A. J., Mulvey, K. L., Hartstone-Rose, A., Winterbottom, M., Joy, A., Law, F., Balkwill, F., Burns, K. P., Butler, L., Drews, M., Fields, G., Smith, H., & Rutland, A. (2022). Gender stereotypes and peer selection in STEM domains among children and adolescents. *Sex Roles, 87*(9-10), 455-470. https://doi.org/10.1007/s11199-022-01327-9

Microsoft Corporation. (2004). *Microsoft Translator*. Retrieved from https://www.microsoft.com/en-us/translator

2250075. https://doi.org/10.1142/S0219649222500757

Kabeer, N. (2015). Gender equality, economic growth, and women's agency: The "endless variety" and "monotonous similarity" of patriarchal constraints. *Feminist Economics, 22*(1), 295–321. https://doi.org/10.1080/13545701.2015.1090009

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*(1), 7-36. https://doi.org/10.1007/s40607-014-0009-9

Kimmel, M. S. (2000). *The gendered society*. Oxford University Press.

Kleinheksel, A. J., Rockich-Winston, N., Tawfik, H., & Wyatt, T. R. (2020). Demystifying content analysis. *American Journal of Pharmaceutical Education, 84*(1), 7113. https://doi.org/10.5688/ajpe7113

Lakoff, R. (1973). *Language and woman's place*. Harper & Row.

Lakoff, R. (2004). *Language and woman's place: Text and commentaries*. Oxford University Press.

Leone, M., Lapucci, E., De Sario, M., & et al. (2019). Social network analysis to characterize women victims of violence. *BMC Public Health, 19*, 494. https://doi.org/10.1186/s12889-019-6797-y

Lindvall-Östling, M., Deutschmann, M., & Steinvall, A. (2020). An exploratory study on linguistic gender stereotypes and their effects on perception. *Open Linguistics, 6*(1), 567-583. https://doi.org/10.1515/opli-2020-0033

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

Loria, S. (2018). *textblob documentation (Release 0.15)*.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McGuire, L., Hoffman, A. J., Mulvey, K. L., Hartstone-Rose, A., Winterbottom, M., Joy, A., Law, F., Balkwill, F., Burns, K. P., Butler, L., Drews, M., Fields, G., Smith, H., & Rutland, A. (2022). Gender stereotypes and peer selection in STEM domains among children and adolescents. *Sex Roles, 87*(9-10), 455-470. https://doi.org/10.1007/s11199-022-01327-9

Microsoft Corporation. (2004). *Microsoft Translator*. Retrieved from https://www.microsoft.com/en-us/translator

Mihaljević, M. (2021). Muško i žensko u hrvatskome jeziku i leksikografiji – stereotipi i jezična diskriminacija. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje, 47*(2). https://doi.org/10.31724/rihjj.47.2.14

Mills, S. (1995). *Language and gender: Interdisciplinary perspectives*. Routledge.

Mills, S. (2008). *Language and sexism*. Cambridge University Press.

Organization for Security and Co-operation in Europe (OSCE). (2023). Analysis of media reporting and the effects of media reporting on gender-based violence against women and girls in family and partner relationships. OSCE. https://www.osce.org/files/f/documents/f/6/558498.pdf

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zheng, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Cristiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. arXiv. https://arxiv.org/abs/2203.02155

Prates, M. O. R., Avelar, P. H. da C., & Lamb, L. C. (2020). Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing and Applications, 32*(1). https://doi.org/10.1007/s00521-019-04144-6

Provalis Research. (2021). *QDA Miner (Version 5.0)*. https://provalisresearch.com/products/qualitative-data-analysis-software/

Rustagi, A., Jonnalagadda, A., Cherukuri, A. K., & Ahmad, A. (2022). Toward sarcasm detection in reviews—a dual parametric approach with emojis and ratings. In *Studies in Computational Intelligence* (pp. 245-257). (Studies in Computational Intelligence; Vol. 988). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-981-16-4713-0_13

Sánchez-Rodríguez, Á., Moreno-Bella, E., & García-Sánchez, E. (2023). Mapping gender stereotypes: A network analysis approach. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1193866

Santoniccolo, F., Trombetta, T., Paradiso, M. N., & Rollè, L. (2023). Gender and media representations: A review of the literature on gender stereotypes, objectification and sexualization. *International Journal of Environmental Research and Public Health, 20*(10), 5770. https://doi.org/10.3390/ijerph20105770

Mihaljević, M. (2021). Muško i žensko u hrvatskome jeziku i leksikografiji – stereotipi i jezična diskriminacija. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje, 47*(2). https://doi.org/10.31724/rihjj.47.2.14

Mills, S. (1995). *Language and gender: Interdisciplinary perspectives*. Routledge.

Mills, S. (2008). *Language and sexism*. Cambridge University Press.

Organization for Security and Co-operation in Europe (OSCE). (2023). Analysis of media reporting and the effects of media reporting on gender-based violence against women and girls in family and partner relationships. OSCE. https://www.osce.org/files/f/documents/f/6/558498.pdf

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zheng, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Cristiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. arXiv. https://arxiv.org/abs/2203.02155

Prates, M. O. R., Avelar, P. H. da C., & Lamb, L. C. (2020). Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing and Applications, 32*(1). https://doi.org/10.1007/s00521-019-04144-6

Provalis Research. (2021). *QDA Miner (Version 5.0)*. https://provalisresearch.com/products/qualitative-data-analysis-software/

Rustagi, A., Jonnalagadda, A., Cherukuri, A. K., & Ahmad, A. (2022). Toward sarcasm detection in reviews—a dual parametric approach with emojis and ratings. In *Studies in Computational Intelligence* (pp. 245-257). (Studies in Computational Intelligence; Vol. 988). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-981-16-4713-0_13

Sánchez-Rodríguez, Á., Moreno-Bella, E., & García-Sánchez, E. (2023). Mapping gender stereotypes: A network analysis approach. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1193866

Santoniccolo, F., Trombetta, T., Paradiso, M. N., & Rollè, L. (2023). Gender and media representations: A review of the literature on gender stereotypes, objectification and sexualization. *International Journal of Environmental Research and Public Health, 20*(10), 5770. https://doi.org/10.3390/ijerph20105770

Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., & Turchi, M. (2021). Gender bias in machine translation. *Transactions of the Association for Computational Linguistics, 9*, 845-874.

Scott, J. (2017). *Social network analysis*. SAGE Publications.

Smith, M. A., & Starke, A. (2019). *NodeXL: Network analysis, design, and visualization*. https://www.smrfoundation.org/nodexl/

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2019). Mitigating gender bias in natural language processing: Literature review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1630–1640). Association for Computational Linguistics.

Sunderland, J. (2006). *Language and gender: An advanced resource book*. Routledge.

Tabassum, N., & Nayak, B. (2021). Gender stereotypes and their impact on women's career progressions from a managerial perspective. *IIM Kozhikode Society & Management Review, 10*(2), 192-208. https://doi.org/10.1177/2277975220975513.

Talbot, M. (1998). *Language and Gender: An introduction*. Polity Press.

Tartaglia, S., & Rollero, C. (2015). Gender stereotyping in newspaper advertisements: A cross-cultural study. *Journal of Cross-Cultural Psychology, 46*(8), 1103-1109. https://doi.org/10.1177/0022022115597068.

Thelwall, M. (2018). Gender bias in sentiment analysis. *Online Information Review, 42*(6). https://doi.org/10.1108/OIR-05-2017-0139

UN Women. (2019). *Progress of the world's women 2019-2020: Families in a changing world*. Retrieved from https://www.unwomen.org/en/digital-library/publications/2019/06/progress-of-the-worlds-women-2019-2020

United Nations. (2020). *Gender equality*. Retrieved from https://www.un.org/sustainabledevelopment/gender-equality/

University of Rijeka. (2021). *Gender equality plan 2021-2025*. Retrieved from https://uniri.hr/wp-content/uploads/2022/05/UNIRI_plan_rodne_ravnopravnosti_HR_2022_final.pdf

Vanmassenhove, E., Hardmeier, C., & Way, A. (2018). Getting gender right in neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3003–3008).

Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., & Turchi, M. (2021). Gender bias in machine translation. *Transactions of the Association for Computational Linguistics, 9*, 845-874.

Scott, J. (2017). *Social network analysis*. SAGE Publications.

Smith, M. A., & Starke, A. (2019). *NodeXL: Network analysis, design, and visualization*. https://www.smrfoundation.org/nodexl/

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2019). Mitigating gender bias in natural language processing: Literature review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1630–1640). Association for Computational Linguistics.

Sunderland, J. (2006). *Language and gender: An advanced resource book*. Routledge.

Tabassum, N., & Nayak, B. (2021). Gender stereotypes and their impact on women's career progressions from a managerial perspective. *IIM Kozhikode Society & Management Review, 10*(2), 192-208. https://doi.org/10.1177/2277975220975513.

Talbot, M. (1998). *Language and Gender: An introduction*. Polity Press.

Tartaglia, S., & Rollero, C. (2015). Gender stereotyping in newspaper advertisements: A cross-cultural study. *Journal of Cross-Cultural Psychology, 46*(8), 1103-1109. https://doi.org/10.1177/0022022115597068.

Thelwall, M. (2018). Gender bias in sentiment analysis. *Online Information Review, 42*(6). https://doi.org/10.1108/OIR-05-2017-0139

UN Women. (2019). *Progress of the world's women 2019-2020: Families in a changing world*. Retrieved from https://www.unwomen.org/en/digital-library/publications/2019/06/progress-of-the-worlds-women-2019-2020

United Nations. (2020). *Gender equality*. Retrieved from https://www.un.org/sustainabledevelopment/gender-equality/

University of Rijeka. (2021). *Gender equality plan 2021-2025*. Retrieved from https://uniri.hr/wp-content/uploads/2022/05/UNIRI_plan_rodne_ravnopravnosti_HR_2022_final.pdf

Vanmassenhove, E., Hardmeier, C., & Way, A. (2018). Getting gender right in neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3003–3008).

Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1334

VERBI Software. (2024). *MAXQDA.* Retrieved from https://www.maxqda.com

Widodo, H. P., & Elyas, T. (2020). Introduction to gender in language education. *Sexuality & Culture, 24*(2). https://doi.org/10.1007/s12119-020-09753-1

Wodak, R. (2015). Gender and language: Cultural concerns. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 698-703). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.64018-7

World Economic Forum. (2023). *Global gender gap report 2023*. Retrieved from https://www.weforum.org/reports/global-gender-gap-report-2023

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2018). Gender bias in coreference resolution: Evaluation and debiasing methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (pp. 15–20). Association for Computational Linguistics.

Zubrzycka-Czarnecka, A. (2021). Discursive construction of the perception of gender identity: The case of tenants and owners in re-privatization/property restitution in Warsaw. *Studia z Polityki Publicznej, 8*(1). https://doi.org/10.33119/KSzPP/2021.1.2

Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1334

VERBI Software. (2024). *MAXQDA.* Retrieved from https://www.maxqda.com

Widodo, H. P., & Elyas, T. (2020). Introduction to gender in language education. *Sexuality & Culture, 24*(2). https://doi.org/10.1007/s12119-020-09753-1

Wodak, R. (2015). Gender and language: Cultural concerns. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 698-703). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.64018-7

World Economic Forum. (2023). *Global gender gap report 2023*. Retrieved from https://www.weforum.org/reports/global-gender-gap-report-2023

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2018). Gender bias in coreference resolution: Evaluation and debiasing methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (pp. 15–20). Association for Computational Linguistics.

Zubrzycka-Czarnecka, A. (2021). Discursive construction of the perception of gender identity: The case of tenants and owners in re-privatization/property restitution in Warsaw. *Studia z Polityki Publicznej, 8*(1). https://doi.org/10.33119/KSzPP/2021.1.2

# Data Spaces in the Republic of Croatia: Potential and Opportunities

**Dino Vlahek**
The Faculty of Electrical Engineering and Computer Science at the University of Maribor
Laboratory for Geospatial Modelling, Multimedia and Artificial Intelligence
dino.vlahek1@um.si

**Dijana Oreški**
Faculty of Organization and Informatics, University of Zagreb
Laboratory for data mining and intelligent systems.
10000 Zagreb, Croatia
dijoresk@foi.hr

**Abstract.** *The digital era has steered in massive production and consumption of diverse data. These data are usually stored in isolated data silos, which can lead to drawbacks such as data spread across multiple silos and data inconsistency due to conflicting or outdated versions. Moving beyond this, the concept of data space offers a flexible, integrated environment, allowing various data types and sources to coexist, enhancing data sharing and processing. This shift is crucial for countries like Croatia to embrace the concept of data space, which can improve their digital transformation and data economy. Using data spaces in public services can enhance transparency, efficiency, and the delivery of high-quality services. For the economy, leveraging data space can benefit sectors like tourism, agriculture, and green energy, fostering growth and innovation.*

**Keywords.** data space, Croatia, public sector, economy, research

## 1 Introduction

Today, humanity produces and consumes large amounts of data. This data is fluid and constantly changing, pivotal in creating innovative products and services ranging from smart agriculture, personalized medicine, earth observation, or ecological research to outdoor and indoor navigation (Constantiou and Kallinikos, 2015). In the conventional paradigm, data is often stored in isolated silos. Such an organization is usually adapted to specific applications and a closed ecosystem of users. This can lead to several issues, such as data spread across multiple silos and data inconsistency due to conflicting or outdated versions. Besides that, users outside the ecosystem often face difficulties in locating and accessing such data, while individual data silos usually provide only one type of data (Manoochehri, 2013; Franklin et al., 2005).

In contrast to the traditional data silos that enforce strict schemas and centralized storage, data space (DS) provides an integrated, loosely coupled environment where data from various sources coexist and interact (Franklin et al., 2005; Otto, 2022). DSs are abstract concepts of middleware systems primarily designed for scalable and efficient data sharing, while data processing is also included in the latest evolution. They represent a collection of unstructured, semi-structured, and structured data and connections between them in catalogs accessible by web services (Otto, 2022). DSs provide integration of diverse data sources that include various data formats simultaneously, for example, text, audio and video files, blockchain, and relation or non-relation databases. They also reinforce richer metadata and semantic annotations, supporting new data sources while reducing data consolidation, which can result in significant cost savings (Kotsev et al., 2020; Otto, 2022).

Similar opportunities can be noted when DSs are integrated with artificial intelligence (AI). On the one hand, this integration enables more efficient use of data, allowing AI researchers access to different sources while not demanding centralized storage. Better data accessibility reduces the time and effort required for data preparation, reducing or eliminating data duplication. AI systems can also continuously learn and adapt using dynamic data from DSs (Zillner et al., 2021), while feature learning algorithms (Vlahek and Mongus, 2023) can extract knowledge from different data sources. On the other hand, AI algorithms can be used in DS optimization and functionality enhancement. For example, AI can help the user discover relevant data or enhance data quality by automation the dataset cleaning process with improved identification of inconsistencies and duplicates (Boukhers et al., 2023).

In this paper, we discussed the European Union's

data strategy along with the concept of data spaces. Furthermore, we explored the implications and opportunities that data spaces could offer for Croatia, focusing on three key sectors: public services, the economy, and the domains of research and science. By delving into these areas, we aim to shed light on the transformative potential that a well-implemented data strategy could have on enhancing efficiency, innovation, and overall societal advancement.

# 2 EU and Data space

## 2.1 The European data strategy

The concept of DSs is continuously evolving, driven by technological advancements, policy changes, and data commercialization. In the early stages (the early 2000s), when the DS concept was not yet defined, the internet and digital technologies rose. This rise highlighted the potential of data as a valuable resource. In this period, the European Union (EU) started recognizing the need for a coordinated approach to data governance as big data initiatives emerged in the public and private sectors. With the Digital Single Market strategy, the EU aimed to create a coherent online environment. In this period, the concept of DSs was not yet entirely defined (Otto, 2022). The critical evolution towards the concept of DS was in 2020 when the European Commission provided a European data strategy. This strategy aims to make the EU a leader in a data-driven society and define the DS concept that enables the free flow of data within the EU. A particular emphasis is on using data between sectors to benefit entrepreneurship, researchers, public administrations, and ordinary residents. The strategy is implemented through legislative and technical measures. The key legislative measures in the DS area are (Commission, 2022):

- the European Data Governance Act,
- the Data act, and
- the AI Act.

This legislative framework specifies explicit rules for data management and sharing and regulates artificial intelligence use. The initial focus was health, agriculture, manufacturing, energy, mobility, finance, public administration, skills, and research. Additional DSs emerged through the years in media and cultural heritage areas. The current goal is to gradually interconnect these DSs, forming a unified market for data across Europe. Common data infrastructures and governance frameworks are being developed to ensure secure, fair, and green data sharing. In Addition, implementing and integrating existing DSs are underway, encouraging fully operational and interconnected DSs. The continuous evolution of technology, such as artificial intelligence and the Internet of Things, pushes additional improvements and innovations in DS. At the same time, the EU continues to refine its policies to adapt to new challenges and opportunities in the data economy (Commission, 2022; Otto, 2022).

## 2.2 Common European Data Spaces

Clear rules support trust in data and consequently grow the data economy. Supporting the data economy is one of the critical objectives of the European Data Strategy, which can also be seen from the funds allocated for this purpose in the 2021-2027 financial perspective. Common European Data Spaces is an EU initiative for data sharing and exchanging. To maximize data's value for Europe's economy and society, the European Commission defined fourteen data areas that are key to the development of the society and the economy (Scerri et al., 2022): agriculture, cultural heritage, energy, finance, green deal, health, language, manufacturing, media, mobility, public administration, research and innovation, skills, and tourism.

These DSs will eventually interconnect, forming a unified market for data. The key features of a common European DS can be summarized as follows (Commission, 2022):

- clear and structured rules for access and use of data,
- secure and privacy-preserved data pooling, accessing, using, and sharing,
- fairness and transparency in non-discriminatory usage and access of data to all,
- reliable data governance mechanisms, and
- promotion of EU values and freedoms with an emphasis on data and consumer protection.

# 3 The data space concept

In continuation, our understanding of generalised DS concept is presented with its components. Fig. 1 provides a design of a DS that identifies the essential component necessary for building and managing it. It serves as a guideline to ensure interoperability, scalability, and efficiency in data integration and utilization. In continuation roles of main components are presented.

Data provider transfers the owner's data to the DS. The transfer is done by data connector that provides a gateway for data or services. DS connectors are usually configured using the Vocabulary services. They provide definition of data used in ds in the form of ontologies and specifications for semantic interoperability. Similarly, Broker provides information in form of metadata about content, structure and other features of the data to the data connector. Data connector has implemented capabilities which ensure trusted and secured environment for participates in DS communication. These capabilities are provided by identity provider. It is in charge for trusted and secured environment for participates by enabling services for providing and managing credentials for trust among the participants of the system. Data consumer process data

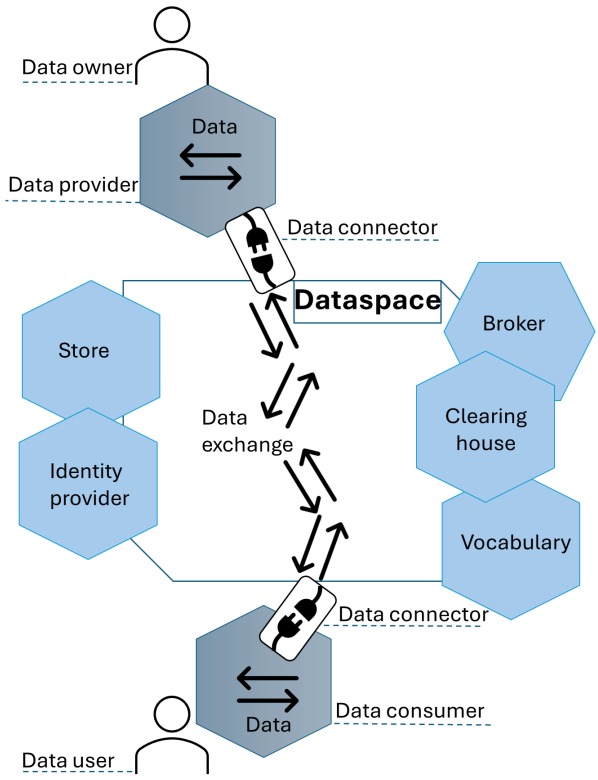



**Figure 1:** Components of a DS

that data user requests. Requested data is offered by data provider in term of they policies and with ensuring that data is quality and reliable. Services that handle the transaction between provider and consumer, including contracting, logging, and data transfer are provided by the Data exchange, while Clearing house services provide clearing and settlement service for all data exchange and financial transactions if data provider's policy request. App provider provides applications that can be deployed at the connector end to execute tasks like data transformation and aggregation, data analytics, machine learning or federated learning (Farahani and Monsefi, 2023).

# 4 Data space opportunities for Croatia

While EU data strategy, legal framework and financial instruments are set, Croatia is standing at a historic crossroads to adopt these directions for digital transformation and data economy. The adoption the EU data strategy by the economy, research organizations, and public service are path to the establishment of data economy, economic growth and progress of civil society. For this reason, it is necessary to identify opportunities for key stakeholders in this process: economy, science and public services. The emphasis for each stakeholder regarding the adoption of the EU data strategy that defines the sharing and use of data brings certain opportunities or added value.

On the other hand, according to the National Development Strategy of the Republic of Croatia, until 2030, digital transformation has been identified as a key element for developing the economy and public administration. Within the strategy, one of the development directions is 'Green and digital transition' which includes the strategic goal 'Digital transition of society and economy'. This strategy emphasizes the importance of digitization as a means of achieving efficiency, innovation, and competitiveness in the economy, as well as improving the quality of public services. In particular, the need to develop infrastructure and capacity for data management is highlighted, which includes the establishment of a system for transparent and efficient data management at the national level.

## 4.1 Opportunities for public services

Besides the corruption and nepotism, the main problem of Croatian public administration is it's inefficiency and lack of transparency (Kopric, 2019). Establishing a single, public DS where citizens can access to the government data could enhance transparency and boost confidence in public institutions. Furthermore, using sophisticated data analysis for public financial data and transaction could help identify and prevent corrupt practices. DS can also assist in improving coordination and information sharing between different government departments and agencies, as well as integrating data from various governmental agencies, enabling Croatia to provide more state-of-the-art services to its citizens and business.

For instance, developing a health DS for patient records integration can enhance healthcare delivery and patient outcomes. This system could integrate health records from hospitals, clinics, and pharmacies. Analyzing health trends and outcomes could improve public health interventions' effectiveness and resource allocation. Likewise, a cohesive DS for government services could significantly enhance operational efficiencies concerning licensing and permits, particularly within the construction industry. This optimization can lead to more strategic infrastructure investments, while DS for citizen services can considerably reduce bureaucratic procedures across a multitude of services (e.g., the process of business registration or the application for social benefits). Adopting smart city technologies, such as Internet of Things (IoT) sensors coupled with DS, can revolutionize urban infrastructure management, ranging from traffic coordination and waste management to optimizing energy consumption. This integration of technology into city planning and operations aims not only to boost the quality of life for its inhabitants by making cities more efficient and sustain-

able but also to reduce operational costs significantly. Furthermore, by enhancing the overall livability and efficiency of urban areas, these innovative solutions have the potential to attract increased investment into cities. Such advancements promise to pave the way for more resilient, eco-friendly, and smarter urban environments in the future. Policy-makers have the opportunity to harness the power of aggregated and anonymous data for guiding decisions in various vital sectors such as public health, education, and urban planning. By leveraging this data, they can not only gain insights into current conditions but also anticipate future challenges. These challenges could range from demographic shifts and economic fluctuations to environmental concerns. Utilizing this data-driven approach enables policy-makers to effectively prepare and strategize for these upcoming changes, ensuring more resilient and responsive governance.

### 4.2 Opportunities for the economy

DSs have the potential to revolutionize industries by fostering a more level playing field for businesses and streamlining the exchange of data securely through well-established set of rules. In the context of Croatia, which profoundly values conventional wisdom, deploying DS in the tourism sector could be particularly transformative. Such systems can facilitate highly targeted marketing strategies by collating data on tourists' preferences, travel habits, hotel occupancy figures, and local attractions. Furthermore, they can significantly enhance the tourist experience and allow for a more effective allocation of resources within the tourism industry. Conversely, Croatia's agricultural sector has witnessed a decline in production despite receiving substantial subsidies (Banski and Kaminska, 2022). Here, a DS that assembles critical data on weather trends, soil quality, crop health, and prevailing market rates could be a game-changer for farmers. This information hub can drive agricultural productivity and sustainable practices forward by enabling precision agriculture and improving market access.

Croatia boasts significant potential for green energy development (Schneider et al., 2007; Smajla et al., 2020), while DS platforms can play a pivotal role in disseminating information regarding energy consumption, the efficiency of energy use, and opportunities within renewable energy sources. This dissemination is essential for encouraging the growth of businesses and stimulating investments in the green energy sector. In addition, Croatia has a strategic location in the corridors of Europe's transportation, including its well-developed Adriatic seaports and excellent connections to major European cities. There is an enormous opportunity for DS logistics companies to offer access to real-time information concerning shipping routes, current traffic conditions, and the availability of warehousing. This would play a significant role in enhancing the efficiency of logistics operations, reducing transportation-related expenses, and ultimately strengthening the resilience of supply chains.

### 4.3 Opportunities for science and research

A centralized data repository in form of DS can serve as a collaborative hub, enabling researchers from different institutions in Croatia to share, access, and leverage an extensive range of datasets, academic publications, and various research outputs. This innovative platform is instrumental in stimulating interdisciplinary research partnerships. Such collaborations pave the way for accelerated scientific advancements by removing barriers to data access and reuse. By facilitating this level of openness, the platform promotes the principles of data sharing and reuse, acknowledging these practices as pivotal drivers for enhancing the overall impact and efficiency of research endeavours. A DS can promote data sharing among Croatian researchers and their international colleagues throughout the EU and beyond. This can bolster global research partnerships while ensuring adherence to international standards and practices. Overall quality of Croatian research and sciencs, leveraging cross-border knowledge and international evaluation. This approach not only aligns with global research trends but also significantly contributes to the advancement of scientific and academic pursuits across various fields.

Furthermore, DS can rid the gap between academia and the commercial sector by offering an expansive repository of information, including patents, cutting-edge technologies, and detailed insights into current market demands. This can optimize the technology transfer process, promoting a synergistic environment that supports innovation by connecting academic research with practical industrial needs. Moreover, researchers can convert scientific findings into viable commercial solutions, accelerating technological advancement and economic growth.

## 5 Conclusion

Croatia's integration of the EU data strategy presents numerous opportunities for catalyzing digital transformation across public services, the economy, and broader civil society. For public services, implementing data spaces can enhance transparency, efficiency, and collaboration between government entities. This approach not only aims to restrain corruption and improve service delivery but also sets the foundation for the economy of data. Economically, DS can level the playing field for businesses, particularly in tourism, agriculture, and green energy, by enabling data-driven decision-making, optimizing resource allocation, and opening new avenues for sustainable practices. By embracing these opportunities, Croatia stands on the brink of a transformative shift, poised to establish a

robust data economy that promotes economic growth, enhances public services, and secures a resilient, prosperous future for its citizens.

# 6 Acknowledgement

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

Banski, J. and Kaminska, W. (2022). Trends for agricultural land-use in the ceecs following the collapse of the eastern bloc. *Land Use Policy*, 112:1–10.

Boukhers, Z., Lange, C., and Beyan, O. (2023). Enhancing data space semantic interoperability through machine learning: a visionary perspective. In *Companion Proceedings of the ACM Web Conference 2023*, WWW 2023 Companion, pages 1462–1467, New York, NY, USA. Association for Computing Machinery.

Commission, E. (2022). Common european data spaces. *EC Staff Working Document*, COM/2020/66 final. Available at `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066`.

Constantiou, I. D. and Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, 30:44–57.

Farahani, B. and Monsefi, A. K. (2023). Smart and collaborative industrial iot: A federated learning and data space approach. *Digital Communications and Networks*, 9(2):436–447.

Franklin, M., Halevy, A., and Maier, D. (2005). From databases to dataspaces: a new abstraction for information management. *SIGMOD REC.*, 34(4):27–33.

Kopric, I. (2019). Public administration reform in croatia: Slow modernization during europeanization of resilient bureaucracy. *Public Administration Issues*, pages 7–26.

Kotsev, A., Minghini, M., Tomas, R., Cetl, V., and Lutz, M. (2020). From spatial data infrastructures to data spaces: A technological perspective on the evolution of european sdis. *ISPRS International Journal of Geo-Information*, 9(3):176.

Manoochehri, M. (2013). *Data just right: introduction to large-scale data & analytics*. Addison-Wesley, 1 edition.

Otto, B. (2022). The evolution of data spaces. In *Designing data spaces: The ecosystem approach to competitive advantage*, page 580. Springer International Publishing Cham.

Scerri, S., Tuikka, T., Vallejo, I., and Curry, E. (2022). *Common European Data Spaces: Challenges and Opportunities*, pages 337–357.

Schneider, D. R., Duic, N., and Bogdan, Z. (2007). Mapping the potential for decentralized energy generation based on renewable energy sources in the republic of croatia. *Energy*, 32(9):1731–1744.

Smajla, I., Crnekovic, R., Sedlar, D., and Bozic, F. (2020). Potential of croatian liquefied natural gas (lng) terminal in supplying regional natural gas markets. *Rudarsko-geolosko-naftni zbornik*, 35:93–101.

Vlahek, D. and Mongus, D. (2023). An efficient iterative approach to explainable feature learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(5):2606–2618.

Zillner, S., Gomez, J., GarcÃa Robles, A., Hahn, T., Bars, L., Petkovic, M., and Curry, E. (2021). *Data Economy 2.0: From Big Data Value to AI Value and a European Data Space*, pages 379–399.

# Social Robots: Applications, Challenges, and Future Directions - literature review

**Ivana Ružić**
University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42 000 Varaždin
I. osnovna škola Čakovec
Ulica kralja Tomislava 43, 40 000 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***. *The integration of AI, robotics, and digital connectivity is transforming various sectors, enhancing operations, reducing costs, and enriching customer interactions. Social robots, equipped with AI and advanced sensors, are prominent in fields such as healthcare, education, and entertainment, offering human-like interactions. This review examines the potential, challenges, and future directions of social robots by analysing experimental studies across diverse domains where social robots have been applied: healthcare, mental health, care for the elderly, psychology, hospitality industries, and other services such as the financial sector, cultural heritage, and retail. It identifies key issues and provides recommendations for effective and ethical implementation.*

**Keywords.** social robot, healthcare, industry

## 1 Introduction

The synergy between the evolution of artificial intelligence (AI), the advancement of robotics, and the proliferation of digital connectivity is profoundly reshaping every sector. Businesses are eagerly integrating these state-of-the-art technologies to streamline operations, cut costs, enrich customer interactions, and pioneer innovative products and services. Undoubtedly, the technological dimension is at the heart of service innovation. Robots, leveraging AI and sophisticated sensors, have permeated diverse industries, spanning manufacturing, search and rescue, entertainment, education, research, assistance, and healthcare. Among these, social (humanoid) robots stand out as a category designed to engage with individuals, enrich interactions, and elevate overall experiences (Spekman et al. 2020). With AI-driven capabilities and sensory awareness, these robots can perceive and respond to their environment, engaging with people in a manner akin to human social dynamics.

Characterized by their humanoid resemblance, featuring recognizable elements such as a head, torso, arms, and legs, social robots hold tremendous promise, particularly in the field of social robotics. Their goal is to seamlessly integrate into human interactions, exhibiting behaviours that foster communication, collaboration, and companionship (Perez-Zuniga et al. 2024). Robotics stands as a technological facet ingrained across diverse life domains (Kalaitzidou & Pachidis (2023)).

The main aim of this paper is to explore the transformative potential of social robots across various areas of life. Therefore, this review will assess the support for claims regarding the potential of social robots in everyday life through an analysis of experimental studies. It will also address the applications of social robots in different domains of human activity: healthcare, mental health, care for the elderly, psychology, hospitality industries, and other services such as the financial sector, cultural heritage, and retail, identify problems and limitations, and provide recommendations.

The paper outlines the general objective and research questions, details the materials and methods utilized, presents the search results, explores the application areas of social robots with significant examples, identifies key issues and provides future recommendations, as well as includes a thorough discussion and conclusion.

## 2 General Objective and Research Questions

This study aims to explore the deployment of social robots across various sectors such as healthcare, psychology, hospitality, and other services. It seeks to identify key research challenges, constraints, and provide recommendations. By understanding

technical, ethical, and societal issues, the study aims to offer insights and strategies for effective and ethical implementation, enhancing the functionality and acceptance of social robots in diverse environments.

The following research questions are at the centre of interest:

RQ1: In which areas of activity are there examples of using social robots in interaction with humans?

RQ2: Which research problems, recommendations and limitations have been identified in the selected papers?

# 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 in Fig. 1.

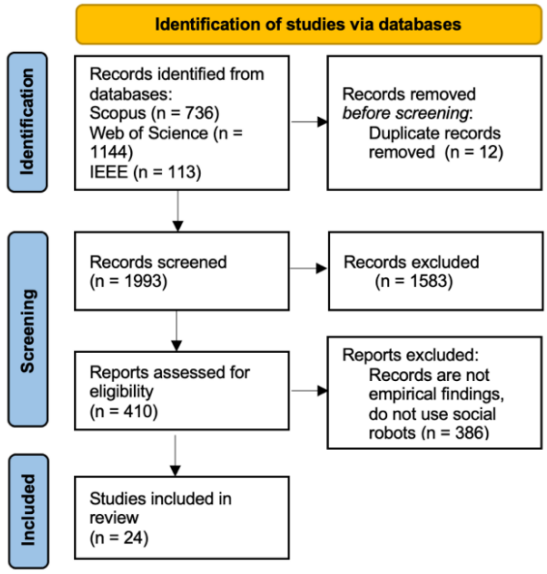



**Figure 1.** The literature review process

In the first identification stage, a structured search strategy was created to be used on the scientific databases IEEE, Scopus and Web of Science. Keywords: social, robot, healthcare and industry were used for database search. The literature review process was conducted using the following search strings:

TS = (social OR humanoid) AND robot* AND healthcare

TS = (social OR humanoid) AND robot* AND industry.

For the second stage, screening, 113 results from IEEE, 1144 results from Web of Science and 736 results from Scopus were identified. The following additional selection criteria were used:

- Published in English,
- Published within the time frame 2020-2024,
- Document type was article,
- Full text was available.

As a result, 47 studies were obtained from IEEE, 118 from Scopus and 257 from Web of Science. 12 duplicates were detected and excluded, and 410 studies were reviewed by titles and abstracts.

In the next stage, the final eligibility criterion was applied – studies should have included empirical findings on the use of social robots in the fields of healthcare and industry. All the studies that research humanoid hands and heads, robot design and performance, frameworks and recommendations have not been included. 386 studies were removed because they did not meet the required criteria. A total of 24 studies were included in the fourth stage of the literature review.

# 4 Results

The research results are presented in Table 1 and they show the type of robot used, the number of participants and their age, the field (healthcare, cultural heritage, psychology, services in commerce and tourism, financial sector, etc.) and the duration of the activity.

**Table 1.** Presentation of the research results

| Authors | Robot | Participants | The field of application | Duration |
|---|---|---|---|---|
| | | No / Type | | |
| Perez-Zuniga et al. (2024) | Qhali | 17 /university students | mental health | 1 session |
| Cantucci et al. (2023) | NAO | 84 / adults | cultural heritage | 1 session |
| Karunarathne et al. (2020) | Robovie | 20 / adults | healthcare | 2 sessions |
| Andtfolk et al. (2022) | Pepper | 264 /patients, relatives, professionals | healthcare | 30 min |
| Fan et al. (2021) | NAO | 26 / adults | care for elderly | 1 session |
| Nertinger et al. (2022) | GARMI | 166 / museum visitors | healthcare | 2 months |
| Kasimoglu et al. (2020) | iRobiQ | 200 / children | healthcare | - |
| Robinson & Kavanagh (2021) | NAO | 18 / adults | healthcare | 2 sessions (60 min) |
| McIntosh et al. (2022) | Pepper | 993 / adults | healthcare | 12 weeks (5 min) |
| Tanioka et al. (2021) | Pepper | 2 / adults | mental health | - |
| Yoshii et al. (2023) | Pepper | 94 / patients and adults | mental health | - |
| Liu et al. (2021) | - | 120/university students | psychology | 1 session |
| Leung et al. (2023) | Ka Ka | 4 / adults | care for elderly | 2 weeks |
| Spekman et al. (2020) | NAO | 211/university students | psychology | 1 session |
| Plotkina et al. (2024) | social robot | 429 / adults | financial sector | - |
| Mingotto et al. (2021) | Pepper | - / customers | hospitality industries | 22 sessions (1 hour) |
| Zhang et al. (2023) | Pepper | 280 / customers | hospitality industries | 6 months |

| | | | | |
|---|---|---|---|---|
| Okafuji et al. (2022) | Sota | 5000+ / pedestrians | mall | 3 weekends (6 hours) |
| Roozen et al. (2023) | Pepper | 537 / customers | store | - |
| Zhu & Chang (2020) | - | 221 / customers | service | - |
| Lo et al. (2022) | Zenbo | 77 / adults | sustainability | - |
| Spekman et al. (2020) | NAO | 101/university students | psychology | - |
| Krakovski et al. (2021) | NAO, Poppy | 26 / adults | psychology | - |
| Ko et al. (2023) | NAO, Pleo | 20 /university students | psychology | 2 sessions |

The results show that NAO and Pepper are the most popular social robots, with each used in seven studies. The robots Pleo, Poppy, Zenbo, Sota, Ka Ka, Robovie, Qhali and iRobiQ were used in one study each. In three studies, it was not specified which social robot was used. The number of participants who actively participated in the research ranges from 2 to as many as 5000. In one study, the number of participants was not specified. Studies are most often conducted among adults: university students, customers, pedestrians, museum visitors, older adults, patients, relatives, and care professionals. Only one study was conducted among children. When we observe the application areas of social robots, we can notice that six studies are from the field of healthcare, five studies are from the field of psychology, three studies are from the field of mental health, and two studies each are from the fields of care for the elderly and hospitality industries. Other services such as the financial sector, cultural heritage, and retail are represented by five studies. Considering the duration of interactions with social robots, both short-term and long-term studies were conducted. Nine studies involved interactions with the robot once or twice. Five studies conducted activities with the robot over several weeks up to 6 months, while in nine studies, the duration of the interaction with the robot was not specified.

## 4.1 Application Areas of Social Robots with Examples

The application of social robots and their interaction with humans has been extensively researched across various contexts, highlighting their significant potential. Below, we outline some cases documented in the literature based on the domains where social robots have been applied: healthcare, mental health, care for elderly, psychology, hospitality industries and other services such as financial sector, cultural heritage, and retail.

### 4.1.1 Healthcare

Karunarathne et al. (2020) conducted an empirical study on older adults, comparing their experiences walking alone and with a robot. Attitudes towards social robots in healthcare among various stakeholders, analysing correlations with demographic characteristics, were investigated by Andtfolk et al. (2022). Nertinger et al. (2022) used a social robot to explore acceptance of caregiving tasks, considering socio-demographics, user beliefs, and robot autonomy. The application of a social robot to reduce children's anxiety during dental treatment was in the centre of interest by Kasimoglu et al. (2020). Robinson & Kavanagh (2021) evaluated a social robot designed to promote health behaviour change related to food intake and weight loss. The impact of a social robot on knowledge of influenza prevention and attitudes toward vaccination were the focus of the research by McIntosh et al. (2022).

In the experiment conducted by Karunarathne et al. (2020) results indicated that participants gave significantly higher ratings to the intention of walking with the robot compared to walking alone. In their research, Andtfolk et al. (2022) found that the majority of participants expressed favourable views on the use of social robots in healthcare, with only a minority holding negative opinions. Among other stakeholders, including healthcare service providers, politicians, individuals with higher education levels, and older adults, attitudes towards social robots were predominantly positive. Nertinger et al. (2022) compiled all pertinent factors regarding acceptance to guide the user-cantered design process of assistive robots. Their findings indicate that, among other factors, trust in the robot and utilitarian variables like perceived usefulness are the most influential factors in robot acceptance. Kasimoglu et al. (2020) demonstrated that the use of robotic technology holds promise in effectively addressing dental anxiety and stress, resulting in improved behaviour among children in dental settings.

Study conducted by Robinson & Kavanagh (2021) indicates that the robot-delivered intervention was found to be helpful and user-friendly, particularly after users became accustomed to interacting with a social robot and experienced initial benefits. Social humanoid robots seem suitable as health coaches for adults, especially for behaviour change. The study conducted by McIntosh et al. (2022) has shed light on the effectiveness of a social humanoid robot in enhancing individuals' understanding of influenza prevention and influencing their attitudes toward influenza vaccination. Following interactions with the humanoid robot, participants experienced immediate increases in knowledge and shifts in attitudes, suggesting that social robots could play a significant role in health promotion efforts related to influenza prevention.

### 4.1.2 Mental Health

Perez-Zuniga et al. (2024) reviewed the expressive social robot Qhali, focusing on its design, components, and validation for telepsychological interventions. The experiences of older patients with schizophrenia and dementia interacting with healthcare social robots and intermediaries were explored by Tanioka et al. (2021). Yoshii et al. (2023) who studied early detection of mild

cognitive impairment through patient-robot conversations, bypassing neuropsychological exams.

The study conducted by Perez-Zuniga et al. (2024) indicates that participants not only experienced improvements in their emotional well-being but also held positive perceptions regarding the psychological intervention facilitated by the social robot. The robot is highly efficient in delivering telepsychological interventions, minimizing the sense of threat, and ensuring engagement with users (Perez-Zuniga et al. (2024). Research findings conducted by Yoshii et al. (2023) indicate the potential to detect patients with mild cognitive impairment through their everyday conversations with a social robot, thereby enhancing the effectiveness of a simple dementia screening test. The study results conducted by Tanioka et al. (2021) suggest that interactions with robots can bring moments of joy to older individuals with schizophrenia and dementia, highlighting the value of healthcare robots in these settings.

#### 4.1.3 Care for the Elderly

Leung et al. (2023) studied humanoid social robots in nursing homes and their interactions with older adults. The socially assistive robot designed to enhance activity and social engagement in older adults were researched by Fan et al. (2021).

Findings from a laboratory-based study involving older adults demonstrate the capabilities of socially assistive robotic, revealing that this system can enable one or multiple older adults to engage in multidomain activities while receiving dynamic guidance; facilitate their participation in robot-mediated tasks and promote human-to-human interaction, and assess their social and activity engagement using multiple sensory modalities (Fan et al. (2021)). In their study, Leung et al. (2023) highlighted the benefits of social robots, including offering emotional support to older adults living alone, adding variety to their daily routines, and strengthening family bonds. The voice of the robot, described as female, soft, and soothing, was perceived as bringing a sense of comfort to older adults.

#### 4.1.4 Psychology

Spekman et al. (2020) examined how emotions and emotional coping influence perceptions of social robots. The impact of human-robot proxemics on concentration-training games with social robots was investigated by Liu et al. (2021). Ko et al. (2023) explored user perceptions of robots in conversational tasks, varying their voice types, appearances, and expressions. The focus of the research conducted by Krakovski et al. (2021) were social robots and how they can fulfil the physical and cognitive training needs of older adults.

The study results conducted by Spekman et al. (2020) revealed an interaction effect between prior emotions and the manipulated coping potential on robot perceptions, contrary to the expected effects from previous studies. Interacting with a robot elicited different reactions, overriding any emotional effects. The findings of Liu et al. (2021) indicate that a distance of 2 meters and a left-front orientation between a human and a robot are optimal for interactive concentration training between humans and robots. Furthermore, females demonstrated superior performance compared to males in human-robot interaction imitation games.

Research findings conducted by Krakovski et al. (2021) revealed that acceptance of the robotic system was influenced by age, attitude, and education. This underscores the significance of tailoring the system to diverse user needs and the value of meaningful feedback. The system exhibited robustness and reliability, showcasing its potential as a personal trainer and a source of motivation for older adults. The research findings conducted by Ko et al. (2023) suggest that accuracy in perceiving emotions varied depending on the emotions presented; participants preferred a regular human voice for its naturalness; however, a characterized voice was more effective in conveying emotions with significantly higher accuracy in emotion perception, and participants exhibited significantly heightened emotion perception.

#### 4.1.5 Hospitality Industries

Mingotto et al. (2021) focused on the evolving roles of frontline employees and customers due to technology adoption, particularly AI-powered conversational agents and robots in tourism companies. In their research Zhang et al. (2023) examined consumer responses to service failures by social versus non-social robots, and how these responses affect brand forgiveness and revisit intentions, considering performance expectations across genders.

The research results conducted by Mingotto et al. (2021) indicate that conversational agents and robots in tourism can serve as an augmentative force, leading to the evolution of frontline employees (FLEs) primarily into enablers - both for customers and technology - as well as innovators and coordinators. Meanwhile, customers may primarily assume the role of enabling the technology. Zhang et al. (2023) in their study demonstrated that consumers exhibit higher performance expectations for non-social robots. These expectations lead to brand forgiveness and revisit intentions among male consumers, while they do not significantly impact forgiveness and revisit behaviours among female consumers.

#### 4.1.6 Other Services

Cantucci et al. (2023) examined user satisfaction during interactions with a robot guided by a computational cognitive model incorporating adjustable social autonomy principles. In the financial sector, Plotkina et al. (2024) investigated how the anthropomorphism and gender of robo-advisors influence social presence, allowing consumers to assess personality traits like competence, warmth, and persuasiveness, thereby contributing to trust. Okafuji et

al. (2022) explored the use of a robot as a social service provider in a shopping mall, examining various robot behaviours. In their research, Roozen et al. (2023) compared perceived service quality in human-robot and human-human interactions in a retail store through hypothetical scenarios, also considering the moderating effects of participants' attitudes towards robots, age, gender, and education level. Zhu & Chang (2020) studied how the anthropomorphism of robotic chefs affects the perception of food quality based on warmth and competence. In their study, Lo et al. (2022) examined the effectiveness of robots in persuading people to recycle through education and encouragement.

Cantucci et al. (2023) found that as the robot's autonomy in task adoption rose, users' satisfaction with the robot declined, yet their satisfaction with the tour itself enhanced. These results underscore the promise of adjustable social autonomy as a framework for creating autonomous adaptive social robots that can enhance user experiences across various real-world domains of human-robot interaction. Plotkina et al. (2024) observed that humanized avatars instil greater trust compared to both gender-neutral cartoonish and anthropomorphized cyborg robo-advisors. This trust is attributed to higher perceived competence and persuasiveness, particularly significant for male robo-advisors, confirming the significance of gender in financial digital services, as hypothesized. Additionally, findings revealed an uncanny valley effect for anthropomorphized cyborg representations of robo-advisors, unlike gender-neutral cartoonish ones. Moreover, highlighting the gender-neutral personality of the robo-advisor "de-humanizes" it and diminishes the positive effect of anthropomorphism on trust.

The findings of Okafuji et al. (2022) suggest that robots' performance in providing information tasks is comparable to humans in controlled environments, indicating their potential effectiveness as labour-support technology in real-world settings. In the study conducted by Roozen et al. (2023) participants rated service quality higher in human-human interactions compared to human-robot interactions. Among those with positive attitudes towards robots, there was no significant difference in perceived service quality between the two interactions. However, participants with lower positive attitudes towards robots rated service quality lower in human-robot interactions. Age, gender, and educational level did not significantly influence participants' perceived service quality in either interaction.

Zhu & Chang (2020) demonstrated that anthropomorphism of robotic chefs influences food quality prediction, mediated sequentially by perceptions of warmth and competence. Age is included as a significant control variable. The results of the study conducted by Lo et al (2022) suggests that due to their anthropomorphic features, robots are more capable of eliciting empathy than tablet computers, potentially making them more effective in encouraging pro-social behaviour.

## 4.2 Identified Issues and Future Recommendations

In their empirical studies, the authors identified key issues and considerations regarding the deployment and integration of social robots. They provided detailed recommendations and outlined plans for future research to address these challenges. Their findings aim to guide further development and ensure effective, ethical, and practical use of social robots across various sectors, enhancing their functionality and acceptance in real-world applications.

### 4.2.1 Identified Issues and Considerations

There is concern about the generalizability of research findings due to potential variations with different robots, contexts, and communication abilities (Karunarathne et al., 2020; Andtfolk et al., 2022; Zhang et al., 2023; Roozen et al., 2023).

Future research should involve participants from diverse age groups, socioeconomic statuses, and cognitive levels across various countries, considering cultural differences in attitudes toward technology and gender (Leung et al., 2023; Plotkina et al., 2024).

Several authors highlight the significant cost of social robot systems as a primary drawback. Damage to these robots, whether hardware or software-related, requires examination by qualified specialists, and maintaining extensive robotic systems demands substantial budgets (Kasimoglu et al., 2020; Leung et al., 2023; Plotkina et al., 2024; Mingotto et al., 2021).

Concerns also include issues related to data privacy, increased intrusiveness, biases among users, and the use of unregulated mechanisms. It is essential to investigate these potential negative aspects of robot services (Plotkina et al., 2024).

### 4.2.2 Recommendations and Plans for Future Research

Andtfolk et al. (2022) underscore the need for further research into the root causes of negative attitudes that could hinder the integration of social robots in healthcare.

Future plans include integrating measures of activity engagement and social interaction to enhance real-time interpersonal interactions and task engagement, thereby improving the adaptive behaviours of robots (Fan et al., 2021).

Further research could explore customers' negative experiences with service robots and measure outcomes (Roozen et al., 2023). Zhu & Chang (2020) emphasize the importance of investigating how anthropomorphism in social robots influences perceptions of their competence.

Longitudinal studies could provide insights into the ways customer attitudes and perceptions of service

quality evolve over time as they become more familiar with service robots (Roozen et al., 2023).

There is a need to enhance technical robustness and individualize robot-delivered interventions, which includes improving the robots' communication abilities to paraphrase sentences, respond to interpersonal cues, and emulate client-centred counselling techniques (Robinson & Kavanagh, 2021).

Future research aims to enhance the linguistic capabilities of robots. Several studies are constrained by robots' challenges in natural interactions with pedestrians in noisy environments, often due to low speech recognition accuracy—a common issue in real-world robot operations. Overcoming these challenges requires improving the robot's ability to accurately recognize speech amidst noise (Yoshii et al., 2023; Okafuji et al., 2022).

Additionally, future experimental designs should incorporate more interactive experiences with real applications. While their study confirmed a significant positive relationship between trust in social robots and intention to use them, exploring other consumer perceptions such as confidence in their decisions and intentions to act on recommendations would provide further insights (Plotkina et al., 2024). Although the intention to use typically predicts actual usage, this relationship needs validation through direct user studies. Given that some functionalities are not fully developed for implementation, these findings will guide future research on user readiness (Nertinger et al., 2022).

The potential influence of novelty is challenging to mitigate. Further research is needed to determine whether this persuasive effect persists over time, beyond the initial decline in user curiosity (Lo et al., 2022).

# 5 Discussion

The study identified several areas of activity with significant examples of using social robots in healthcare, mental health, care for the elderly, psychology, hospitality industries, and other services. Additionally, the study highlighted research problems, provided recommendations, and stated limitations found in the selected papers. Here it should be noted that NAO and Pepper are very popular social robot in the context of interactions with humans (e.g. Cantucci et al. (2023), Fan et al. (2021), Andtfolk et al. (2022), Tanioka et al. (2021), Yoshii et al. (2023), etc.). The range of participants varied widely, and the predominant focus was on adults. In terms of research issues, the centre of interest for researchers was human-robot interaction.

Many papers (11) confirmed the positive relationship between human-robot interaction, behavior and attitudes. For example, Andtfolk et al. (2022) found that the majority of participants expressed favorable views on the use of social robots in healthcare. The study by Perez-Zuniga et al. (2024) indicates that participants experienced improvements in their emotional well-being and had positive perceptions of the psychological intervention facilitated by the social robot. Leung et al. (2023) emphasized the benefits of social robots, such as providing emotional support to older adults living alone, introducing variety into their daily routines, and strengthening family bonds.

Successful interaction between humans and robots is significantly influenced by the characteristics of both the humans and robots, as well as their spatial relationship. The acceptance of the robotic system was influenced by factors such as age, attitude, and education what underscores the importance of tailoring the system to diverse user needs and the value of obtaining meaningful feedback (Krakovski et al. (2021)). Liu et al. (2021) indicate that a distance of 2 meters and a left-front orientation between a human and a robot are optimal for interactions between humans and robots. The voice of the robot also plays a significant role; Ko et al. (2023) found that participants preferred a regular human voice due to its naturalness, while Plotkina et al. (2024) highlighted that the gender-neutral personality of the robo-advisor "dehumanizes" it and diminishes the positive effect of anthropomorphism on trust.

In terms of main findings, it should be noted that most of the research had relatively small samples, which means that some results should be confirmed by others. Additionally, most studies are short-term, with participants interacting with the robot for a short period or even just once.

We can note that authors highlight concerns about research generalizability due to robot, context, and communication variations persist (Karunarathne et al., 2020; Andtfolk et al., 2022; Zhang et al., 2023; Roozen et al., 2023). Future studies should diversify participants by age, socioeconomic status, and cognitive level across cultures (Leung et al., 2023; Plotkina et al., 2024). Additionally, high costs, maintenance demands, privacy risks, and biases are significant drawbacks emphasized by Kasimoglu et al., 2020; Leung et al., 2023; Plotkina et al., 2024; Mingotto et al., 2021.

We observe that the authors emphasize the need to investigate barriers to robot integration (Andtfolk et al., 2022), the need to improve real-time interactions and task engagement, which could enhance robot adaptability (Fan et al., 2021). Additionally, several authors mention the negative experiences of users interacting with robots and the need to understand these experiences, as well as the importance of exploring the impact of anthropomorphism on the perception of robot competence (Roozen et al., 2023; Zhu & Chang, 2020).

It can be observed that several authors highlight the importance of conducting longitudinal studies to track evolving attitudes and user perceptions of the quality of services provided by robots (Roozen et al., 2023).

These studies could also address novelty effects and provide more information on the long-term impact and intention to use robots (Lo et al., 2022; Plotkina et al., 2024; Nertinger et al., 2022). Additionally, the authors emphasize the importance of improving the technical capabilities of robots and personalizing robot interventions, including enhancing linguistic abilities, especially in real-world and noisy environments (Robinson & Kavanagh, 2021; Yoshii et al., 2023; Okafuji et al., 2022).

While the integration of social robots into various aspects of human activities offers promising opportunities, it also brings significant ethical and data privacy challenges. A major concern is that none of the analyzed papers address these issues, even though some acknowledge their importance. For example, Boch et al. (2020) identified several risks, including lack of transparency, data privacy issues, robot dependency, diminished human interaction, and potential job displacement, all of which require careful consideration. These challenges underscore the need for the responsible and safe deployment of social robots. Data privacy is a critical concern, as noted by Boch et al. (2020), necessitating clear protocols for data collection, processing, storage, and informed consent. Newton & Newton (2019) emphasize the importance of evaluating the appropriateness of the assumptions, values, and beliefs embedded in robots' instructional methods. Proper management and secure disposal of collected data, whether handled by robots or humans, are essential for maintaining privacy and safety. The creation of specific regulations to address new technological features that facilitate enhanced data collection highlights the need for ongoing policy development. These issues point to the necessity for developing regulations and public policies for social robots, a topic that will be crucial for future study and research (Subramanian, 2017). Collaboration among policymakers, developers, and privacy experts is essential to establish clear guidelines and policies that prioritize human well-being and rights (Boch et al., 2020).

# 6 Conclusion

The synergy between the evolution of artificial intelligence (AI), robotics, and digital connectivity is profoundly reshaping numerous sectors. Businesses are integrating these technologies to streamline operations, reduce costs, enhance customer interactions, and innovate products and services. A key facet of this technological transformation is the rise of social robots, designed to engage with individuals, enrich interactions, and elevate experiences across various domains. These robots, equipped with AI-driven capabilities and sensory awareness, can perceive and respond to their environment, engaging with people in a human-like manner. Social robots, characterized by their humanoid features such as heads, torsos, arms, and legs, hold significant promise, particularly in social robotics. Their design aims to seamlessly integrate into human interactions, exhibiting behaviors that foster communication, collaboration, and companionship.

The transformative potential of social robots is being explored in various sectors, including healthcare, mental health, care for the elderly, psychology, hospitality, and other services like the financial sector and cultural heritage. In healthcare, social robots have been used to reduce children's dental anxiety, promote health behavior changes, and enhance influenza prevention knowledge. Mental health applications include facilitating telepsychological interventions and early detection of cognitive impairments. For elderly care, robots offer emotional support and engage older adults in multidomain activities. In psychology, research focuses on robots' influence on emotional coping and concentration training. In the hospitality industry, robots serve as conversational agents, influencing consumer responses to service failures and shaping frontline employee roles.

Despite their potential, social robots face several challenges, including high costs, maintenance requirements, data privacy concerns, and issues related to user biases and unregulated mechanisms. Future research should address these challenges, explore the root causes of negative attitudes towards robots, and enhance robots' technical robustness and communication abilities. Longitudinal studies are needed to understand how user perceptions evolve over time, and interactive experiences with real applications should be incorporated to validate findings. Addressing these challenges and recommendations will ensure the effective and ethical integration of social robots, enhancing their functionality and acceptance in various environments.

# References

Andtfolk, M., Nyholm, L., Eide, H., Rauhala, A., & Fagerström, L. (2022). Attitudes toward the use of humanoid robots in healthcare—a cross-sectional study. AI & SOCIETY, 37(4), 1739-1748.

Boland, A., Cherry, G., & Dickson, R. (2017). Doing a Systematic Review: A Student's Guide. Sage.

Cantucci, F., Falcone, R., & Marini, M. (2023). Redefining User Expectations: The Impact of Adjustable Social Autonomy in Human–Robot Interaction. Electronics, 13(1), 127.

Fan, J., Mion, L. C., Beuscher, L., Ullal, A., Newhouse, P. A., & Sarkar, N. (2021). SAR-connect: a socially assistive robotic system to support activity and social engagement of older adults. IEEE Transactions on Robotics, 38(2), 1250-1269.

Kalaitzidou, M., Pachidis, T.P. (2023). Recent Robots in STEAM Education. Educ. Sci. 13, 272. doi: 10.3390/educsci13030272

Kasimoglu, Y., Kocaaydin, S., Karsli, E., Esen, M., Bektas, I., Ince, G., & Tuna, E. B. (2020). Robotic approach to the reduction of dental anxiety in children. Acta Odontologica Scandinavica, 78(6), 474-480.

Karunarathne, D., Morales, Y., Nomura, T., Kanda, T., & Ishiguro, H. (2020). Will older adults accept a humanoid robot as a walking partner?. International Journal of Social Robotics, 11, 343-358.

Ko, S., Barnes, J., Dong, J., Park, C. H., Howard, A., & Jeon, M. (2023). The effects of robot voices and appearances on users' emotion recognition and subjective perception. International Journal of Humanoid Robotics, 20(01), 2350001.

Krakovski, M., Kumar, S., Givati, S., Bardea, M., Zafrani, O., Nimrod, G., ... & Edan, Y. (2021). "Gymmy": designing and testing a robot for physical and cognitive training of older adults. Applied Sciences, 11(14), 6431.

Leung, A. Y., Zhao, I. Y., Lin, S., & Lau, T. K. (2022, December). Exploring the presence of humanoid social robots at home and capturing human-robot interactions with older adults: Experiences from four case studies. In Healthcare (Vol. 11, No. 1, p. 39). MDPI.

Liu, L., Liu, Y., & Gao, X. Z. (2021, July). Impacts of human robot proxemics on human concentration-training games with humanoid robots. In Healthcare (Vol. 9, No. 7, p. 894). MDPI.

Lo, S. Y., Lai, Y. Y., Liu, J. C., & Yeh, S. L. (2022). Robots and sustainability: robots as persuaders to promote recycling. International Journal of Social Robotics, 14(5), 1261-1272.

McIntosh, C., Elvin, A., Smyth, W., Birks, M., & Nagle, C. (2022). Health promotion, health literacy and vaccine hesitancy: the role of humanoid robots. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 59, 00469580221078515.

Mingotto, E., Montaguti, F., & Tamma, M. (2021). Challenges in re-designing operations and jobs to embody AI and robotics in services. Findings from a case in the hospitality industry. Electronic Markets, 31(3), 493-510.

Nertinger, S., Kirschner, R. J., Naceri, A., & Haddadin, S. (2022). Acceptance of remote assistive robots with and without human-in-the-loop for healthcare applications. International Journal of Social Robotics, 1-20.

Okafuji, Y., Ozaki, Y., Baba, J., Nakanishi, J., Ogawa, K., Yoshikawa, Y., & Ishiguro, H. (2022). Behavioural assessment of a humanoid robot when attracting pedestrians in a mall. International journal of social robotics, 14(7), 1731-1747.

Pérez-Zuñiga, G., Arce, D., Gibaja, S., Alvites, M., Cano, C., Bustamante, M., ... & Cuellar, F. (2024). Qhali: A Humanoid Robot for Assisting in Mental Health Treatment. Sensors, 24(4), 1321.

Plotkina, D., Orkut, H., & Karageyim, M. A. (2024). Give me a human! How anthropomorphism and robot gender affect trust in financial robo-advisory services. Asia Pacific Journal of Marketing and Logistics.

Robinson, N. L., & Kavanagh, D. J. (2021). A social robot to deliver a psychotherapeutic treatment: Qualitative responses by participants in a randomized controlled trial and future design recommendations. International Journal of Human-Computer Studies, 155, 102700.

Roozen, I., Raedts, M., & Yanycheva, A. (2023). Are retail customers ready for service robot assistants?. International Journal of Social Robotics, 15(1), 15-25.

Spekman, M. L., Konijn, E. A., & Hoorn, J. F. (2021). How physical presence overrides emotional (coping) effects in hri: Testing the transfer of emotions and emotional coping in interaction with a humanoid social robot. International Journal of Social Robotics, 13(2), 407-428.

Tanioka, T., Betriana, F., Yokotani, T., Osaka, K., Locsin, R. C., King, B., & Schoenhofer, S. (2021). The experience of older persons with mental health conditions who interact with healthcare robots and nurse intermediaries: The qualitative case studies. Belitung Nursing Journal, 7(4), 346.

Yoshii, K., Kimura, D., Kosugi, A., Shinkawa, K., Takase, T., Kobayashi, M., ... & Nishimura, M. (2023). Screening of mild cognitive impairment through conversations with humanoid robots: Exploratory pilot study. JMIR Formative Research, 7(1), e42792.

Zhang, M., Cui, J., & Zhong, J. (2023). How consumers react differently toward humanoid vs. nonhumanoid robots after service failures: a moderated chain mediation model. International Journal of Emerging Markets.

Zhu, D. H., & Chang, Y. P. (2020). Robot with humanoid hands cooks food better? Effect of robotic chef anthropomorphism on food quality prediction. International Journal of Contemporary Hospitality Management, 32(3), 1367-1383.

# Learning Analytics, AI and Technologies in Education

Martina Holenko Dlab, Gordan Đurović, Marko Horvat
**Perception and Use of GenAI Chatbots in Education by Students at the University of Rijeka**

Patrik Komljenović, Mia Čarapina
**Tablet Math Application for Supporting Independent Simultaneous Interaction of Colocated Users**

Sonja Nidogon Višnjić
**The relationship between student work and academic success: a case study**

Marko Matus, Igor Balaban
**Relationship between factors that contribute to student success in the e-learning environment**

Barbi Svetec, Blaženka Divjak, Damir Horvat, Katarina Pažur Aničić
**Predicting the Outcomes of PBL: Comparison of Two Methods**

Blaženka Divjak, Barbi Svetec, Bart Rienties, Paul Astles, Catriona Matthews, Alexander Tillmann, Michael Eichhorn, Julia Schmitt
**Prioritizing Learning Outcomes in Different Learning Design Contexts**

Bojan Radišić, Ivan Dunđer, Sanja Seljan
**Predictive Factors for Completion or Dropping Out of Professional Studies**

Matija Novak, Marija Maček
**Student perspectives on using Artificial Intelligence tools for solving programming assignments**
**Perspektiva studenata o korištenju alata umjetne inteligencije za rješavanje programskih zadataka**

# Perception and Use of GenAI Chatbots in Education by Students at the University of Rijeka

**Martina Holenko Dlab**
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, 51000 Rijeka, Croatia
mholenko@inf.uniri.hr

**Gordan Đurović**
University of Rijeka
Trg braće Mažuranića 10, 51000 Rijeka, Croatia
gdurovic@uniri.hr

**Marko Horvat**
University of Zagreb
Faculty of Electrical Engineering and Computing
Unska 3,10000 Zagreb, Croatia
marko.horvat3@fer.unizg.hr

***Abstract***. *The development of generative artificial intelligence (GenAI) technologies offers opportunities for enhancing education, particularly through AI-driven tools like chatbots. This study explores University of Rijeka students' perceptions of GenAI chatbots (N=300), examining their motivation, usage patterns, acceptance of this technology, and the perceived value it adds to their educational experience. Our results show that a substantial majority of students have accepted and are actively using GenAI chatbots in their study routines. Despite this acceptance, our findings also highlight the need for educational institutions to provide formal education and structured integration of these tools in the teaching and learning processes. These insights inform the development of an improved version of the educational recommender system ELARS, to support personalized learning and efficient chatbot interaction in STEM education.*

**Keywords.** chatbot, LLM, ELARS recommender system, generative artificial intelligence, higher education.

## 1 Introduction

The development of generative artificial intelligence (GenAI) technologies has created numerous opportunities to fundamentally improve education which is why the application of this technology in teaching is the subject of many studies (Labadze et al., 2023), (Chiu, 2024).

As GenAI continues to pervade various industries and is expected to be an integral part of professional practice in the 21st century (Al-Zahrani, 2023), it is necessary to provide students with opportunities to acquire the skills to use it and to explore effective ways to include this technology into learning and teaching processes. Although proficiency in leveraging these tools can enhance students' career prospects and empower them to handle complex tasks and develop solutions to real-world problems, it is equally important for students to be aware of the challenges of using AI and its potential negative impacts in order to make informed decisions as future professionals. Furthermore, familiarity with ethical standards is crucial for the responsible use of AI technologies (Luo, 2024). Thus, the chance to employ GenAI within formal education through well-designed learning activities can educate students on best practices for utilizing this technology and offer them a positive experience they will (want to) repeat.

The study presented in this paper aims to explore perceptions of University of Rijeka students on the use of GenAI chatbots in learning. Chatbots are computer applications that enable human-like interaction using natural language or text. They can support students in the acquisition of complex concepts as they enable the dynamic generation of responses based on large language models (e.g., GPT-4, Gemini), are always available, and stimulate interactive conversations and student activities (Haindl & Weinberger, 2024). One of the most well-known chatbot is [1]ChatGPT by OpenAI (ChatGPT, 2024), offering advanced natural language processing capabilities (Shoufan, 2023). Other popular chatbots are [2]Copilot by Microsoft (Copilot, 2024), [3]Gemini by Google (Gemini, 2024), and [4]Claude by Anthropic (Claude, 2024).

The study was conducted within the project "Support for personalized learning in STEM based on learner personas and recommendations". The purpose of the project is to develop innovative approaches to enhance the quality of e-learning and the acquisition of knowledge and skills needed for 21st century jobs.

[1] ChatGPT: https://chatgpt.com
[2] Copilot: https://copilot.microsoft.com
[3] Gemini: https://gemini.google.com
[4] Claude AI: https://claude.ai

Results presented in this paper contribute to the body of research on the implementation of the GenAI technologies in education by identifying students' perceptions towards GenAI chatbots and factors influencing their motivation to use this technology. These findings will inform the development of the improved version of the educational recommender system ELARS (Holenko Dlab et al., 2019) that will encourage students to interact with chatbots using recommendations based on learner personas (Tudor, Holenko Dlab & Đurović, 2024). In addition, these findings could be valuable for educators aiming to use AI-based educational tools in their practice.

# 2 Related Work

Among the various applications of GenAI in education, chatbots have proven to be a particularly promising tool. Recent research underscores the transformative potential of Generative AI (GenAI) chatbots in education, revealing their wide-ranging applications and benefits. This trend indicates an increasing interest in utilizing GenAI technologies to innovate teaching methodologies (Al-Zahrani, 2023).

Labadze et al. (2023) identified key benefits of chatbots for students in their systematic literature review, including homework and study assistance, personalized learning experiences, and skill development, particularly in problem-solving and critical thinking. Chatbots offer customized content, aiding students in understanding complex concepts and enhancing flexible, personalized learning. As learning companions, chatbots provide explanations and clarifications, helping students to complete assignments more effectively.

Haindl & Weinberger (2024) provide empirical evidence that chatbots can enhance learning in undergraduate programming courses by offering personalized assistance and clarifying complex concepts. Šarčević et al. (2024) similarly found that interacting with large language models (LLMs) improves academic support, learning efficiency, and personalized learning. Xu et al. (2023) highlighted ChatGPT's potential to enhance personalization within personalized learning environments (PLEs), promoting a learner-centered approach compatible with formal education requirements.

Beurer-Kellner et al. (2023) address the technical aspects of interacting with LLMs through prompt engineering, essential for optimizing chatbot responses. Their work highlights the importance of developing effective prompts to maximize the benefits of chatbots in an educational context.

Chiu (2024) argues for involving students in research to understand their perspectives on GenAI's educational impact, which generally show positive attitudes toward chatbots. Such research shows that students find chatbots interesting, useful, and easy to use, though they note the importance of having good background knowledge due to occasional inaccuracies in chatbot responses (Shoufan, 2023). Stöhr et al. (2024) emphasize the need to consider diverse student backgrounds when integrating GenAI technologies in education. Their study found significant differences in attitudes based on gender, with female students generally expressing more negative views, and by field of study, with technology students using these tools more frequently than those in humanities.

Esiyok et al. (2024) highlight the potential of chatbots tools to enhance learning and productivity, suggesting that universities should update curricula and promote self-directed learning with technology. Al-Abdullatif (2023) also recommends encouraging chatbot use in education and calls for further research on their long-term effects on student outcomes and engagement. Chan & Hu (2023) underscore the interactive benefits of chatbots but point out significant challenges related to ensuring accuracy, relevance, and the ethical use of these tools in academic settings.

# 3 Methodology

The aim of the research is to investigate how students from University of Rijeka accept and use GenAI, particularly chatbots, how they perceive its value, and what are the key motivational factors behind the use of this technology from their point of view.

Research questions were:

- RQ1: What motivates university students to use GenAI chatbots in their studies?
- RQ2: Do university students accept and use GenAI chatbots technology?
- RQ3: How do university students perceive the value of GenAI chatbots in their studies?
- RQ4: What are the potential benefits and disadvantages associated with using GenAI chatbots in teaching and learning, as perceived by university students?
- RQ5: Do university students think that GenAI chatbots should be included into organized educational activities and why?

## 3.1 Research Instrument

To address the research questions, this study employed a paper-based questionnaire as the primary research instrument. Utilizing the paper format ensured the complete anonymity of the participating students. In the questionnaire, students had to indicate their study program level (undergraduate or graduate, year of study), age, gender and whether they had used GenAI chatbots before the survey.

If the students had used GenAI chatbots, they were asked to indicate their attitude using 18 statements (see Table 1) on a Likert scale (1 – strongly disagree, 2 – disagree, 3 – no opinion, 4 – agree, and 5 – strongly agree). In addition, they were asked to answer two

open-ended questions about the advantages and disadvantages of using GenAI chatbots as part of their studies and whether GenAI chatbots should be included in organized educational activities.

In order to examine how the students perceive GenAI chatbots, three standard frameworks were used: the Motivational Model (MM) (Aytekin et al., 2022) the Technology Acceptance Model (TAM) and the Value-Based Adoption Model (VAM) (Al-Abdullatif, 2023). These models are used to understand user acceptance and adoption of technology by analyzing different influencing factors.

The motivation model (MM) stems from broader motivation theory and enables the examination of intrinsic and extrinsic motivational factors. Intrinsic motivation (IM) can be seen as the extent to which an individual engages in an activity for its own sake, because it is inherently interesting or enjoyable. Extrinsic motivation (EM), on the other hand, can be seen as the extent to which an individual engages in an activity because they believe it contributes to achieving a valuable outcome that is different from the activity itself (Marques et al., 2011). In this study, extrinsic motivation was examined through external regulation (EM-ER) and introjected regulation (EM-IR).

The Technology Acceptance Model (TAM) is a theoretical model that aims to explain how students come to accept and use a technology, in this research GenAI chatbots, in educational activities (Al-Abdullatif, 2023). In this study, the following components of TAM were used: perceived usefulness (PU), perceived ease of use (PEOU), behavioral intention to use (BI), attitude toward using (ATU), and actual use (AU). The extent to which a student believes that using GenAI chatbots would enhance their performance was explored through PU. PEOU was used to explain the extent to which a student believes that using GenAI chatbots would be free from effort, while BI was used to explain the extent to which a student has formulated conscious plans to use or not use GenAI chatbots. Using ATU, students' general affective response to using GenAI chatbots were examined and using AU whether students have actually used GenAI chatbots.

The Value-Based Model (VAM) incorporates the concept of perceived value in understanding of GenAI adoption by students. In this study, it was used to examine perceived value (PV), perceived sacrifices (PS), and perceived benefits (PB) of GenAI chatbots. PV was examined using utility value (PV-UV) and hedonic value (PV-HV). Perceived sacrifice was explored through effort cost (PS-EC), psychological cost (PS-PC), social cost (PS-SC), time cost (PS-TC), and perceived value over cost (PS-PVOC) from the students' perspective. Using PB, the positive outcomes associated with the adoption of GenAI chatbots (e.g., convenience, improved performance) were explored.

## 3.2 Participants

Students enrolled in study programs in the technical and social science fields at the University of Rijeka participated in the research. Participants were selected based on availability, using a convenience sampling approach. Overall, 350 students participated in an anonymous survey, of which 300 had used generative AI chatbots prior to the survey.

Among the surveyed students with prior experience with GenAI chatbots, there were 230 male and 70 female students. The students' ages ranged from 18 to 27 (average age was 20.48).

## 3.3 Data Analysis

After the students had been surveyed, the questionnaires were transferred to digital form for analysis. The average (Avg) and standard deviation (StD) were calculated for each statement, both overall and by gender. Percentages for each response option (1 to 5) were calculated as well. The D'Agostino-Pearson test, using a significance level of $p<0.05$, was employed to assess whether the results for female and male students followed a normal distribution. Based on the outcomes of this normality test, the Mann-Whitney U test was chosen for comparing scores between genders for each statement. Additionally, Pearson correlation coefficients were calculated to examine the strength and direction of statistically significant linear relationships.

Answers to the open-ended questions were analyzed to identify the most common answers.

# 4 Results

Out of 350 students that have been surveyed, 300 of them have used GenAI chatbots prior to the survey, resulting in a TAM-AU score of 85.71%. For these students the results are presented in Table 1. Comparison using the Mann-Whitney U test with a significance level of $p<0.05$ revealed no significant statistical differences between the results for all statements, except for two statements from the VAM, PB (S11), $p=0.03$, and PS-EC (S12), $p=0.02$.

The results for statements related to MM are presented in Table 1 and Fig. 1. The analysis shows that GenAI chatbots are not currently a mandatory part of study programs, with an average score of 2.23 (EM-ER). Further analysis shows an average score of 2.17 for introjected regulation (EM-IR), suggesting that students are not significantly motivated by internalized pressures or expectations to use GenAI chatbots. Strong intrinsic motivation (IM) towards GenAI chatbots was found, with an average score of 4.09 indicating independent use of chatbots for study purposes. Additionally, the chatbots were perceived as inherently interesting, which is reflected in an average score of 3.40. However, it is notable that 35% of

students were neutral regarding their interest in the chatbots. Answer to RQ1 based on these results is: students are motivated to use GenAI chatbots in their studies primarily through intrinsic motivation.

The results presented in Table 1 and Fig. 2 reveal key insights into students' experiences with GenAI chatbots related to the TAM. The actual use of GenAI chatbots among students is notably high, with 85.71% suggesting that they actively use these tools. Student attitudes toward GenAI chatbots are generally positive, with an average score of 3.93. However, it is important to note that 25.33% of students were neutral, and 5.66% expressed a negative attitude, indicating that approximately 31% of students do not have a fully positive opinion on these tools. Regarding perceived usefulness, students rated GenAI chatbots with an average score of 3.67. Nonetheless, 33.66% of students did not find them particularly useful, pointing to inconsistencies in how effectively chatbots support learning activities. Additionally, there were concerns about the accuracy of chatbot responses, with an average score of 2.13 reflecting that some students experienced issues with incorrect answers, though 34.33% did not report this problem. In terms of perceived ease of use, average score of 4.25 indicates that most students find them easy to learn but only 52.33% of students felt that instructing the chatbots to provide desired answers was straightforward, with 33.33% choosing to remain neutral. Finally, average score of 3.97 reflects a strong likelihood that they will continue to use these tools throughout their studies, although 31.66% of students are uncertain about their future use of chatbots. Based on the presented results, it can be concluded that the answer to RQ2 is positive, but there is room for additional improvement. The majority of students have accepted and are using chatbots in their studies, yet they would benefit from formal education activities that will teach them how to use this technology correctly and more efficiently.

The results presented in Table 1 and Fig. 3 address RQ3 through statements of VAM. The majority of students believe that GenAI chatbots assist them in achieving their study goals, with an average score of 3.59 (PB). However, 42.33% of students do not support this belief, indicating that a significant portion of students remains unconvinced about the benefits of these tools. The results of comparison using Mann-Whitney u test indicated that male students agree more strongly with this statement compared to female students ($U=6738.50$, $p=0.03$).

Perceived costs were assessed through different types of sacrifices reported by students. The perceived effort cost (PS-EC) was rated low, with an average score of 1.83. Female students perceive using GenAI chatbots as requiring less mental effort than male students ($U=6621.00$, $p=0.02$). A statistically significant correlation with gender was found only for this statement. The correlation coefficient $r=0.1264$ ($p=0.03$) indicate a very weak positive correlation that suggest that gender has only a minor effect on how much mental effort is perceived when using chatbots.

The psychological cost (PS-PC) was rated higher at 2.77. A notable 29.33% of students expressed frustration when encountering issues with GenAI chatbots, and 26.33% were neutral on this issue. 7.66% of students have stopped using the chatbots, and 13.66% were neutral, totaling 21.33% of students who either discontinued use or are undecided. Students reported minimal concerns about perceived social cost (PS-SC), with an average score of 1.58, and generally do not view chatbots as time-consuming (PS-TC). Nevertheless, 43.33% of students did not fully support this view, with 29.33% neutral and 14% acknowledging some perceived time costs. When evaluating the perceived value (PV) of chatbots, opinions were mixed. Only 37.67% of students agreed that the chatbots make learning fun, while 27.33% disagreed and 35% were neutral. The average score for perceived value in terms of quality improvement (PV-UV) was 3.76, with 60.67% of students believing that chatbots enhance the quality of their work. 28.33% of students were neutral, and 11% disagreed. Presented results give answer to RQ3: students see value and benefits in using chatbots for learning, while they do not find the perceived sacrifices alarming.

Students have answered two open-ended questions to address RQ4 and RQ5. The first question was: *In your opinion, what are the main advantages and disadvantages of using GenAI chatbots during studying?* List of five advantages identified by students includes the following:

1. *Quick Access to Information*: Students highlight using chatbots for quick access to information they need. This allows them to save time in research and preparation for assignments and exams.
2. *Simplicity and Efficiency*: The tools are user-friendly, enabling students to easily get answers to their queries, often with explanations and context.
3. *Aid in Understanding Material*: Chatbots are useful for explaining concepts and terms that students may not fully grasp. Many students appreciate the ability to ask additional questions and receive clear explanations.
4. *Reduced Mental Effort*: These tools simplify information retrieval, allowing students focus on more important aspects of learning.
5. simplify information retrieval, enabling students to focus on key learning aspects
6. *Support in Problem Solving*: Chatbots assist students in solving math and programming tasks, often providing step-by-step instructions that facilitate understanding and problem-solving.

List of five disadvantages identified by students includes the following:

1. *Inaccuracy of Information*: One of the most frequently mentioned issues is the inaccuracies in data generated by GenAI chatbots. Students

emphasize the need for additional verification of all information obtained through these tools.
2. *Dependency on Technology*: There is concern that excessive use of chatbots may reduce students' critical thinking abilities and their capacity for independent problem-solving. Relying too much on technology could lead to lack of effort and reduced deeper understanding of the material.
3. *Confusing or Incomplete Information*: In some cases, chatbots may provide confusing/incomplete information, which can hinder students' ability to properly comprehend or solve a problem.
4. *Potential for Academic Misconduct*: The use of chatbots may encourage some students to engage in plagiarism or cheating rather than genuine learning and understanding of the material.
5. *Language and Contextual Barriers*: The tools are often better developed for English language use, which may pose challenges for students seeking information in Croatian or having specific requirements in their local context.

The answer to RQ4 is that in students' opinion, the use of chatbots offers significant advantages in terms of quicker and easier access to information, as well as assistance in learning and task-solving, but there is also a need for a critical approach and additional verification of information to avoid inaccuracies and potential pitfalls of excessive reliance on technology.

**Table 1.** Results of questionnaire (N = 300)

| Model Code | | Statement | Female (N = 70) | | Male (N = 230) | | All (N = 300) | |
|---|---|---|---|---|---|---|---|---|
| | | | Avg | StD | Avg | StD | Avg | StD |
| Motivational model (MM) | | | | | | | | |
| EM-ER | S1 | The use of GenAI chatbots is mandatory in my study program. | 2.30 | 1.15 | 2.21 | 1.22 | 2.23 | 1.21 |
| EM-IR | S2 | I use GenAI chatbots to avoid feeling like I'm falling behind. | 2.01 | 1.16 | 2.21 | 1.22 | 2.17 | 1.21 |
| IM | S3 | I independently used GenAI chatbots for my study needs. | 4.00 | 1.13 | 4.11 | 1.17 | 4.09 | 1.17 |
| IM | S4 | I enjoy using GenAI chatbots because they are interesting. | 3.19 | 1.03 | 3.46 | 1.15 | 3.40 | 1.13 |
| Technology Acceptance Model (TAM) | | | | | | | | |
| PU | S5 | The answers provided by GenAI chatbot are accurate and do not need to be verified. | 2.16 | 1.01 | 2.13 | 0.99 | 2.13 | 0.99 |
| PU | S6 | GenAI chatbot help me execute my study tasks more efficiently. | 3.71 | 1.06 | 3.65 | 1.21 | 3.67 | 1.18 |
| PEOU | S7 | Learning to use GenAI chatbot is easy. | 4.30 | 0.80 | 4.23 | 0.86 | 4.25 | 0.85 |
| PEOU | S8 | It's easy for me to instruct GenAI chatbot to do what I want. | 3.43 | 0.93 | 3.53 | 0.94 | 3.50 | 0.94 |
| BI | S9 | I will definitely use GenAI chatbot throughout my studies. | 3.91 | 1.04 | 3.98 | 1.16 | 3.97 | 1.14 |
| ATU | S10 | My attitude towards using GenAI chatbot is positive. | 3.84 | 0.84 | 3.96 | 0.96 | 3.93 | 0.94 |
| Value-Based Model (VAM) | | | | | | | | |
| PB | S11 | I use GenAI chatbot because I believe they help me achieve my goals. | 3.37 | 1.04 | 3.66 | 1.19 | 3.59 | 1.16 |
| PS-EC | S12 | GenAI chatbot require a lot of mental effort. | 1.63 | 0.85 | 1.90 | 0.90 | 1.83 | 0.89 |
| PS-PC | S13 | I feel frustrated when I encounter a problem with a GenAI chatbot. | 2.91 | 1.32 | 2.72 | 1.23 | 2.77 | 1.25 |
| PS-PVOC | S14 | I used GenAI chatbot but I no longer use them. | 1.81 | 0.99 | 1.81 | 1.04 | 1.81 | 1.03 |
| PS-SC | S15 | GenAI chatbot isolate me from my colleagues at university. | 1.44 | 0.86 | 1.62 | 0.93 | 1.58 | 0.92 |
| PS-TC | S16 | It takes a lot of time to effectively use GenAI chatbot. | 2.26 | 1.13 | 2.35 | 1.04 | 2.33 | 1.06 |
| PV-HV | S17 | GenAI chatbots make learning fun. | 3.26 | 1.05 | 3.05 | 1.17 | 3.10 | 1.15 |
| PV-UV | S18 | GenAI chatbots improve the quality of my work. | 3.61 | 0.96 | 3.80 | 1.10 | 3.76 | 1.07 |

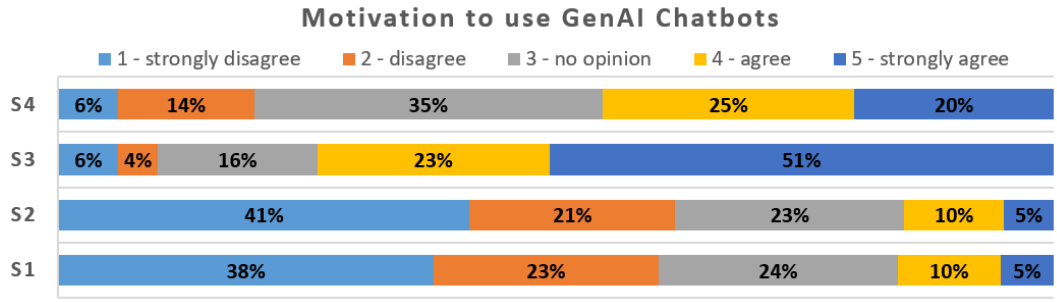



**Figure 1.** Motivational model (MM) results

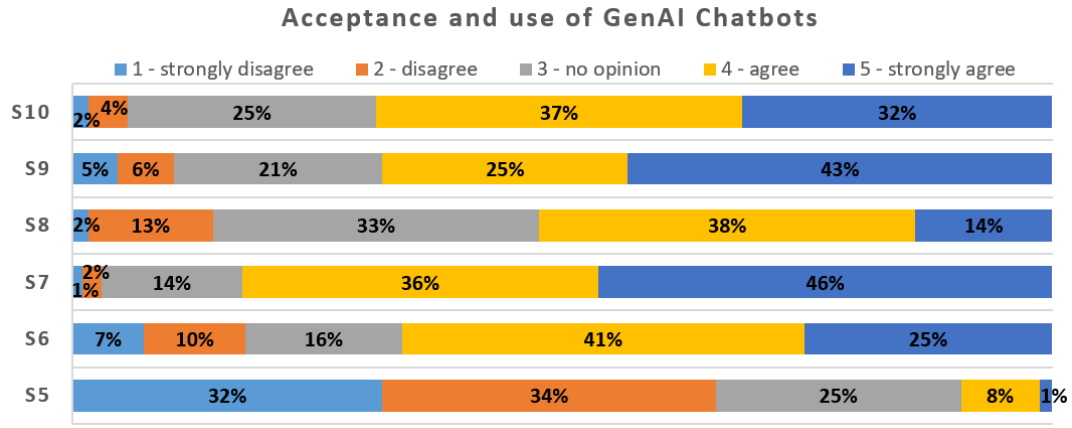



**Figure 2.** Technology Acceptance Model (TAM) results

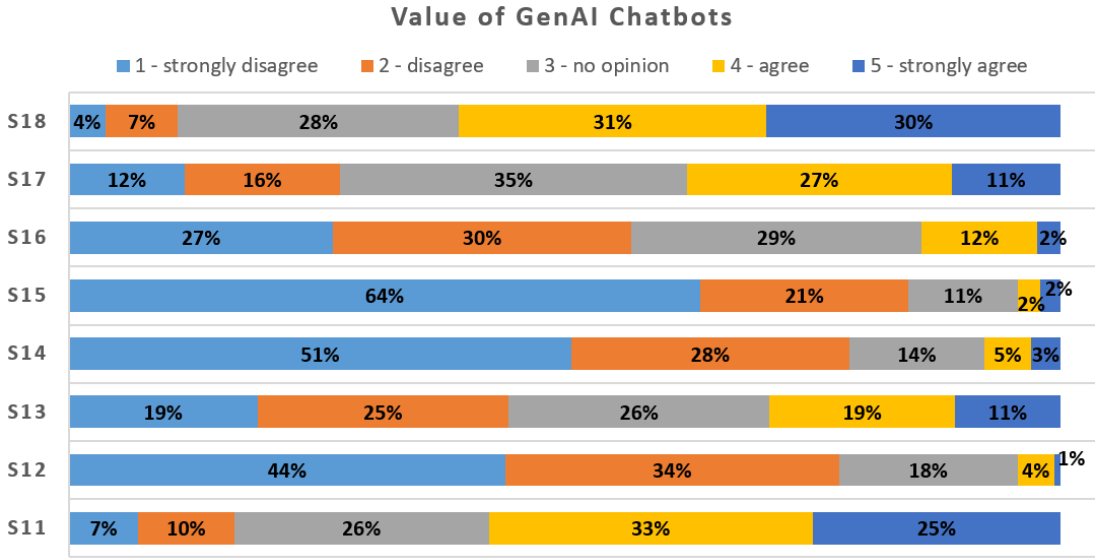



**Figure 3.** Value-Based Adoption Model (VAM) results

The second open-ended question was: *In your opinion, should the requirement to use GenAI chatbots during studying be increased (by incorporating them into organized educational activities), and why?* Five most common answers were:

1. *Support for Increased Use*: Students believe that the use of chatbots enhance their learning. They argue that quick access to information and explanations make learning more efficient.
2. *Support for Optional Use*: Students feel that while GenAI chatbots are beneficial, their use should remain optional rather than mandatory. They suggest that students should be free to choose whether or not to use these tools.
3. *Concern Over Dependency and Creativity*: Students have expressed concern that mandatory use of GenAI chatbots could reduce students'

creativity and critical thinking skills, making them overly reliant on technology.
4. *AI as a Future Necessity*: Students have highlighted that AI represents the future, and students should be familiarized with these tools to be prepared for their professional lives.
5. *Educational Value and Proper Use*: Students also emphasized the importance of learning to use GenAI chatbots correctly to maximize their benefits and avoid potential pitfalls.

The answer to RQ5 based on these results is that students would like for GenAI chatbots to be included in organized educational activities but that they should also be properly educated in a correct way of using them in order to achieve the desired effect in teaching and learning activities.

# 5 Discussion

## 5.1 Motivation to Use GenAI Chatbots in Studies

To address RQ1, the results indicate that GenAI chatbots are not currently a mandatory part of study programs. This suggests that, at present, the integration of GenAI chatbots into educational settings is largely driven by students' personal interest, as reported in (Lee et al., 2022), rather than by formal requirements or to avoid negative consequences. However, the observation that 35% of students selected "no opinion" regarding the interest in GenAI chatbots indicates there is considerable uncertainty or indifference towards the value of these tools. This highlights the need to enhance the perceived relevance and attractiveness of GenAI chatbots within the educational context. Additional efforts should be made to motivate students to incorporate GenAI chatbots into their self-directed learning activities, as noted by Esiyok et al. (2024).

Demonstrating the practical benefits of using chatbots and providing personalized recommendations based on students' individual needs could enhance the motivational aspects associated with these tools. As there are no significant differences between male and female students, the strategies can be developed without gender-specific adaptations.

## 5.2 Acceptance and Use of GenAI Chatbots

In relation to RQ2, the results indicate that the use of GenAI chatbots is widespread among students, suggesting a general acceptance and integration of this technology into their study routine but also the growing role of GenAI chatbots in education. However, the results also highlight areas where further support and training could be beneficial. A significant proportion of students neither agreed nor disagreed with the perceived usefulness and ease of use of these tools. This is in line with the findings of Stöhr et al. (2024) and Haindl & Weinberger (2024), who suggest that the usefulness and ease of use of these tools are still perceived differently. These findings suggest that more structured guidance is needed to ensure effective use.

While many students find the chatbots easy to learn, the results also suggest that more personalized support is needed to help students interact with this technology. Results indicate that strategies to enhance interaction are likely to impact male and female students similarly, but this needs to be verified in future studies since research by Stöhr et al. (2024) reported differences in attitudes across genders.

The observation that some students receive incorrect responses suggests that they need to be better trained in writing prompts. This emphasizes the importance of teaching students not only how to use chatbots, but also how to critically evaluate the answers provided. This is in line with related research done by Al-Abdullatif (2023) and Chiu (2024) who also emphasize the need for structured educational initiatives to improve students' ability to effectively use chatbots.

## 5.3 Value of GenAI Chatbots for Studies

Results related to RQ3 reveal that although a significant number of students believe that GenAI chatbots assist in achieving their study goals, a notable proportion remain unconvinced about their overall benefits. In addition, there is a substantial number of students who do not find them as effective or valuable. Students are divided on whether these tools make learning enjoyable but the overall impression is that chatbots can enhance the quality of their work. According to Al-Abdullatif (2023), by making GenAI chatbots more compelling and relevant to students' educational experiences, educators can potentially increase their perceived value.

The analysis of perceived costs reveals that the effort required to use chatbots is seen as minimal, but the psychological costs are higher, as some students express frustration when encountering problems. This indicates the need for enhanced support and training to help students navigate these tools more effectively and reduce potential problems, which will also help to increase the perceived value (Al-Abdullatif, 2023). Integration into formal education is certainly a good opportunity to achieve that. Perceptions of chatbots in terms of time spent and social impact vary. Although many students do not perceive chatbots as particularly time-consuming or socially burdensome, some students do, suggesting that the impact of chatbots on students' time management and social interactions may be shaped by personal experiences and specific contexts. Providing personalized recommendations and feedback may contribute to overall satisfaction, but future research is needed to understand what makes chatbots attractive to students in order to maintain their interest in using them (Shoufan, 2023).

### 5.4 Advantages and Disadvantages of GenAI Chatbots

When analyzing the results from open-ended questions, it can be concluded that students are aware of the advantages and disadvantages of GenAI chatbots. From their answers five advantages were identified: *Quick Access to Information*, *Simplicity and Efficiency*, *Aid in Understanding Material*, *Reduced Mental Effort* and *Support in Problem Solving*. On the other hand, five disadvantages were identified: *Inaccuracy of Information*, *Dependency on Technology*, *Confusing or Incomplete Information*, *Potential for Academic Misconduct* and *Language and Contextual Barriers*. Identified advantages and disadvantages are in line with those reported by Lee et al., (2022) and Chan & Hu (2023). Although listed advantages are indicated in the results given to statements with Likert scale, some of the disadvantages were not. Students independently indicated that there was a possibility of academic misconduct when using GenAI chatbots and that some of them encountered language and contextual barriers.

These findings also suggest that students should be formally trained in the proper use of chatbots for learning in order to engage them in their learning activities in a way that helps them in desired ways while minimizing the stated drawbacks. Tailoring interactions with chatbots to students' needs and learning styles could significantly reduce the perceived disadvantages and encourage a more positive reception of these tools.

### 5.5 Introduction of GenAI chatbots in Educational Activities

The results addressing RQ5 indicate that while students support the integration of GenAI chatbots into structured educational activities, there is a clear need for comprehensive training on their effective use to maximize their impact on teaching and learning outcomes. Students' opinions vary on whether introduction of chatbots should be mandatory or optional. Students have expressed their concern over dependency and reduced creativity if GenAI chatbots are used but are also aware that they are a future necessity and cannot be avoided. A particularly important opinion stated by students is that they see value in educational use of chatbots if used in a proper way, and they do expect to be taught how to use them in a formal and organized way. This result is consistent with results indicating that at least a third of the students would greatly benefit from formal education regarding the proper use of chatbots.

Several strategies and resources can be utilized to effectively implement in practice the recommendation of appropriate prompts in educational activities. First, recommendations of interactive tutorials and online courses focused on effective communication with GenAI chatbots can provide students with flexible and self-paced learning opportunities which, in addition to developing skills, will also increase their motivation to use these tools. These resources may include video demonstrations, interactive simulations, and quizzes designed to reinforce key concepts and techniques. Second, appropriate prompt libraries could be created, containing a repository of example prompts and responses that can be used as a valuable resource by both educators and students. The quality of the language model response is related to the structure and precision of the query (Beurer-Kellner et al., 2023), so recommending appropriate prompts could help students to communicate effectively with conversational agents. Those recommendations could be in accordance with their characteristics and preferences (learner persona). For this purpose educators can develop a set of guidelines or frameworks that outline the characteristics of effective prompts and that are used to generate recommendations for students. These frameworks should include templates for well-structured questions, the use of specific keywords, and strategies for refining queries to obtain more precise and relevant responses from the chatbots. Finally, within educational recommender system, students could receive feedback on their prompts with recommendations on how to improve them.

## 6 Conclusions and Future Work

The study presented in this paper was conducted within the project "Support for Personalized Learning in STEM Based on Learner Personas and Recommendations" with the aim to explore University of Rijeka students' perceptions of using GenAI chatbots in learning. The results indicate that most students use GenAI chatbots for their studies. While external motivational factors are not significant, many students show intrinsic motivation. There is potential for improvement through external motivation, such as integrating tools into study programs. Regarding perceived value, students see value in using GenAI chatbots, though many remain uncertain.

The results suggest that formal education on using GenAI chatbots would benefit students. While students acknowledge both the advantages, like quick information access, and disadvantages, such as inaccuracy, they generally do not find them alarming. Most students support integrating GenAI chatbots into learning, given proper training. Despite intrinsic motivation among many, some students are indifferent or face challenges, highlighting the need for structured integration and education on these tools.

The results from this paper and future work will enhance personalized STEM education, preparing students for 21st-century careers. By generating recommendations for interaction with GenAI chatbots based on student characteristics and educational personas, personalized learning support for STEM

education will be implemented in the new version of the educational recommender system ELARS.

In terms of limitations, this study's convenience sampling limits generalizability to all University of Rijeka students. Future research should include a more diverse student population from various faculties and universities in Croatia. Additionally, the study assumes a uniform level of technological familiarity, which may vary. Future studies should analyze results by subgrouping students based on this criteria and field of study.

## Acknowledgments

The research has been co-funded by the University of Rijeka under project "Support for personalized learning in STEM based on learner personas and recommendations" (uniri-iskusni-drustv-23-92944).

## References

Al-Abdullatif, A. M. (2023). Modeling Students' perceptions of chatbots in learning: Integrating technology acceptance with the value-based adoption model. Education Sciences, 13(11), 1151.

Al-Zahrani, A. M. (2023). The impact of generative AI tools on researchers and research: Implications for academia in higher education. Innovations in Education and Teaching International, 1-15.

Aytekin, A., Özköse, H. & Ayaz, A., (2022). Unified theory of acceptance and use of technology (UTAUT) in mobile learning adoption: Systematic literature review and bibliometric analysis. COLLNET Journal of Scientometrics and Information Management, 16(1), 75.

Beurer-Kellner, L., Fischer, M., & Vechev, M. (2023). Prompting is programming: A query language for large language models. Proceedings of the ACM on Programming Languages, 7(PLDI), 1946-1969.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20(1), 43.

Chiu, T. K. (2024). Future research recommendations for transforming higher education with generative AI. Computers and Education: Artificial Intelligence, 6, 100197.

Esiyok, E., Gokcearslan, S., & Kucukergin, K. G. (2024). Acceptance of Educational Use of AI Chatbots in the Context of Self-Directed Learning with Technology and ICT Self-Efficacy of Undergraduate Students. International Journal of Human–Computer Interaction, 1-10.

Haindl, P., & Weinberger, G. (2024). Students' Experiences of Using ChatGPT in an Undergraduate Programming Course. IEEE Access, 12, 43519-43529.

Holenko Dlab, M., Đurović, G., Hoić-Božić, N., & Botički, I. (2019). Support for knowledge assessment in STEM education using ELARS recommender system. In 10th International Conference on e-Learning (pp. 55-59).

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education, 20(1), 56.

Lee, Y. F., Hwang, G. J., & Chen, P. Y. (2022). Impacts of an AI-based cha bot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. Educational technology research and development, 70(5), 1843-1865.

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. Assessment & Evaluation in Higher Education, 1-14.

Marques, B.P., Villate, J.E. & Carvalho, C.V. (2011). Applying the UTAUT model in engineering higher education: Teacher's technology adoption. In 2011-*6th* Iberian Conference on Information Systems and Technologies (pp. 1-6).

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. IEEE Access, 11, 38805-38818.

Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. Computers and Education: Artificial Intelligence, 100259.

Šarčević, A., Tomičić, I., Merlin, A., & Horvat, M. (2024). Enhancing Programming Education with Open-Source Generative AI Chatbots. In MIPRO 2024-47 th ICT and Electronics Convention (pp. 2367-2372).

Tudor, I., Dlab, M. H., & Đurović, G. (2024, May). Learner Personas in Technology-Enhanced Learning. In 2024 47th MIPRO ICT and Electronics Convention (MIPRO) (pp. 311-316). IEEE.

Xu, X., Wang, X., Zhang, Y., Zhang, H., & Wu, Y. (2023). Applying ChatGPT to Tackle the Side Effects of Personal Learning Environments in Higher Education: A Teacher and Teaching Perspective. In Machine Learning and Artificial Intelligence (pp. 73-88). IOS Press.

Yetişensoy, O., & Karaduman, H. (2024). The effect of AI-powered chatbots in social studies education. Education and Information Technologies, 1-35.

# Tablet Math Application for Supporting Independent Simultaneous Interaction of Colocated Users

**Patrik Komljenović, Mia Čarapina**
Zagreb University of Applied Sciences
Department of Information Technology and Computing
Vrbik 8, 10020 Zagreb, Croatia
{pkomljeno, mcarapina}@tvz.hr

***Abstract***. *This paper presents an Android application for tablets that allows multiple students to solve mathematical assignments simultaneously while using a single device. To facilitate this, the tablet's screen can be split into two, three, or four sections, enabling individual interaction. Still, a single student can use the device without applying the split screen feature. Following the process from the idea to the final realization, this paper provides insight into the technology and techniques used to implement the split screen feature. In addition, the application was evaluated with first-grade primary school pupils in Croatia. Pupils interacted with the application under two different conditions: individually and in pairs or a group of three. The results indicate that even though pupils liked working in both settings, a higher number preferred working alone rather than sharing a device in pairs.*

**Keywords.** Mobile learning, primary education, tablet sharing, Android application development, split screen

## 1 Introduction

Education and the approaches to teaching students constantly evolve. The introduction of mobile technology in schools (Quinn, 2001) offered students a new tool to engage with educational content dynamically and seamlessly (Wong & Looi, 2011). Throughout the last two decades, various initiatives promoted the 1:1 (one device-per-one student) learning (Balanskat et al., 2013; Richardson et al., 2013). However, many schools worldwide still do not have enough resources to meet this device distribution (European Commission, 2019; Heinrich et al., 2020; Tamim et al., 2015). Under these circumstances, educational activities can be organized in a way that students share a device. Yet educational applications are rarely designed and modeled to support the interaction of multiple students sharing a device. Examples of supporting interactions among multiple users on a single device can be observed in the single display groupware (SDG) domain (Stewart et al., 1999), which focuses on developing systems that allow people physically located next to each other to interact simultaneously with one common output interface. However, the SDG model has been used mostly in activities on personal computers and shared large displays (Kumar, 2008; Nussbaum et al., 2008).

This paper presents an educational application (Komljenović, 2023) developed to support multiple colocated users' simultaneous use of a single tablet by dividing the screen into multiple independent segments. The application is designed to help users practice and learn basic mathematical operations (addition, subtraction, multiplication, and division).

A similar application was previously developed for iPads and evaluated with pupils attending the first four grades of primary education in Croatia (Čarapina & Pap, 2023). In their study, the authors explored differences in students' performance and perceived satisfaction completing similar mathematical tasks on paper, tablet in 1:1 (one-per-one), and tablet in 1:m (one-per-many) settings. The results showed a statistically significant difference in task performance between the paper and tablet conditions in favor of paper treatment. However, the study suggested that this might be attributed to the students' unfamiliarity with the application and their limited experience with using tablets in the classroom. Still, no significant differences were found between the 1:1 and the 1:m conditions. In addition, no difference was found between the three evaluated 1:m settings (i.e., 1:2, 1:3, and 1:4). Moreover, the students' perceived satisfaction was significantly higher while working with tablets than during the paper condition. Still, the results showed no significant difference between the 1:1 and the 1:m conditions. Since one of the highlighted limitations of the study in question was combining the different 1:m settings (1:2, 1:3, and 1:4) altogether, in this study, the authors aimed to explore in more detail the perceived satisfaction of students while working alone on a tablet compared to working in pairs, i.e., 1:2 settings.

## 2 Tablet Split Screen Design and Implementation

Android OS supports multiple ways to simultaneously use and display two applications (*Multi-Window Support*, 2023). One way is to use a split screen mode to view two applications side by side. However, there is no possibility to display more than two applications at the same time. Other ways include using picture-in-picture mode, in which one application is opened in a smaller element and displayed over the main application that is in focus. This mode is mainly used for playing videos. Some applications support the floating display feature, allowing them to be used in a floating window on top of other applications. These features supported in native Android development did not cater to our specific needs. Since we wanted to support more than two students using a single application simultaneously on a single tablet, we looked for other solutions, and a new application was developed using Android's fragments.

The tablet math application's development started in Figma, where a split screen prototype was designed to support up to four users on independent segments (Franković et al., 2023). Several things were considered, such as students taking control of the device or the layout orientation and rotation of graphical elements if students sit one across from the other. Each segment was visually distinguished, with different backgrounds, buttons, and font colors.

The application was developed natively using the Kotlin programming language and Android Studio IDE to support learning and practicing basic math operations. Besides using it in split screen mode, a single user can also interact with the application (Fig. 1). The split screen feature was implemented by using fragments, which allowed the division of the screen into several segments that function independently, as shown in Fig. 2, Fig. 3., and Fig. 4. A fragment represents a part of the application interface that can be used multiple times. It has its appearance and lifetime and can process the events that occur due to input. They cannot exist alone but must be triggered by another activity or fragment.

In addition, in the developed application, each fragment can be rotated and adjusted in orientation before and during the activity. Fragments retain their orientation through activity changes by saving the rotation button's unique identifier. Each time the rotate button is clicked, the fragment undergoes a 90-degree rotation, altering the orientation of the content container accordingly. This method ensures that the content inside the container remains intact and properly aligned with each 90-degree rotation. By preserving the layout and integrity of the content, users can seamlessly rotate the fragment without encountering any disruptions.

**Figure 1.** Single user mode

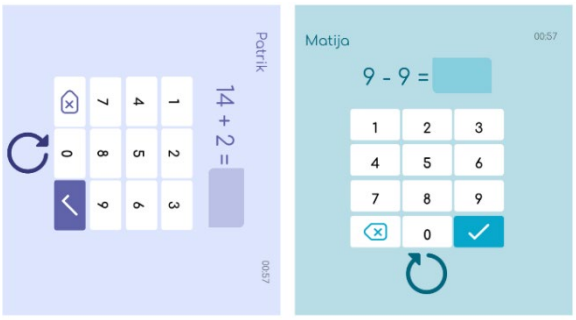



**Figure 2.** Two fragments of the math activity

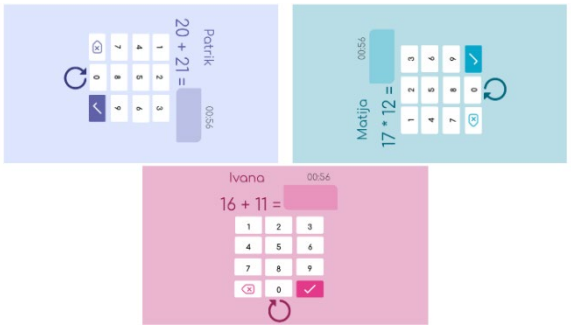



**Figure 3.** Three fragments of the math activity

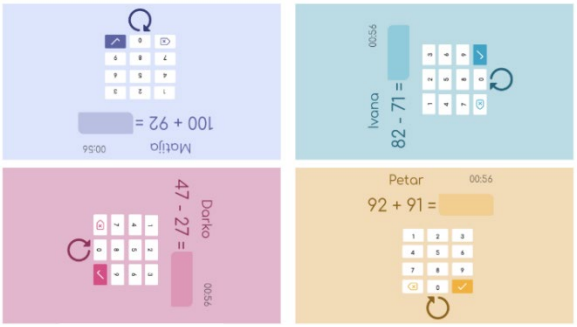



**Figure 4.** Four fragments of the math activity

## 3 Application's Workflow

When the application starts, the students see a welcome screen (Fig. 5) that allows them to start the math activity or select a button that opens the settings screen (Fig. 6). In addition, the currently selected settings for the math activity are displayed in the bottom right

corner of the welcome screen. In the application's settings, users can change the lower and upper limits of the numbers to be generated, modify the countdown timer, increase the number of users, and enable or disable available operators.

**Figure 5.** Welcome screen

**Figure 6.** Settings screen

The application's personalization screen depicted in Fig. 7 looks different depending on the number of players selected. It displays text fields where students enter their names, enabling the personalization of the segments.

**Figure 7.** Personalization screen for students' name inputs (this example is for splitting the screen into four segments)

After the activity starts, the students' names are displayed on each segment, and instructions for the task are given. The math activity starts after each student confirms he or she is ready to solve the tasks. The math activity displays the student's name in each segment, the countdown timer, and in the middle, the generated math task that the student has to solve by entering a number using the numpad keyboard (Fig. 1). Submitting an answer automatically generates another math task until the time runs out. After the time runs out, a student is presented with the results (Fig. 8). On the showcase, they can see the number of correct answers and scroll through them.

**Figure 8.** Results screen

# 4 Application Evaluation

The application was evaluated in school settings to determine the next steps for improving the application's design and to explore students' perceived satisfaction when sharing a tablet with a split screen feature in pairs compared to working alone.

## 4.1 Research Methodology and Participants

During a 45-minute school hour in May 2024, 15 first-grade pupils (7 to 8 years of age) from one primary public school in Zagreb, Croatia, solved a similar set of generated mathematical tasks (addition of numbers from 5 to 15) during two different 5-minute conditions. During the intervention, a class teacher was present but did not interfere with the students in the classroom. It should be noted that pupils who participated in this study are not used to tablets as part of the in-school assignments, and this was the first time they had used the developed math application. In one condition, they were using a tablet alone, and in another condition, they shared a tablet with another pupil (i.e., they worked in pairs). However, one group was formed out of three students due to the odd number of present students in the classroom.

Since the pupils in the classroom were distributed across two rows of tables, during the first 5 minutes the first row was assigned to a single-use condition and the second row was grouped in pairs (Fig. 9). In the following 5 minutes, the conditions were switched, and the second-row pupils solved the assignment using a tablet alone, and pupils in the first row shared a tablet in pairs and one group of three (Fig. 10).

**Figure 9.** Research settings during the first condition where the first row used tablets in 1:1 distribution and the second-row students worked in pairs

**Figure 10.** A group of three students sharing a tablet

After completing the assignments, they were asked to complete an anonymous questionnaire with three questions (Fig. 11). The content was read aloud to ensure that everyone clearly understood the questions that needed to be answered.

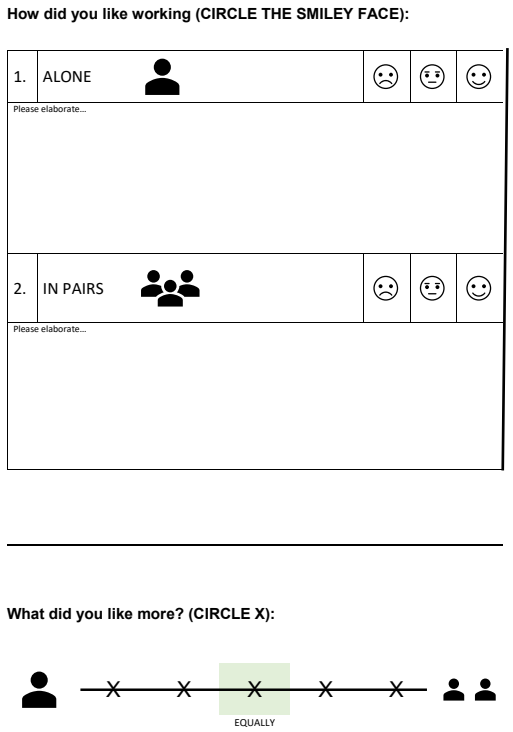



**Figure 11.** Questionnaire

The first two questions were a simple Likert-type three-point scale with smiley faces (sad, neutral, and happy) and open-ended questions to rate their experience working alone and sharing a tablet. In the third question, they were asked to rate their overall preference between working alone and in a group sharing a tablet on a Likert-type five-point scale, where 1 indicated they preferred working alone over sharing a tablet, 3 indicated they liked both settings the same, and 5 indicated they preferred sharing a tablet over working alone.

## 4.2 Results and Discussion

Seven pairs and one group of three students participated in the study. The following sections present the satisfaction questionnaire results and overall observations regarding the application's usability and performance.

### 4.2.1 Questionnaire Analysis

As each questionnaire was filled out correctly, the responses of all 15 students were included in the analysis. Table 1 presents a descriptive statistic of the results.

**Table 1.** Descriptive statistics of the questionnaire responses

| | Mean | Std. Deviation |
|---|---|---|
| Working alone | 2.53 | 0.834 |
| Sharing a tablet | 2.33 | 0.617 |
| Preference | 2.47 | 1.598 |

The answers regarding satisfaction of working alone and sharing a tablet show overall positive attitudes. Still, the responses indicate that the rate of positive responses was higher for the individual activity (Fig. 12) than for the shared tablet activity (Fig. 13). However, the rate of negative responses (i.e., sad smiley face) was also higher for the individual activity (20%) compared to sharing a tablet (6.67%).

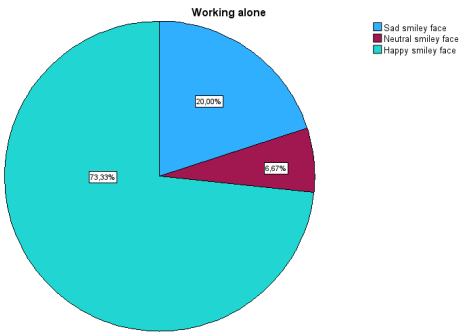



**Figure 12.** Ratings for the single-user activity

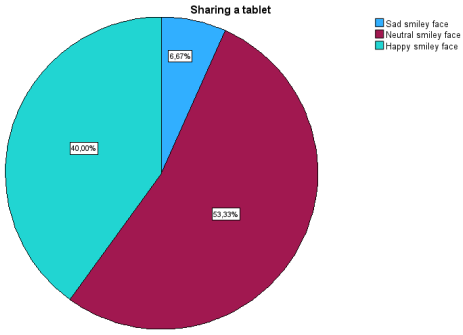



**Figure 13.** Ratings for the shared tablet activity

Since the pupils wrote no descriptive comments to elaborate on reasons for their responses, this should be explored in more detail. Additional interviews are recommended because most pupils of that age are still not fluent in reading and writing.

However, it was observed that some students were physically dominant over the shared device during the paired activity, which could have influenced the satisfaction of both parties (Fig. 14).

**Figure 14.** An example of a dominant student (on the right) in a pair

In the third question, on the preference scale, pupils responded by circling 1, indicating they preferred working alone over sharing; 3, indicating they liked both settings equally; or 5, indicating they preferred sharing over working alone. No one circled the in-between values of 2 and 4. The average result on the preference scale was 2.47, showing that a higher percentage of pupils preferred working alone (46.67%) compared to those who preferred sharing a tablet (20%). This can be observed in the rating percentage depicted in Fig. 15. Yet, 33.33% of pupils preferred both conditions equally.

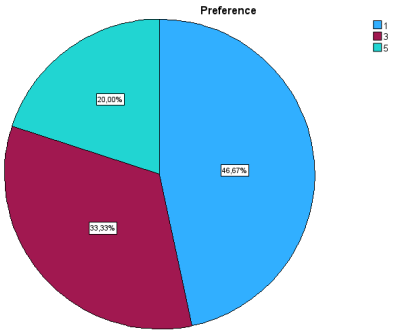



**Figure 15.** Settings preference analysis

### 4.2.2 Application Usability and Performance

Regarding the application's usability, a group of three had problems understanding what order to enter their names when prompted (Fig. 3). Enabling the application to inverse the prompt positions for three students to enter their names would help students align the application's layout to their seating position.

Moreover, since we wanted to evaluate the application in a regular classroom setting, we observed some limitations in the application's setting page regarding the feature of setting up the range of numbers to be used. To be more precise, the minimum and maximum values for each operand could be adjusted, yet the range for the result cannot be set, which turned out to be as important as the previous two.

In addition, it was observed that the timer resets after the application is closed (intentionally or unintentionally). Since there were a few cases of pupils unintentionally putting out the application, making the activity timer run even when the user leaves the application would solve the timer reset that happens if a user refreshes the activity. This would provide more adaptability for various learning stages and activities.

## 4.3 Research Limitations

The current application allows four students to solve math problems simultaneously. However, some limitations should be addressed. The evaluation was conducted with only 15 first-grade students, limiting the findings' generalizability. A larger and more diverse sample would provide a more comprehensive understanding of the application's effectiveness and user satisfaction.

Moreover, the evaluation was conducted over a 45-minute session, which may not be sufficient to fully understand the long-term impact of using the application on students' learning outcomes and preferences. Participants were not accustomed to using tablets for educational activities, which could influence their responses and overall experience with the application. This novelty effect might not reflect the typical classroom setting where students are more familiar with the technology.

Using anonymous questionnaires limits the ability to follow up on individual responses to understand the reasons behind certain preferences or dissatisfaction.

Additionally, the questionnaire did not capture detailed qualitative feedback due to the participants' young age and limited writing skills.

It should also be highlighted that the study included one group of three students due to the odd number of participants, which introduces variability that may have affected the results. The differences in dynamics between pairs and groups of three were not separately analyzed, potentially confounding the results.

# 5 Conclusion and Future Work

This paper presents an Android mobile application for generating math tasks. The application supports the shared use of a single tablet among colocated users. This is enabled with the implemented split screen feature, which divides the tablet's display into two, three, or four segments. Each segment can be adjusted in rotation and display an independent math activity. Fragments are the basis of this application, and their implementation supports splitting the screen on mobile devices into more than two parts. This enabled a parallel display of independent math activities.

The application was evaluated with first-grade pupils to gain insights into possible improvements and pupils' attitudes while interacting with it in shared mode. The questionnaire results indicate that most pupils preferred working alone over sharing a device in pairs.

Future improvements to the application include redesigning the interaction for the group of three to be more position-independent, adding more math activity adjustment options on the settings page, and improving the application's general performance. The application currently focuses on basic mathematical operations (addition, subtraction, multiplication, and division). This narrow scope might limit its applicability across different educational stages and subjects. In addition, adding features such as progress tracking or adaptive learning algorithms that analyze areas where the student struggles could improve the learning experience.

# References

Balanskat, A., Bannister, D., Hertz, B., Sigillò, E., & Vuorikari, R. (2013). *Overview and analysis of 1:1 learning initiatives in Europe*. https://doi.org/10.2791/20333

Čarapina, M., & Pap, K. (2023). Exploring colocated synchronous use of tablets based on split screen feature. *IEEE Access*, *11*, 123418–123432. https://doi.org/10.1109/ACCESS.2023.3329478

European Commission. (2019). *2nd Survey of Schools: ICT in Education*. https://data.europa.eu/euodp/data/storage/f/2019-03-19T084831/FinalreportObjective1-BenchmarkprogressinICTinschools.pdf

Franković, R., Čarapina, M., & Uglješić, V. (2023). Dizajn vizualnog identiteta edukacijske aplikacije za djecu niže osnovnoškolske dobi. In J. Žiljak Gršić (Ed.), *Tiskarstvo i dizajn 2023* (pp. 67–75). Akademija tehničkih znanosti Hrvatske (HATZ).

Heinrich, C. J., Darling-Aduana, J., & Martin, C. (2020). The potential and prerequisites of effective tablet integration in rural Kenya. *British Journal of Educational Technology*, *51*(2), 498–514. https://doi.org/https://doi.org/10.1111/bjet.12870

Komljenović, P. (2023). *Izrada Android aplikacije za generiranje matematičkih zadataka za niže razrede osnovne škole*. Tehničko veleučilište u Zagrebu.

Kumar, D. (2008). *Study of split screen in shared-access scenarios- optimizing value of PCs in resource-constrained classrooms in developing countries.* [UC San Diego]. https://escholarship.org/uc/item/7x8081jw

*Multi-window support*. (2023, November 2). Android Developers. https://developer.android.com/guide/topics/large-screens/multi-window-support

Nussbaum, M., Infante, C., Hidalgo, P., Reyes, T., Weitz, J., Gottlieb, A., Echeverria, A., Calderón, J., & Bravo, C. (2008). Collaborative 1:1 with emerging markets available ICTs. *Proceedings - ICCE 2008: 16th International Conference on Computers in Education*, 609–616.

Quinn, C. (2001). *mLearning: Mobile, Wireless, In-Your- Pocket Learning*.

Richardson, J., Mcleod, S., Flora, K., Sauers, N., Kannan, S., & Sincar, M. (2013). Large-scale 1:1 computing initiatives: An open access database. *The International Journal of Education and Development Using Information and Communication Technology*, *9*, 4–18.

Stewart, J., Bederson, B. B., & Druin, A. (1999). Single display groupware: A model for co-present collaboration. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 286–293. https://doi.org/10.1145/302979.303064

Tamim, R. M., Borokhovski, E., Pickup, D., & Bernard, R. M. (2015). *Large-Scale, Government- Supported Educational Tablet Initiatives*. https://oasis.col.org/server/api/core/bitstreams/3348c70b-eee3-47d9-ac29-83d595520b3a/content

Wong, L.-H., & Looi, C.-K. (2011). What seams do we remove in mobile-assisted seamless learning? A critical review of the literature. *Computers & Education*, *57*, 2364–2381. https://doi.org/10.1016/j.compedu.2011.06.007

# The relationship between student work and academic success: a case study

**Sonja Nidogon Višnjić**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin
svisnjic@student.foi.hr

***Abstract.*** *Working during academic studies is very important because most students gain their first work experience during their academic education. Some students work occasionally while studying, most often during the summer period when they do not have any academic obligations, while others work continuously throughout their academic education. This research aims to investigate, using a practical example, whether students who work during their academic education achieve better academic success, measured by a higher number of earned ECTS credits and regular progression from year to year.*
*The research was conducted based on data collected from the employment agency database for the student population in the Republic of Croatia. As a result of the research conducted on a sample of 1,113 respondents, it was found that the positive aspects of student work on academic success are observed only at the graduate level of study. The research results imply that it is necessary to further investigate the impact of student work on academic success.*

**Keywords.** students, student work, academic success

## 1 Introduction

Student work in the Republic of Croatia (RH) is regulated by the Law on Performing Student Jobs, which governs the relationships between the job provider, the intermediary, and the job performer (Zakon o izmjeni zakona o obavljanju studentskih poslova, 2020; Zakon o obavljanju studentskih poslova, 2018). This law equates part-time students with full-time students in terms of their right to work (Zakon o izmjeni zakona o obavljanju studentskih poslova, 2020; Zakon o obavljanju studentskih poslova, 2018) contributing to an increase in the number of students who can work while studying. According to the National Development Strategy 2030 in RH, four development directions are evident: sustainable economy and society, strengthening resilience to crises, green and digital transition, and balanced regional development. One of the strategic goals within the direction of a sustainable economy and society is to educate and employ people (Hrvatski Sabor, 2021).

Future highly educated professionals in RH often gain their first work experience during their academic education by finding employment through student service intermediaries. Western Balkan countries, including the RH, have retained the concept of student work as a special employment status category. In Slovenia, this activity is handled by employment agencies, whereas in RH, the mediation of student work is under the jurisdiction of registered student centers. In RH, student services operate within 13 registered student centers, which are public institutions functioning as other components of universities (Zajednica studentskih centara, n.d.).

The Law on Performing Student Jobs defines the working conditions for students and other individuals (those in the process of enrolling in studies or those who have completed their studies) during their academic and post-academic education in RH (Zakon o obavljanju studentskih poslova, 2018).

According to data from Eurostat (ICT Specialists' Workforce Growing in 2020 - Products Eurostat News - Eurostat, n.d.), the National Development Strategy 2030 in RH (Vlada Republike Hrvatske, 2021), and the student employment mediation data from the Student Center in Varaždin (Studentski centar Varaždin, n.d.), the potential demand for workers in information and communication technology (ICT) and ICT graduates shows continuous growth. There is an increasing need for student labor in all business segments.

The research aim was to determine whether students who work during their academic education achieve better academic success, measured by a higher number of earned ECTS credits and regular progression from year to year.

The Agency for Science and Higher Education conducted a study on the employability of graduates, which included individuals who completed higher education in the academic year 2021/2022. The employment rate one year after graduation was 84.3% (Agency for Science and Higher Education, 2023). The question remains open as to how much paid student work during academic studies contributes to employability.

This research aims to answer the research question: How does paid work through authorized student employment agencies affect academic success?

# 2 Theoretical framework

Student work can be viewed from the perspectives of students, employers, and higher education institutions. This paper focuses on the student, examining the benefits and risks associated with working during studies in the context of academic education. There is a need to investigate how student work affects their academic obligations and whether appropriate policies and strategies should be implemented to adapt academic institutions to the new conditions of student self-financing.

For this research, student work refers to paid student employment mediated by authorized intermediaries in RH for the time of their academic studies. Numerous studies examine the relationship between student work and academic success, with some indicating that students enrolled with government funding achieve greater academic success and shorter average study times compared to self-financing students. Additionally, more students are working, and work has been shown to impact academic success positively (Išljamović et al., 2016; Logan et al., 2016; Nidogon Višnjić et al., 2024).

## 2.1 Previous research

According to the results of the Eurograduate survey on the employability of graduate students, which involved eight European countries, including Croatia, personal drivers of early employability and the study of ICT, as well as prior employment and work experience related to studies, have the greatest impact on the labor markets of individual countries. The research highlights the growing importance of the connection between higher education institutions and labor markets and the necessity to balance traditional methods of academic education with the validation of theoretical knowledge through student work (The Agency for Science and Higher Education, 2000).

In Croatia, as well as at the European Union (EU) level, the issue of employability for young, highly educated individuals and their preparation for transitioning to the labor market is currently significant (European Commission, 2019; Hrvatski Sabor, 2021).

Working during academic education allows students to gain their first work experience. According to Eurostudent, an average of 17 % of all students have difficulties fulfilling their academic obligations due to student work commitments, with those working more than 20 hours per week being particularly affected. This is especially true for students in Estonia, Poland, Croatia, and Ireland (European Commission, 2019).

According to the Systematic Literature Review (SLR), which encompasses 347 studies on the impact of student work on academic success, researchers have demonstrated both positive and negative aspects of this influence. However, the positive effects have been less studied. The positive relationship primarily involves the students' improved ability to connect theory with practice, while the negative impact on academic success is associated with the increased time dedicated to work. Early employment of young people results in better long-term work experiences, and combining education with work during studies positively affects employment and financial independence (Nidogon Višnjić et al., 2024).

Some researchers confirm that student work enriches students by expanding their network of social contacts and support. However, it does not contribute to student-parent bonding because work encourages student independence (Wang et al., 2010).

The increasing number of students who work while studying presents a challenge for higher education institutions. Students who work during their studies can transfer their theoretical knowledge to industry, creating mutual benefits from student work (Davies, 2012; Rochford et al., 2009). The most prevalent group of specific positive impacts of student part-time jobs on academic success relates to the connection between theory and practice that such jobs facilitate (Nidogon Višnjić et al., 2024).

The distinctiveness and diversity of different cultures, the psychosocial characteristics of students, and the financial capabilities of students are crucial factors when deciding whether to work while studying. The best strategy for successfully combining student work with satisfactory academic achievements is part-time work related to the fields of academic education (Mitchell, 2020; Schoffstall & Arendt, 2014).

The asymmetry between students' involvement and satisfaction in their university life and paid work can be altered through curricula designed to link theories and practices, ensuring that students' work experiences contribute to their academic success (Lingard, 2007).

According to the SLR, student work is influenced by market and economic factors of the economic system, the socio-demographic profile of students, the structure of financial support they receive, and the number of obligations they have at higher education institutions (Nidogon Višnjić et al., 2024).

Early employment of young people results in better long-term work experiences, and combining education with work during studies has a positive impact on employment and financial independence. Working during studies improves academic success, but only if paid work does not exceed the threshold of 22 hours per week (Adams, 1978; Applegate & Daly, 2006).

Research indicates that working during studies can have positive effects on academic success; it is not the work itself that negatively impacts learning outcomes, but the number of hours students spend working while managing academic obligations. Work experiences are significantly more beneficial when students are

employed in jobs related to their field of study (Nidogon Višnjić et al., 2024; Rochford et al., 2009; Smith & Schoffstall, 2020; Watts & Pickering, 2000).

The SLR was focused on indicators of academic success, the relationship between student part-time work and academic success, and the drivers of student work. The SLR established that researchers have confirmed better academic success and enhanced generic skills that contribute to improved academic performance as positive characteristics of student work, along with personal and professional development and better execution and mastery of academic obligations (Nidogon Višnjić et al., 2024).

# 3 Methodology

For this study, research was conducted on a random sample consisting of 10 % of all members of the Student Service of the Student Center in Varaždin who had active membership during the year 2020.

Members of student employment services can be full-time and part-time students, as well as other individuals in the process of enrolling in a study program or those who have completed their studies.

However, the latter can exercise their right to work through student employment services for up to three months after the end of the academic year in which they completed their secondary education, until the end of the academic year in which they completed their studies, or until the end of the three months after finishing their studies. An additional condition is that members of student employment services do not have an established employment relationship and/or do not engage in independent activities in trade, freelance professions, and/or agriculture and forestry (Zakon o obavljanju studentskih poslova, 2018).

The initial random sample consisted of 1,113 individuals and members of the Student service. The secondary data were collected from various sources, including databases of the Student Service in Varaždin and the Ministry of Science, Education, and Youth of the Republic of Croatia. Data was collected from student employment service databases, which included: anonymized demographic information, employment status, type of job performed, number of hours worked, and duration of employment for each student contract, university, and study program. Data was collected from the Ministry of Science, Education, and Youth databases about an earned number of ECTS credits, academic year of first enrollment, current academic year of study, and level of study.

During the research process, it was necessary to aggregate all input data by academic year (data was collected in such a way that the information from the student employment service databases was taken for the academic year rather than the calendar year), because student status and academic indicators are tracked on an academic year basis, while financial indicators and student employment status are tracked on a calendar year basis.

Incomplete and outlier data from the initial sample were discarded based on predefined criteria because they were found to be the result of administrative errors and significantly skewed the initial sample. Incomplete and outlier data, such as students with 270 earned ECTS credits, students enrolled in double-major programs (e.g., Faculty of Humanities and Social Sciences, University of Zagreb; Faculty of Humanities and Social Sciences, University of Rijeka; Croatian studies, University of Zagreb), changes in university names during the same academic year (e.g., University of Slavonski Brod), students who study at two different universities, students studying outside Croatia, and older non-traditional students, had to be excluded from the statistical analysis.

For this case study, a hypothesis was formulated.

H: Student work during academic education is positively associated with academic success.

Academic success in this study is observed through the regular progression of students and the number of earned ECTS credits. Regular progress in this work is considered if a student enrolls in the next year of study each academic year, meaning they do not repeat the same year multiple times.

An important assumption of scientific work is obtaining all necessary consent from data owners for data processing, analysis, and interpretation for scientific purposes, as well as ensuring that the data are fully anonymized.

Consent has been obtained from the Croatian Personal Data Protection Agency regarding the method of data collection and processing. Students represented the observations, while academic success was defined as the dependent variable.

# 4 Research results

The collected data were processed using the SPSS software interface.

The normality of the distribution was tested using the Shapiro-Wilk test, and it was found that the empirical results do not significantly deviate from a normal distribution. Therefore, means were calculated as measures of central tendency, and standard deviations were used as measures of dispersion for student work.

**Table 1.** The number of total and valid cases, the number of students who worked and who did not work

| | Did the student work? | Primary Case | Total |
|---|---|---|---|
| VALID | NO | 117 | 201 |
| | YES | 996 | 1,949 |
| **Total** | | **1,113** | 2,150 |

Table 1 shows the total number of cases and the valid number of observations of students who entered the random sample of the research subject (of students who did and did not work yes/no in Table 1).

The initial sample of 2,150 records from the Student Employment Service in Varaždin was reduced to 1,113 members. This reduction occurred because some users had a higher number of completed student contracts or a greater number of jobs performed during the observed period. The sample of 1,113 students was additionally reduced to 912 students because only students with registered or at least 1 ECTS credit or a rolled academic year were considered (Table 2).

**Table 2.** The number of students by the number of years of study

| Year of study | Number of students | Percentage |
|---|---|---|
| 1 | 236 | 25,88% |
| 2 | 251 | 27,52% |
| 3 | 203 | 22,26% |
| 4 | 116 | 12,72% |
| 5 | 58 | 6,36% |
| 6 | 22 | 2,41% |
| 7 | 14 | 1,54% |
| 8 | 3 | 0,33% |
| 9 | 4 | 0,44% |
| 10 | 1 | 0,11% |
| 11 | 2 | 0,22% |
| 12 | 2 | 0,22% |
| Total | 912 | 100% |

Table 2 provides an overview of the obtained results according to the number of years of study. It presents the number of students for each observed calendar year at the time of sampling from the student database. The initial sample of 1,113 students was reduced to 912 students because only students with registered or at least 1 ECTS credit or a rolled academic year were considered.

**Table 3.** The number of students in the current year of study

| Currently enrolled year of study | Number of students | Percentage |
|---|---|---|
| 1 | 261 | 28,62% |
| 2 | 301 | 33,00% |
| 3 | 281 | 30,81% |
| 4 | 39 | 4,28% |
| 5 | 27 | 2,96% |
| 6 | 3 | 0,33% |
| Total | **912** | 100% |

Table 3 presents an overview of the obtained results, including the number and percentage of students in the currently enrolled year of study at the time of data collection from the student database.

**Table 4.** The number of students by level of study

| Level of study | Number of students | Percentage |
|---|---|---|
| Professional study | 11 | 1,21% |
| Undergraduate studies | 547 | 59,98% |
| Integrated undergraduate and graduate studies | 115 | 12,61% |
| Specialist graduate study | 28 | 3,07% |
| Graduate study | 211 | 23,14% |
| Total | 912 | 100,00% |

Table 4 provides an overview of the number of enrolled students and their percentage by level of study. In the first phase of testing students at the undergraduate level, a sample of 553 students of professional and undergraduate studies studying for up to 7 years is taken (Table 4, first two lines).

Students studying for more than 7 years at the undergraduate level (in this case, 5 of them) were excluded from the sample. In this way, from a sample of 558 students, we reached 553 students. Necessary

tests were conducted for the obtained sample as prerequisites for further analysis, i.e., measures of central tendency.

**Table 5.** Mean, standard deviation, and standard error mean

| | St. contract | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Number of achieved ECTS | yes | 285 | 96,31 | 53,862 | 3,19 |
| | no | 228 | 99,44 | 54,62 | 3,617 |
| Regular transition to the next year of study | yes | 302 | 2,54 | 1,303 | 0,075 |
| | no | 243 | 2,75 | 1,504 | 0,097 |

Table 5 displays measures of central tendency for the academic success of students observed through the number of earned ECTS credits and regular progression, i.e., enrollment in the next year of study. Students who had achieved ECTS and students who had a regular transition to the next year of study were observed separately.

**Table 6.** Testing the difference depending on the number of earned ECTS credits and regular enrollment in the next year of study for all undergraduate students

| | | T-test | | |
|---|---|---|---|---|
| | | t | df | p |
| Number of achieved ECTS | Equal variances assumed | 0,650 | 511 | 0,516 |
| | Equal variances not assumed | 0,649 | 483,621 | 0,517 |
| Regular transition to the next year of study | Equal variances assumed | 1,766 | 543 | 0,078 |
| | Equal variances not assumed | 1,739 | 481,346 | 0,083 |

According to Table 6, there is no significant statistical difference when considering all undergraduate students in the sample collectively.

**Table 7.** The number of achieved ECTS and the regular transition to the next academic year for first, second, and third-year undergraduate students

| | | Undergraduate level | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | First year | | | Second year | | | Third year | | |
| | | t | df | p | t | df | p | t | df | p |
| number of achived | equal variances assumed | 1,672 | 124 | 0,097 | 1,151 | 134 | 0,252 | 0,665 | 249 | 0,507 |
| | equal variances not assumed | 1,569 | 83,929 | 0,120 | 1,141 | 120,519 | 0,256 | 0,653 | 215,395 | 0,515 |
| regular transition | equal avriances assumed | 1,317 | 143 | 0,190 | 0,267 | 138 | 0,790 | 0,149 | 258 | 0,252 |
| | equal variances not assumed | 1,255 | 104,067 | 0,212 | 0,261 | 115,619 | 0,795 | 1,135 | 232,284 | 0,258 |

The data analysis shows that there is no statistically significant difference in the number of ECTS credits earned and regular enrollment in the next academic year between first-year, second-year, and third-year undergraduate students who worked through student services during their studies and those who did not, as shown in Table 7.

**Table 8.** The number of achieved ECTS and the regular transition to the next academic year for first-year and second-year graduate students

| | | Graduate level | | | | | |
|---|---|---|---|---|---|---|---|
| | | First year | | | Second year | | |
| | | t | df | p | t | df | p |
| Number of achieved ECTS | equal variances assumed | 0,137 | 89 | 0,892 | 0,318 | 131 | 0,751 |
| | equal variances not assumed | 0,145 | 86,952 | 0,885 | 0,328 | 119,352 | 0,744 |
| Regular transition to the next year of study | equal variances assumed | 3,172 | 97 | 0,002 | 1,913 | 135 | 0,058 |
| | equal variances not assumed | 3,668 | 60,345 | 0,001 | 2,060 | 134,342 | 0,041 |

The data analysis shows that there is a statistically significant difference in the regular transition to the next academic year between first-year graduate students who were employed during their studies and those who were not ($t = 3,172$, $df = 97$, $p < 0.05$). However, no statistically significant difference was found in the regular transition to the next academic year between second-year graduate students who worked and those who did not work, and overall, there is no significant difference in the number of ECTS credits earned between students who worked and those who did not work during their graduate studies.

# 5 Discussion

Based on this case study, whose objective was to determine whether students who work during their academic education achieve better academic performance as observed through a higher number of earned ECTS credits and regular progression, it can be confirmed that there are no differences in study success (ECTS, regular progression) between those of students who have and those who have not worked at the undergraduate level of studies.

On the other hand, regarding graduate-level students, a statistically significant difference was found in the regular transition to the next academic year between first-year graduate students who worked through the student service during their studies and those who did not work. As for the earned ECTS credits, no significant difference was found between first and second-year graduate students. Therefore, we can say that the initial hypothesis was only partially confirmed.

The research results show that students who worked during the observed period did not earn a higher number of ECTS credits compared to students who did not work during the same period, regardless of whether all undergraduate students are considered collectively or individually by academic years. A trend can also be observed that the regular transition to the next year of study and the higher number of achieved ECTS are slightly higher for students who worked for graduate-level students.

The results of the study provide insight into the proportion of variance in student work engagement, as a criterion for academic success, and it has been concluded that predictions for more extensive research can be obtained by examining the variance.

It is noticeable that the smaller sample of respondents by the observed years of study leads to a smaller difference according to the observed groups and that the coefficient of asymmetry is high in the protruding data. These findings are by (Logan et al., 2016; Rochford et al., 2009) that student work has no negative consequences on academic success. The results of this case study confirm the conclusions of the SLR (Nidogon Višnjić et al., 2024) that student work conditionally contributes to better academic success and that this is a positive characteristic of student work.

The limitation of this work is the fact that the number of working hours of students in the research sample was not observed. Additionally, the social background of the participants was not considered in this case study.

The case study has also confirmed that for more precise research, a larger statistical sample and historical data should be taken into account due to the specific circumstances of the pandemic in 2020 (COVID-19).

According to the results obtained in this case study, the hypothesis that student work during academic education is positively related to academic success can only be partially confirmed, and this is evident at the graduate level. However, there remains a need for more comprehensive research in this area.

# 6 Conclusion

Working during academic education has both positive and negative aspects and can impact students' academic achievements. It is necessary to investigate whether the positive or negative effects of work prevail.

This case study on a sample of 1,113 students aimed to prove the positive impact of student work during academic education on academic success. The positive impact is visible only in the first year of graduate studies if it is observed through regular progression. However, there is no negative impact of student work on academic success at the undergraduate level. It is obvious that students have enough time for occasional work in addition to regular academic duties, but graduate-level students probably balance academic and work duties better.

The limitation of this research is it a too small a random sample that cannot be considered representative, and there remains a need to repeat the research on a representative sample of students and include additional variations, i.e. the number of working hours, historical data and the social background of the participants. The contribution of this work lies in the fact that it lays the foundation for a comprehensive study that needs to consider all factors influencing student work during academic education.

Given that an increasing number of students are forced to work during their academic studies, it is assumed that academic obligations in certain areas need to be adjusted to students' work commitments. It remains for students to balance their academic and work obligations and try to align them in the best possible way, with support from higher education institutions and employers themselves.

# References

Adams, A. V. |And O. (1978). *The Lingering Crisis of Youth Unemployment.* (First Edit). W. E. Upjohn Institute for Employment Research, 300 South Westnedge Avenue, Kalamazoo, Michigan

Agencija za znanost i visoko obrazovanje. (2000). *Objavljeni rezultati europskog istraživanja o zapošljivosti diplomiranih studenata Eurograduate*. Retrieved April 15, 2024, from https://www.azvo.hr/hr/azvo-vijesti/2395-objavljeni-rezultati-europskog-istrazivanja-o-zaposljavanju-diplomiranih-studenata-eurograduate-u-kojem-je-sudjelovala-i-hrvatska

Agencija za znanost i visoko obrazovanje. (2023). *Naslovna - AZVO*. Rezultati Istraživanja o

Zapošljivosti Studenata Koji Su Diplomirali Akademske Godine 2021. / 2022. Retrieved April 20, 2024, from https://www.azvo.hr/

Applegate, C., & Daly, A. (2006). The impact of paid work on the academic performance of students: A case study from the University of Canberra. *Australian Journal of Education*, *50*(2), 155–166. https://doi.org/10.1177/000494410605000205

Davies, J. (2012). Using part-time students to improve the student experience. *Strategies for Transition, Induction, and Retention of Part-Time Students*.

European Commission. (2019). *The social dimension of student life in the European higher education area in 2019*. Retrieved April 15, 2024, from http://creativecommons.

Hrvatski Sabor. (2021). *Nacionalna razvojna strategija Republike Hrvatske do 2030. godine*. Narodne Novine. Retrieved April 15, 2024, from https://narodne-novine.nn.hr/clanci/sluzbeni/2021_02_13_230.html

*ICT specialists' workforce growing in 2020 - Products Eurostat News - Eurostat*. (n.d.). Retrieved January 24, 2024, from https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/ddn-20210716-1

Išljamović, S., Jeremić, V., & Lalić, S. (2016). Pokazatelji uspješnosti studiranja u korelaciji s tipom financiranja prilikom upisa na fakultet. *Croatian Journal of Education*, *18*(2), 583–606. https://doi.org/10.15516/cje.v18i2.1003

Lingard, H. (2007). Conflict Between Paid Work and Study: Does it Impact upon Students' Burnout and Satisfaction with University Life? *Journal for Education in the Built Environment*, *2*(1), 90–109. https://doi.org/10.11120/jebe.2007.02010090

Logan, J., Hughes, T., & Logan, B. (2016). Overworked? An Observation of the Relationship between Student Employment and Academic Performance. *Journal of College Student Retention: Research, Theory and Practice*, *18*(3), 250–262. https://doi.org/10.1177/1521025115622777

Mitchell, J. (2020). Juggling employment and studies: Nursing students' perceptions of the influence of paid employment on their success. *Nurse Education Today*, *92*, 104429. https://doi.org/10.1016/J.NEDT.2020.104429

Nidogon Višnjić, S., Pažur Aničić, K., & Divjak, B. (2024). A systematic review of the literature on student work and academic performance. *Industry and Higher Education*. https://doi.org/10.1177/09504222241241974

Rochford, C., Connolly, M., & Drennan, J. (2009). Paid part-time employment and academic performance of undergraduate nursing students. *Nurse Education Today*, *29*(6), 601–606. https://doi.org/10.1016/j.nedt.2009.01.004

Schoffstall, D. G., & Arendt, S. W. (2014). Benefits and Challenges Encountered by Working Students. *Journal of Hospitality & Tourism Education*, *26*(1), 10–20. https://doi.org/10.1080/10963758.2014.880614

Smith, R. A., & Schoffstall, D. G. (2020). Employment impact on hospitality and tourism students' academic journey: the HBCU students' perspective. *Journal of Teaching in Travel & Tourism*, *20*(2), 105–120. https://doi.org/10.1080/15313220.2019.1706696

Studentski centar Varaždin. (n.d.). *Studentski centar Varaždin*. Retrieved April 27, 2024, from http://www.scvz.unizg.hr/

Vlada Republike Hrvatske. (2021). *Nacionalna razvojna strategija Republike Hrvatske*. Retrieved April 29, 2024, from https://hrvatska2030.hr/

Wang, H., Kong, M., Shan, W., & Vong, S. K. (2010). The effects of doing part-time jobs on college student academic performance and social life in a Chinese society. *Journal of Education and Work*, *23*(1), 79–94. https://doi.org/10.1080/13639080903418402

Watts, C., & Pickering, A. (2000). Pay as you learn: Student employment and academic progress. *Education + Training*, *42*(3), 129–135. https://doi.org/10.1108/00400910010372670

Zajednica studentskih centara. (n.d.). *Zajednica studentskih centara RH*. Retrieved April 27, 2024, from https://www.zsc.hr/

Narodne Novine. (n.d.-b) Zakon o izmjeni Zakona o obavljanju studentskih poslova, Pub. L. No. 6/2020 (2020). Retrieved April 22, 2024, from https://narodne-novine.nn.hr/clanci/sluzbeni/2020_02_16_329.html

Narodne Novine. (n.d.-b). Zakon o obavljanju studentskih poslova. Pub. L. No. 96/2018 (2018). Retrieved April 22, 2024, from https://narodnenovine.nn.hr/clanci/sluzbeni/2018_10_96_1851.

# Relationship between factors that contribute to student success in the e-learning environment

**Marko Matus, Igor Balaban**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{marko.matus, igor.balaban}@foi.unizg.hr

***Abstract**. This qualitative research study identifies and investigates the relationship between key factors influencing student success in e-learning environments. It emphasizes the importance of identifying key factors to enhance educational practices and student outcomes. By applying Interpretive Structural Modelling (ISM) to factors already identified as crucial in previous literature, this research reveals the interdependencies and impact levels among factors such as faculty guidance, resource availability, and continuous assessment. The analysis, supported by Structural-Self Interaction Matrix (SSIM) and MICMAC diagram highlights that 10 of 12 factors are interconnected and therefore depend on each other, that is, the change of one factor affects the others. Low attendance was identified as a factor that depends on all other factors.*

**Keywords.** E-learning, Interpretive Structural Modelling, ISM, student performance, student success

## 1 Introduction

Ensuring quality and enhancing student performance in the e-learning context had become key factors in monitoring the progress of student attributes. Tracking student performance was not merely about collecting data on students but served as a mechanism for evaluating learning and adjusting teaching materials and methods to develop students' full potential. E-learning systems like Moodle typically collected large amounts of data that needed to be analysed. However, it was very challenging to isolate the precise factors that required special attention. Raw data was often used to predict students' test results or various forms of assessments, but it was less commonly used to identify factors or behavioural patterns (Chen, 2023) that needed attention to identify students who required additional support or to further enhance successful practices and encourage the continuous improvement of the educational process (Kumar, 2021).

By using different factors such as students' historical performance and behaviour, and studying the relationship between these factors and grades, it was possible to form a prediction model to assess students' learning effectiveness and then infer future student performance. This provided a crucial basis for academic warnings, adjustment of teaching strategies, optimisation of educational resource allocation, and personalised customisation of learning plans for students (Yan, 2022). Monitoring student performance was important not only for students but also for teachers and broader educational institutions such as universities. To improve learning success and evaluate the effectiveness of educational methods, it was essential to understand the social, cognitive, and behavioural aspects of students, enabling teachers to enhance their teaching by adapting to the individual needs of students, and educational institutions to assess and improve the effectiveness of their programmes by predicting academic success and adjusting the curriculum according to industry requirements (Kuna & Prasad, 2019).

In this paper, we focused on the factors identified in the study proposed by Bharadi & Awasthi (2022), specifically those marked as having a positive or negative impact on student performance, to examine whether and how these factors affected their success. For this purpose, Interpretive Structural Modelling (ISM) was used. The method was interpretive in that the group's judgement decided whether and how items were related; it was structural in that, based on the relationship, an overall structure was extracted from the complex set of items; and it was modelling in that the specific relationships and overall structure were portrayed in a diagram model (Sage, 1977).

## 2 Research objectives and methodology

The aim of this research is:

- To identify how selected factors contribute to student performance.

- To determine the relationships between those factors.

And to answer the following research question:

RQ: In what way do the identified factors contribute to student success?

Interpretive Structural Modeling (ISM) is a computer-aided method for developing graphical representations of system composition and structure (Attri et al., 2013). ISM not only aided in understanding the relative relationship between the critical success factors but also helped to build their interdependence while implementing sustainability (Sreenivasan et al., 2023). When a problem was identified within a system, the first step involved identifying all relevant elements or factors that were part of the system being studied. Having decided on the factor set and the contextual relation, a Structural Self-Interaction Matrix (SSIM) was developed based on a pairwise comparison of factors (Attri et al., 2013). The SSIM was vital in ISM as it systematically captured and represented the direct relationships between elements within a system. This matrix served as the foundation for constructing the reachability matrix, which in turn allowed for the hierarchical structuring of elements, enabling the identification of key drivers and dependencies within complex systems. Once transitivity embedding was complete, a matrix model was obtained. This allowed for the partitioning of elements and the extraction of the structural model known as ISM (Agarwal et al., 2007).

As mentioned in the previous section, factors selected for the ISM are taken from Bhardi & Asasthi (2022). The authors, based on a literature review, identified 33 potential independent variables that could influence student performance in an e-learning environment. After a more detailed analysis, the following factors were identified as having either positive or negative impact on student performance (Bhardi & Asasthi, 2022):

1) Trained Faculty guidance
2) Resource Availability (Labs, Library etc)
3) Continuous Assessment (Tests) During Course
4) Student's focus on career and self-grooming
5) Use of New Educational Tools and Techniques in Teaching (Smartboards, Portals etc)
6) Prior in-depth knowledge about current subjects (Prerequisites)
7) Score/Performance in Prior Course
8) Extra-Curricular Activities Participation
9) Social Media Platforms utilization
10) Peer Pressure
11) Parent's involvement in student's study
12) Low Attendance.

As presented by Bhardi & Asasthi (2022), in the e-learning context, factors 1-7 had a positive impact on student's performance, while factors 8-12 provided a negative impact. It was also found that those factors had different impacts on a student's educational performance. For example, Trained Faculty Guidance and the Use of New Educational Tools and Techniques (like smartboards and online portals) are critical in delivering effective and engaging online education. On the other hand, external factors such as Peer Pressure or Parent's involvement in student's study may also impact student's performance, but not to such an extent. It can also be noted that the factors the authors consider positive are related to the teaching process (teaching staff, resources, assessments, student career development, new educational technologies...) while the factors identified by the authors as having a negative impact on student performance are related to external factors (extracurricular activities, social media use, peer pressure, parental involvement...).

# 3 Participants and data collection

To identify the correct relationships between the individual factors, it was necessary to ensure understanding and seriousness when completing the SSIM matrix. To this end, efforts were made to avoid undergraduate students, as they could not reflect on those factors since they do not have adequate experience and consequently, cannot develop proper attitudes towards different factors influencing their study. Therefore, postgraduate students were deemed more representative for identifying relevant factors in the SSIM analysis, as they already possess certain research knowledge, and some have even experience working with other students. Additionally, all postgraduate students have completed their master's degree, which allows them to assess better which factors more or less influence student success. In contrast, first-year students often lack sufficient experience in theoretical and practical research aspects, which could lead to more superficial or inaccurate results. For this purpose, a task was prepared in the LMS Moodle, which the students downloaded to their computers and then re-uploaded to the Moodle system once completed. To ensure students had enough time to complete the task, the deadline for submitting the task was set to 2 days, which aimed to ensure accurate and honest responses. A total of 17 SSIM matrices were completed, with 17 students participating in the research. The SSIM matrix factors refer to the 12 factors previously mentioned, which were identified by Bhardi & Asasthi (2022) as factors that can significantly influence student performance within an e-learning environment.

# 4 The ISM procedure

The procedure started with identification and analysis of relationships and interactions between different factors identified as important for student success in previous literature. The relationships between factors established by the participant's responses were presented using a standard ISM notation:
V: factor i will assist to reach factor j;
A: factor j will assist to reach factor i;
X: factor i and j will assist to reach each other;
O: factors j and i are unrelated (Pramod & Banwet, 2010).

Using the ISM notation described above, a SSIM was generated (Table 1) where the relationships between factors are represented in terms of rows (i) and columns (j) along with their respective connections. After the students expressed the degree of connection between the two factors (using standard ISM notation symbols), their responses were compared, and the most common values were selected using the mode function. The final SSIM table is based on the mode function, meaning that for each value at the intersection of (i) row and (j) column the most frequent value was used. The final SSIM is presented in Table 1 where factors from 1-12 present independent variables mentioned before.

**Table 1.** SSIM

| Factors | (j) | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (i) | | | | | | | | | | | | | |
| 1 | | O | O | O | O | V | X | O | X | X | X | X | 1 |
| 2 | | V | O | O | O | X | X | O | X | V | V | 1 | |
| 3 | | V | O | O | O | O | X | A | X | V | 1 | | |
| 4 | | O | O | O | O | X | A | X | X | 1 | | | |
| 5 | | O | O | O | X | O | O | O | 1 | | | | |
| 6 | | V | O | O | O | O | A | 1 | | | | | |
| 7 | | V | O | O | O | X | 1 | | | | | | |
| 8 | | V | O | A | X | 1 | | | | | | | |
| 9 | | V | O | X | 1 | | | | | | | | |
| 10 | | V | O | 1 | | | | | | | | | |
| 11 | | O | 1 | | | | | | | | | | |
| 12 | | 1 | | | | | | | | | | | |

## 4.1 Reachability matrix

The next stage in the ISM methodology involved creating an initial reachability matrix from the SSIM. This process entailed transforming the SSIM into the Initial Reachability Matrix (IRM) by replacing the four symbols (V, A, X, and O) with 1s or 0s. IRM is created by converting symbolic SSIM matrices into binary format using the ISM rules (Tabziri et al., 2010). The Initial Reachability Matrix is created from the SSIM by applying the following rules: if the SSIM entry is "V," the matrix entry is (1,0); if "A," it is (0,1); if "X," it is (1,1); and if "O," it is (0,0), where the pairs represent the direction of influence between the factors. With the mentioned guidelines, the initial reachability matrix was constructed. To address any gaps in the opinions gathered during the creation of the structural self-interaction matrix, 1* entries were added to ensure transitivity. Transitivity can be described as follows: if factor A leads to another factor B and if factor B leads to a third factor C, as per the rule of transitivity A leads to C also exists (George & Pramod., 2014). Once the transitivity concept had been applied as outlined, the Final Reachability Matrix (FRM) was obtained, and presented in Table 2. An example of transitivity is observed when Trained Faculty Guidance (factor 1) influences Resource Availability (factor 2), and Resource Availability (factor 2) influences Continuous Assessment (factor 3); therefore, Trained Faculty Guidance (factor 1) indirectly influences Continuous Assessment (factor 3) through Resource Availability (factor 2). This example presents that 1* would be placed at the intersection of Trained Faculty Guidance (1) and Continuous Assessment (3) in the Final Reachability Matrix, indicating that although there is no direct relationship between these two variables, there is an indirect influence established through Resource Availability (2).

**Table 2.** FRM

| Factors | (j) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Driving power |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (i) | | | | | | | | | | | | | | |
| 1 | | 1 | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1* | 1* | 0 | 1* | 11 |
| 2 | | 1 | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1* | 1* | 0 | 1 | 11 |
| 3 | | 1 | 1* | 1 | 1 | 1 | 1* | 1 | 1* | 1* | 1* | 0 | 1 | 11 |
| 4 | | 1 | 1* | 1* | 1 | 1 | 1 | 1* | 1 | 1* | 1* | 0 | 1* | 11 |
| 5 | | 1 | 1 | 1 | 1 | 1 | 1* | 1* | 1* | 1 | 1* | 0 | 1* | 11 |
| 6 | | 1* | 1* | 1 | 1 | 1* | 1 | 1* | 1* | 1* | 1* | 0 | 1 | 11 |
| 7 | | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1 | 1* | 1* | 0 | 1 | 11 |
| 8 | | 1* | 1 | 1* | 1 | 1* | 1* | 1 | 1 | 1 | 1* | 0 | 1 | 11 |
| 9 | | 1* | 1* | 1* | 1* | 1 | 1* | 1* | 1 | 1 | 1 | 0 | 1 | 11 |
| 10 | | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1 | 1 | 1 | 0 | 1 | 11 |
| 11 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 12 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Dependence power | | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 1 | 11 | |

From the FRM, reachability and antecedent sets were derived. The reachability set contained the factor itself and any other factors it might affect, while the antecedent set included the factor itself and any other factors that may affect it. The intersection of these sets was then determined for all factors, allowing for the identification of different factor levels. Factors where the reachability and intersection set match were placed at the top level of the ISM hierarchy. Top-level factors did not influence any factors above their level. Once identified, they were excluded from further consideration. This process was repeated to determine the next level of factors. In our example shown in Table 3, only 2 levels were detected which later will correspond to the levels of the diagram.

**Table 3.** Level Partitioning (LP)

| Factor | Reachability Set | Antecedent Set | Intersection | Level |
|---|---|---|---|---|
| 1 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 2 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 3 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 4 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 5 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 6 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 7 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 8 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 9 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 1,2,3,4,5,6,7,8,9,10 | 2 |
| 11 | 11 | 11 | 11 | 1 |
| 12 | 12 | 1,2,3,4,5,6,7,8,9,10,12 | 12 | 1 |

# 5 Results

The Conical Matrix (CM) presented in Table 4. helped construct hierarchical linkages among factors and is a restructured version of the reachability matrix. It made the representation of direct and indirect relationships simpler by organising factors according to their degrees of influence. A more organised and perceptive examination of complex systems is made possible by this matrix rearrangement, which improves the clarity of factor interdependencies. In summary, driving power indicates how strongly a factor influences other factors, while dependence power shows how strongly is one factor influenced by other factors (Balaban, 2020).

**Table 4.** Conical Matrix (CM)

| F (j) / (i) | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | DRP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **11** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **12** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **1** | 0 | 1* | 1 | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1* | 1* | 11 |
| **2** | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1* | 1* | 11 |
| **3** | 0 | 1 | 1 | 1* | 1 | 1 | 1 | 1* | 1 | 1* | 1* | 1* | 11 |
| **4** | 0 | 1* | 1 | 1* | 1* | 1 | 1 | 1 | 1* | 1 | 1* | 1* | 11 |
| **5** | 0 | 1* | 1 | 1 | 1 | 1 | 1 | 1* | 1* | 1* | 1 | 1* | 11 |
| **6** | 0 | 1 | 1* | 1* | 1 | 1 | 1* | 1 | 1* | 1* | 1* | 1* | 11 |
| **7** | 0 | 1 | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1 | 1* | 1* | 11 |
| **8** | 0 | 1 | 1* | 1 | 1* | 1 | 1* | 1* | 1 | 1 | 1 | 1* | 11 |
| **9** | 0 | 1 | 1* | 1* | 1* | 1* | 1 | 1* | 1* | 1 | 1 | 1 | 11 |
| **10** | 0 | 1 | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1 | 1 | 1 | 11 |
| **DEP** | 1 | 11 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | |
| **Level** | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | |

* DEP: Dependence Power; DRP: Driving Power; F: Factors

## 5.1 Classification of factors

To analyse and classify factors based on their driving power and dependence, a Matrics d'Impacts Croises-Multiplication Applique an Classment (MICMAC) diagram was constructed and presented in Fig. 1. The diagram categorizes factors into four main groups:

I. Autonomous factors: Have weak driving power and weak dependence.
II. Dependent factors: Have high dependence but weak driving power.
III. Linkage factors: Have both high driving power and dependence. These factors are considered unstable because any intervention on them affects other factors and triggers feedback effects on themselves (Balaban, 2020).
IV. Independent factors: Have strong driving power and weak dependence.

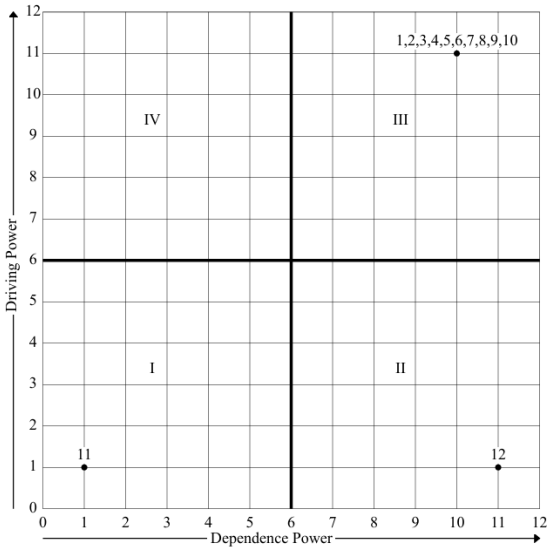



***Figure 1. MICMAC diagram***

As seen in Fig. 1, in the first quartile there was only a factor number 11 which had weak driving power and also weak dependence. This factor is called “Parent's involvement in student's study”. This is also a factor with a negative impact on student performance divided by Bhardi & Asasthi (2022). It means that “Parent's involvement in student's study” does not heavily influence other factors, and also is not heavily influenced by other factors. Also, in the “Dependent factor” quartile, there was only one factor called “Low Attendance”, which means that this factor has high dependence but weak driving power and was also one with a negative impact on student performance. This also means that this factor is very influenced by other factors but does not significantly influence other factors. The third quartile is a home of factors from 1 to 10. Those factors have both high driving power but

also dependence. Those factors are very influenced by other factors in the system, but also very influence other factors: 1) Trained Faculty guidance 2) Resource Availability (Labs, Library etc), 3) Continuous Assessment (Tests) During Course, 4) Student's focus on career and self-grooming 5) Use of New Educational Tools and Techniques in Teaching ( Smart boards, Portals etc) 6) Prior in-depth knowledge about current subjects (Prerequisites) 7) Score/Performance in Prior Course 8) Extra-Curricular Activities Participation 9) Social Media Platforms utilization 10) Peer Pressure. In comparison with the results of the authors Bhardi & Asasthi (2022), all our factors from 1 to 10 are highly dependent on each other, which does not align with the author's previous findings, and give an answer to our research question. Differences in research results may arise due to the different research contexts as well as the methodology, as the authors Bhardi & Asasthi (2022) did not use ISM. Additionally, our research was conducted with postgraduate respondents who had a slightly longer period to complete the matrix. Additionally, the factors identified as having a positive impact on student success by Bhardi & Asasthi (2022) are present in the 3rd quartile in this study and strongly influence each other. The same is true for factors 8-10, which have been identified as factors that negatively impact student success.

## 5.2 Relationship between factors

The final step in ISM was creating the ISM diagram which was utilized to illustrate the complex relationships between factors within a system. Nodes in this directed graph represented factors, and the directed lines indicated the relative influence of each factor on other factors. The diagram facilitated identification of important factors and dependent factors by helping to see the hierarchical structure and interdependencies. It was created from the Conical Matrix shown in Table 4 and was presented in Fig. 2.

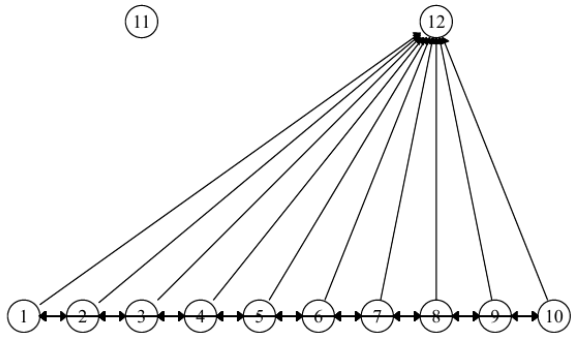



**Figure 2.** ISM diagram

# 6 Conclusion

The study has effectively employed Interpretive Structural Modelling (ISM) to analyse the factors impacting student performance in an educational setting. By utilizing the ISM methodology, we were able to systematically identify and understand the interrelationships among various factors that contribute to students' performance.

The results, encapsulated in the reachability and conical matrices, as well as the MICMAC diagram, provide a clear hierarchical structure and classification of these factors based on their driving and dependence power. Notably, the study highlights that most factors (1-10) have fallen into the third (Linkage) quadrant in which factors have both high driving power and dependence. These factors are considered unstable because any intervention on them affects other factors and triggers feedback effects on themselves (Balaban, 2020). Also, it can be concluded that factors: 1) Trained Faculty guidance 2) Resource Availability (Labs, Library etc), 3) Continuous Assessment (Tests) During Course, 4) Student's focus on career and self-grooming 5) Use of New Educational Tools and Techniques in Teaching ( Smart boards, Portals etc) 6) Prior in-depth knowledge about current subjects (Prerequisites) 7) Score/Performance in Prior Course 8) Extra-Curricular Activities Participation 9) Social Media Platforms utilization 10) Peer Pressure depend on each other in a great extent, and any action on any of them can affect other factors.

On the other hand, 'Parent's involvement in student's study' emerged as an autonomous factor with weak driving and dependence power, while 'Low Attendance' was identified as a dependent factor, heavily influenced by other factors but exerting minimal influence itself.

The ISM diagram further illustrated the hierarchical and directional relationships among the factors, providing a visual representation that aids in comprehending the underlying dynamics. This model serves as a valuable tool for educators and administrators, enabling them to pinpoint key areas for intervention and strategic improvements.

In conclusion, it can be noted that by identifying the key drivers and dependencies, educational institutions can tailor their strategies to foster a more effective learning environment, thereby improving educational outcomes and addressing the specific needs of their student populations. Future research could expand on this study by incorporating a larger and more diverse sample size, exploring additional factors, and integrating longitudinal data to further validate and refine the ISM model. Along that, questions for future research could be, how do the key factors in the ISM model interact with each other and what long-term effects do they have on students' performance?

# References

Agarwal, A., Shankar, R., & Tiwari, M. K. (2007). Modeling agility of supply chain. *Industrial marketing management*, 36(4), 443-457.

Attri, R., Dev, N., & Sharma, V. (2013). Interpretive structural modelling (ISM) approach: an overview. *Research journal of management sciences*, 2319(2), 1171.

Balaban, I. (2020). An empirical evaluation of E-Portfolio critical success factors. International *Journal of Emerging Technologies in Learning (iJET)*, 15(4), 37-52.

Bharadi, V., & Awasthi, S. P. (2022, December). Variables identification for Students Performance Prediction. *In 2022 IEEE Bombay Section Signature Conference (IBSSC)* (pp. 1-6). IEEE.

Bharadi, V., & Awasthi, S. P. (2022, December). Variables identification for Students Performance Prediction. *In 2022 IEEE Bombay Section Signature Conference (IBSSC)* (pp. 1-6). IEEE.

Chen, X. (2023). Raw Data Processing Method. In Application of Gray System Theory in Fishery Science (pp. 21-34). Singapore: Springer Nature Singapore.

George, J. P., & Pramod, V. R. (2014). An interpretive structural model (ISM) analysis approach in steel re rolling mills (SRRMS). *International Journal of Research in Engineering & Technology*, 2(4), 161-174.

Kumar, K. (2021). Exploratory Data Analysis for Predicting Student's Grades. *In Soft Computing: Theories and Applications: Proceedings of SoCTA 2020*, Volume 2 (pp. 359-367). Springer Singapore.

Kuna, L., & AV, K. P. (2019). Research Methodologies for Student Performance Evaluation Using Educational Analytics Tools and Approaches. *International Journal of Recent Technology and Engineering (IJRTE)*, 8, 2277-3878.

Pramod, V. R., & Banwet, D. K. (2015). ISM for understanding the enablers of telecom service supply chain. *International Journal of Business Excellence*, 8(5), 537-565.

Sage, A.P., *Interpretive Structural Modelling: Methodology for Large Scale Systems*, McGraw-Hill, New York, NY, 1977, pp. 91-164.

Sreenivasan, A., Ma, S., Nedungadi, P., Sreedharan, V. R., & Raman, R. R. (2023). Interpretive structural modeling: research trends, linkages to sustainable development goals, and impact of COVID-19. *Sustainability*, 15(5), 4195.

Tabrizi R. S., Foong P. Y., Ebrahimi N. (2010). Using Interpretive Structural Modelling to Determine the Relationships among Knowledge Management Criteria inside Malaysian Organizations. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*. 4(12):2270–2275.

Yan, C. (2022). Research on Student Academic Performance Prediction Methods. *Highlights in Science, Engineering and Technology*, 24, 257-263.

# Predicting the Outcomes of PBL: Comparison of Two Methods

**Barbi Svetec, Blaženka Divjak, Damir Horvat, Katarina Pažur Aničić**

University of Zagreb

Faculty of Organization and Informatics

Pavlinska 2, 42000 Varaždin, Croatia

{barbi.svetec,blazenka.divjak,damir.horvat1, katarina.pazur}@foi.unizg.hr

***Abstract****. We used the Random Forest and Decision Trees algorithms to develop predictive learning analytics models for problem- and project-based learning (PBL). Predictive modeling was done on the data collected from two university courses, with a total sample of 309 students. Different phases of PBL were analyzed, with problem-solving found to be more predictable than problem-posing or peer-assessment. Students were divided into three classes based on their performance in PBL, and the RF-based models were found to be the most efficient in predicting the lower performing students. The DT-based models were generally found to be less efficient, but more interpretable.*

**Keywords.** predictive learning analytics, random forest, decision trees, problem-based learning, project-based learning, assessment

# 1 Introduction

Today, in the age of education highly supported and enhanced by technologies, higher education institutions (HEIs) have access to more ample and more diverse student data than ever before (Prinsloo, 2020). Exploiting this opportunity, predictive learning analytics (LA) can contribute to the quality of education by using historical data to make predictions and inferences about possible future outcomes (Susnjak et al., 2022).

Problem- and project-based learning (PBL), in which students work to solve real-life problems, have been recognized as innovative approaches that foster critical thinking, creativity, autonomy, collaboration and interdisciplinarity (Brassler & Dettmers, 2017; Dole et al., 2015; Savery, 2006). Nevertheless, educators are still challenged by course structuring, student progress monitoring, and giving guidance. LA provides a promising new perspective in terms of course design and monitoring. (Wang et al., 2023)

Some studies have been conducted at the intersection of PBL and LA. For example, a study investigating the key factors predicting online PBL performance (Wang et al., 2023) identified self-regulation, posted messages, message words, and peer-learning engagement as predictors. However, the overall body of research combining PBL and LA to support the development of data-driven student-centered courses, and especially linking LA with particular steps of PBL, is still relatively scarce (Wang et al., 2023).

The aim of this study was to investigate the predictive power of parts of the assessment program on PBL performance, and to analyze the efficiency of predictive models based on two prediction algorithms (Random Forest and Decision Trees).

# 2 Theoretical Background

## 2.1 Predictive Learning Analytics

An important area in LA refers to forecasting academic outcomes based on patterns in past and present data. This *predictive* LA strongly relies on machine learning and deep learning algorithms, learning from historic datasets and using various kinds of student data to make predictions. (Sghir et al., 2023; Susnjak et al., 2022) Predictive LA can provide important support to achieving LOs, informing learning design, identifying at-risk students, as well as increasing students' satisfaction (Sghir et al., 2023). Predictive modeling, as an important practice in LA, is done with the use of several algorithms: in particular, the most often used Artificial Neural Networks are followed by the Random Forest (RF) and Gradient Boosting algorithms, and these algorithms were also found to have the highest prediction accuracy compared to other algorithms (Sghir et al., 2023). Particularly, an educational research study (Kabathova & Drlik, 2021) found that the RF classifier was associated with the

best accuracy and precision compared to Logistic Regression, Support Vector Machine, Decision Tree (DT), Neural Networks and Naïve Bayes. The DT algorithm was among the algorithms that came out second best.

Some research has been conducted to identify the most important predictors. Importantly, it has been found that cognitive data (entry test and quiz scores) present the best predictors, as opposed to basic LMS data (Tempelaar et al., 2015). Research has also shown that formative assessment is an important predictor of students' performance (Bulut et al., 2023; Divjak et al., 2024).

## 2.2 Problem- and Project-Based Learning

Problem- and project-based learning are student-centered, constructivist approaches to teaching and learning (Dole et al., 2015). Both are based on activities done by students collaboratively, independently, in order to achieve a mutual goal (Brassler & Dettmers, 2017; Savery, 2006). Problem-solving is usually done in teams, and this peer-learning is followed by peer-assessment. While in problem-based learning students investigate to solve ill-structured problems, not necessarily leading to a concrete artifact (Savery, 2006), in project-based learning, students' problem-solving is oriented towards the production of a final artifact (Dole et al., 2015). Moreover, while guidance is important in both approaches, in project-based learning, teachers provide more direct instructions to students on how to produce the final artifact, but still fostering self-regulated learning (Savery, 2006). In the problem-solving context, we distinguish between problem-posing and problem-solving, with the former found to be more demanding for students, as it is something they are less exposed to (Divjak, 2015). Problem-posing generally refers to students generating new problems, which can be based on given situations or existing problems (Cai & Rott, 2024).

For the aim of this paper, the two approaches will be referred to in conjunction (using the abbreviation "PBL"), as the aspects relevant for this research are common to both.

# 3 Methodology

The focus of this study has been on two research questions:

1. In these PBL case studies, how do different prediction methods compare in terms efficiency and explainability?
2. What elements of an assessment program are the best predictors of PBL performance?

## 3.1 Study Setting and Data Collection

The study was conducted at the University of Zagreb, Faculty of Organization and Informatics, a HEI offering study programs in ICT, and included data from two courses, collected in the academic year 2022/2023. The courses were chosen so as to represent two different levels of study (undergraduate and graduate) and two different subject areas (ICT and Mathematics), include a considerable PBL experience for students, as well as to have a number of students large enough to enable meaningful machine learning data analysis.

At the undergraduate level, the study included the Informatics Services Management (ISM) course (second year). The course has a student workload of 4 ECTS credits. The assessment program includes formative assessment (assignments done in laboratory exercises) and summative assessment (two periodical exams, a PBL assignment in the middle of the semester and at the end). In the PBL assignment, students work in groups on projects given by the course teacher, based on client requirements, focused on developing a prototype of an IT service. In the first phase, students have the task to determine the characteristics of end-users and make a service proposal in the form of a wireframe (a type of problem-posing). The first version of the student solution is presented to the teacher in the middle of the semester, in the form of a business meeting, in which teachers have the role of potential clients. In the second part of the semester students work on the solution (service prototype as a type of problem-solving) according to the feedback received in the first business meeting. The PBL assignment contributes 50% to the final grade. The sample included in this study consisted of 191 students.

At the graduate level, the study included the Discrete Mathematics with Graph Theory (DMGT) course (first year). The course has a student workload of 6 ECTS credits. The course's assessment program comprises formative assessment (weekly quizzes) and summative assessment (two periodical exams, a PBL assignment). The extensive PBL assignment includes five weeks of independent work done by students in teams, and has three stages. In the first stage, students find and describe a real-world situation in which graph theory and discrete mathematics can be applied to solve a problem, and submit a project description with the characteristics of solutions (problem-posing). Feedback and grades are provided by course teachers, and these project proposals are distributed to other teams to solve. In the second stage, students work on projects, solving the problems posed by other teams (problem-solving). Finally, each team assesses the solutions to their project proposals (peer-assessment), together with the course teachers. The final grades consist of aggregated students' peer-assessment grades and teachers' grades (which have a higher weight). Additionally, students also receive a grade for the quality of work on peer-assessment, based on its consistency with the teacher's grade. The PBL assignment contributes 40% to the final grade. The sample included in this study consisted of 118 students.

For both courses, we collected assessment data from the Moodle LMS. The data were anonymized and students' identities protected.

## 3.2 Data Analysis

For each of the courses, the assessment data were randomly distributed, so that 75% of the data were used for training, and 25% were used for testing.

The data were analyzed using two supervised machine learning algorithms used for classification and regression purposes: Random Forest (RF) and Decision Trees (DTs). The analyses were performed in R, using the *tidymodels* package with *rpart* engine.

DTs are a simple method, but successful in predicting and explaining relationships (Rokach & Maimon, 2014). A DT is a mathematical graph, starting with a single node, with edges branching into next-level nodes based on the probability of a certain outcome. Each node refers to a possible decision, while edges represent their consequences. It consists of repeated tests on the input variable, with the results of each test determining the next test, until the result of the function is certain. DTs can be prone to overfitting if the number of nodes is high relative to the quantity of data. A higher number of nodes decreases the training error, but increases the generalization error. (Rokach & Maimon, 2014)

RF is an ensemble method, more complex, based on multiple DTs, combining their results for final classification. While building DTs, the RF randomly selects various subsets of attributes. RF presents "a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest" (Breiman, 2001). Due to the Law of Large Numbers, with RFs, there is no problem of overfitting. Considering all this, RFs have been found effective in making predictions. (Breiman, 2001; Kursa & Rudnicki, 2010) Another advantage of RFs is that they offer the possibility of variable importance measures (VIMs), used for identification of relevant features or selection of variables, including impurity and permutation importance (Nembrini et al., 2018). Regardless of the advantages, the random nature of RFs means they are "not always intuitive and comprehensible", as different trees can lead to inconsistency in interpretations. In this sense, the advantage of single DTs is their interpretability. (Rokach & Maimon, 2014)

We started the analysis with cross-validation to choose the optimal combination of hyperparameters. For RF, we tested 500 randomly chosen combinations of hyperparameters (*mtry, min_n, trees*), and for DTs, we tested 5000 randomly chosen combinations of hyperparameters (*tree_depth, min_n, cost_complexity*) with Latin hypercube sampling. Then, we performed RF and made single DTs on the training datasets, and tested the efficiency of each algorithm on the testing datasets.

For each of the two courses, we built an RF-based and a DT-based model for each phase of the PBL assignment. We tested the efficiency of each of the models based on several metrics, and most importantly using the area under the Receiver Operating Characteristics curve (ROC_AUC), presenting a plot of the true positive vs. the false positive rate. Although the ROC curve is normally used for assessing the performance of binary classification models, its use also extends to multi-class classification (Hand & Till, 2001; Mandrekar, 2010).

In each model, students were divided into three classes, based on their assessment results (0 - 33%, 33 - 67%, 67 - 100% of points obtained in the respective part of the PBL assignment).

We analyzed the confusion matrices for each of the models to assess the models' performance.

Furthermore, we analyzed the importance of predictors, based on three VIMs: the Gini index, permutations and the Boruta extension (Kursa & Rudnicki, 2010).

# 4 Results

## 4.1 Efficiency of the Models

The efficiency of the models was tested, most importantly, based on the ROC_AUC value. The values for both courses and both algorithms are presented in tables 1 and 2.

If we look at the training datasets, the RF results indicate a higher discriminatory power than the DT results. Namely, for both courses, the ROC_AUC values in the RF-based models are generally (around or) above 0.9, which is considered outstanding (Hosmer & Lemeshow, 2000). Looking at the DT-based models, the performance is somewhat lower, with ROC_AUC values ranging from (around) 0.7, which is considered acceptable (Hosmer & Lemeshow, 2000), to (above) 0.9. Generally, the models' performance on the training datasets indicates that patterns have been learnt well, but with the RF-based models outperforming the DT-based models.

If we consider the testing datasets, the models turned out to be less efficient, with ROC_AUC values often below 0.7 (less than acceptable). However, looking at the models for problem-solving in both DMGT and ISM, the ROC_AUC values around or above 0.8 indicate excellent efficiency (Hosmer & Lemeshow, 2000) of both RF-based and DT-based models, with the RF-based again performing slightly better. In the case of peer-assessment, this applies only to the DT-based DMGT course model.

The confusion matrices were analyzed for all the models, and here we present those related to problem-solving in DMGT (Fig. 1) and ISM (Fig. 2), RF-based models, as the most efficient ones.

**Table 1.** DMGT course: ROC_AUC values

| Random Forest | | | | | | Decision Trees | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Problem-posing | | Problem-solving | | Peer-assessment | | Problem-posing | | Problem-solving | | Peer-assessment | |
| Train | Test | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test |
| 0.850 | 0.612 | 0.955 | 0.826 | 0.996 | 0.687 | 0.689 | 0.640 | 0.883 | 0.795 | 0.939 | 0.784 |

**Table 2.** ISM course: ROC_AUC values

| Random Forest | | | | | | Decision Trees | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Problem-posing | | Problem-solving | | Peer-assessment | | Problem-posing | | Problem-solving | | Peer-assessment | |
| Train | Test | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test |
| 0.875 | 0.561 | 0.999 | 0.895 | 1.00 | 0.685 | 0.733 | 0.576 | 0.882 | 0.835 | 0.700 | 0.548 |

The DMGT course confusion matrix in Figure 1 suggests that the model is the most efficient in classifying class 1 students (lower-performing), with 9 students classified correctly, although additional 3 class 1 students were misclassified as class 3. There was some more confusion when it came to the neighboring classes 2 and 3, as 7 students were classified correctly in each of the two classes, but 4 were misclassified as belonging to the neighboring class.

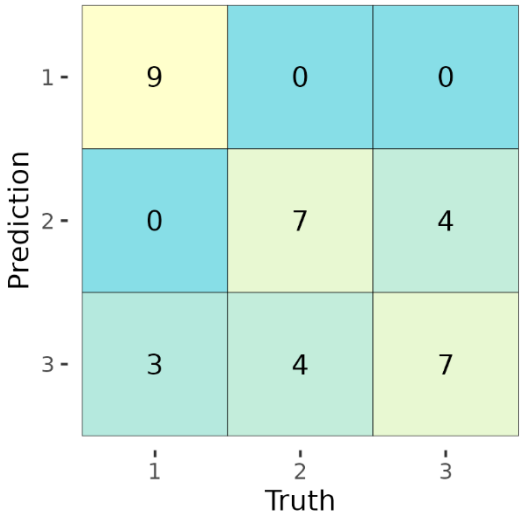



**Figure 1.** DMGT course: RF confusion matrix (problem-solving)

The ISM confusion matrix in Figure 2 shows the model performed best in predicting the (lower-performing) class 1, with 13 students classified correctly, but it also misclassified 3 class 1 students as the neighboring class 2. Similarly, it correctly classified 13 students as class 2, although it also misclassified 4 respective students as class 1. The model was the least successful with class 3, with 10 students classified correctly, but additional 5 students misclassified as either class 1 or 2.

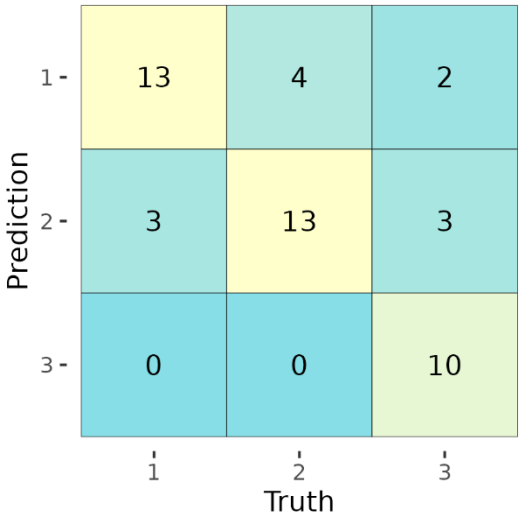



**Figure 2.** ISM course: RF confusion matrix (problem-solving)

## 4.2 Predictors

The importance of predictors was analyzed for all the models (phases of PBL). Here we present those related to *problem-solving* in DMGT and ISM, as the most efficient models.

**DMGT.** In the RF-based model, results in *problem-posing* turned out to be the most important predictor of the results in *problem-solving*, according to all the three VIMs (Gini, permutations, Boruta). *Problem-posing* was followed (from the most important) by:

*exam 2, exam 1, quizzes* (Gini, permutations) or *exam 1, quizzes, exam 2* (Boruta).

The DT-based model gave a similar picture, with *problem-posing* coming out as the most important predictor, followed by *exam 1, exam 2*, and, lastly, *quizzes*.

Looking at all the results, what comes out as the best predictor is generally *problem-posing*, followed by *exams*, and, finally, *quizzes*.

The DT (Fig. 3) gives a visual representation of criteria-based decisions about student classification. For example, students with less than 4.5 points in *problem-posing* are likely to end up in class 1. If their result in *problem-posing* was still equal to or above 1 point, they can end up in either class 1 or 2 depending on their *quiz* results. Interestingly, if their *problem-posing* results were below 1 point, they could still end up in class 3, provided that their *exam 1* results were above 16.3 points. Other paths through the DT can be interpreted in a similar way. Looking at the leaves of the DT, we can interpret the structure of the three classes.

**ISM.** In the RF-based model, similarly to DMGT, *problem-posing* was found to be the most significant predictor according to the tree VIMs. The order of

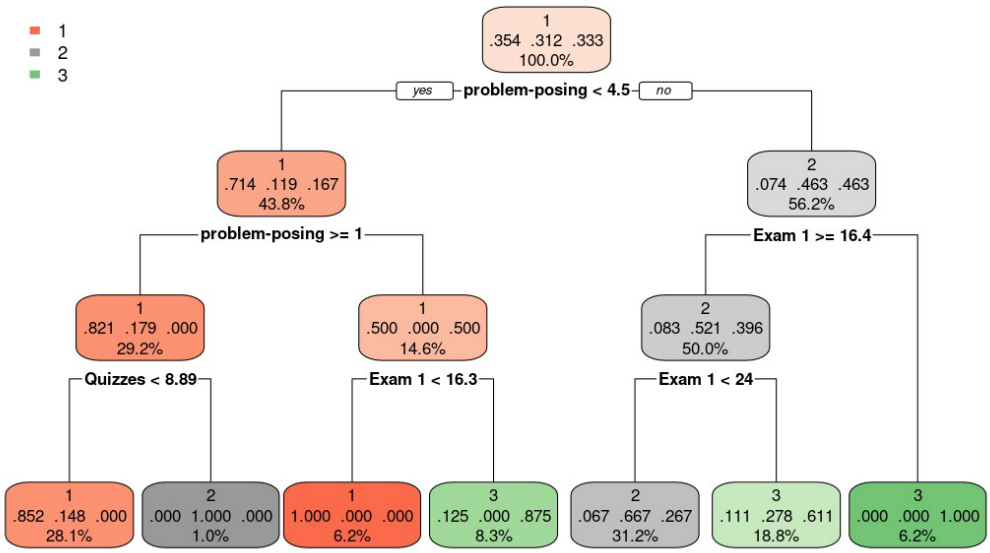



**Figure 3.** DMGT course: decision tree

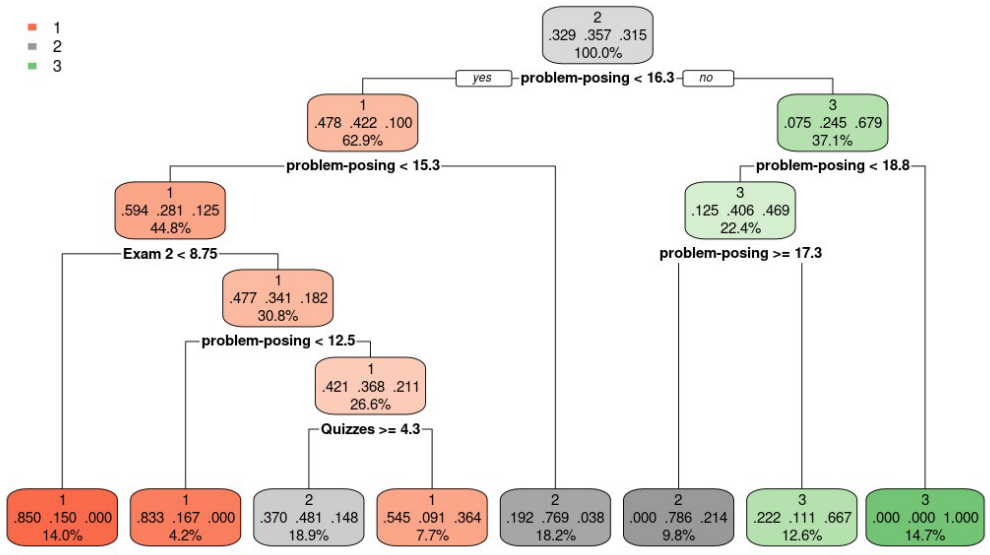



**Figure 4.** ISM course: decision tree

other predictors varied depending on the VIM, as follows (from the most important): *quizzes, exam 2, exam 1* (Gini); *exam 1, quizzes, exam 2* (permutation); *quizzes, exam 1, exam 2* (Boruta).

The DT-based model confirmed *problem-posing* as the most important predictor, followed by *exam 1, quizzes,* and *exam 2*.

Overall, the results undoubtedly point to *problem-posing* as the most important predictor, followed by *quizzes* and *exam 1*, with *exam 2* seeming the least important.

The DT (Fig. 4) shows that students with less than 16.3 points in *problem-posing* are likely to end up in class 1. If their result in *problem-posing* is above 15.3 (and still less than 16.3) they would end up in class 2, regardless of their results in other activities. If the student's result in *problem-posing* is above 16.3, they are most likely to end up in class 3. Only in the case when the result in *problem-posing* is above or equal to 17.3., students would end up in class 2.

# 5 Discussion

The comparison of prediction models based on the two algorithms, expectedly, suggests a higher efficiency of the RF compared to the DT. However, when it comes to problem-solving, both algorithms gave acceptable results. As for the other two phases of PBL, namely problem-posing and peer-assessment, the lower efficiency of the models may be related to the nature of these two PBL phases. For example, problem-posing is a type of activity in which students are generally less experienced, and which differs in terms of required competences from other parts of the assessment program. In fact, some research has pointed out that problem-posing is generally harder for students (Divjak, 2015). For example, identifying project requirements in real-life situations does not necessarily require the same competences as solving exams. Therefore, it makes sense that exams might not be good predictors of performance in problem-posing. On the other hand, this is not the case for problem-solving, as in this phase students necessarily use techniques, theory and models covered by exams. Moreover, problem-solving builds on problem-posing, so it is expected that problem-posing is the most important predictor. Although the potential of problem-posing in enhancing learning and understanding of concepts is clear (Cai, 2022), problem-posing, as a less researched part of PBL, remains a mystery. In that respect, our research is in line with previous studies (Cai & Rott, 2024). Furthermore, when it comes to peer-assessment, the better efficiency of the predictive models related to the graduate-level course may suggest that graduate-level students' performance in peer-assessment is more consistent with other parts of the assessment program. This may also be related to the fact that in the DMGT course, students peer-assessed the projects solving the problems they had posed themselves, allowing them to better understand the problem and feel more ownership.

As for other predictors, their importance may depend on the level of LOs and the learning context. For example, in the DMGT course, problem-solving tasks require students to use the algorithms and approaches which are covered by exams, but are more complex than those included in quizzes. Therefore, exams are a better predictor of PBL than quizzes. On the contrary, in the ISM course quizzes are more related to PBL, which is reflected in the importance of the predictors. Previous research (Divjak et al., 2024) showed that both formative and summative assessment are important for the prediction of final summative results. Here we showed that similar applies to PBL.

If we look at the confusion matrices, we notice that the models perform best when it comes to lower-performing (class 1) students. This has an important implication for identification and advising of students at risk from failing the course.

Moreover, the confusion among higher classes can be explained if we consider the nature of PBL, which includes teamwork. Namely, the data reflect the grades obtained by the students in teams. However, this does not reflect the exact contribution of each of the students within a team, whose performance in other parts of the assessment program might differ significantly. This, however, does not mean that teamwork is not valuable for peer-learning.

Furthermore, when it comes to explainability, DTs offer clear visual representations of how decisions are made in classification. However, DTs should be interpreted carefully, as they do not consider the timeline of activities, and do not present activities which are not nodes (vortices) of a given DT. In that sense, it is advisable to combine the DT results with other representations of students' progression, like the braided graph (Sankey diagram), which offer insights in the fluctuations of students between classes.

In terms of practical implications, the tested predictive models could be used by practitioners to inform learning design and innovate assessment.

# 6 Limitations and further research

The main limitation of this study is a relatively small sample and a limited number of courses. Data collection through several years and in several courses is needed to obtain more comprehensive and conclusive results. Furthermore, the use of other prediction methods and algorithms, like Artificial Neural Networks, Gradient Boost or Support Vector Machine, may elicit different results, which could shed a different light. Finally, including digital trace data, as well as multimodal data, in LA, is important in order to build predictive LA models which could be used as the basis for recommendation systems.

# 7 Conclusion

This paper contributes to the growing body of research on the use of learning analytics to understand and enhance problem- and project-based learning (PBL). We presented predictive analytics models based on the Random Forest (RF) and Decision Tree (DT) algorithms. We confirmed that while RF enables better efficiency, DTs are more interpretable. The models provided the best predictions for the lower performing students in PBL. Importantly, the study found that students' performance in problem-posing is an important predictor of problem-solving. Other predictors from the assessment program, including both formative and summative assessment, can be used as predictors, but fine-tuned according to the learning context. Finally, problem-posing cannot be easily predicted based on formative and summative assessment results.

# Acknowledgments

This work has been supported in part by the Croatian Science Foundation under the project IP-2022-10-2854 (*Trustworthy Learning Analytics and Artificial Intelligence for Sound Learning Design* – TRUELA).

# References

Brassler, M., & Dettmers, J. (2017). How to Enhance Interdisciplinary Competence—Interdisciplinary Problem-Based Learning versus Interdisciplinary Project-Based Learning. *Interdisciplinary Journal of Problem-Based Learning*, *11*(2). https://doi.org/10.7771/1541-5015.1686

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324

Bulut, O., Gorgun, G., Yildirim-Erbasli, S. N., Wongvorachan, T., Daniels, L. M., Gao, Y., Lai, K. W., & Shin, J. (2023). Standing on the shoulders of giants: Online formative assessments as the foundation for predictive learning analytics models. *British Journal of Educational Technology*, *54*(1), 19–39. https://doi.org/10.1111/bjet.13276

Cai, J. (2022). What Research Says About Teaching Mathematics Through Problem Posing. *Éducation et Didactique*, *16*, 31–50. https://doi.org/10.4000/educationdidactique.10642

Cai, J., & Rott, B. (2024). On understanding mathematical problem-posing processes. *ZDM – Mathematics Education*, *56*(1), 61–71. https://doi.org/10.1007/s11858-023-01536-w

Divjak, B. (2015). Assessment of complex, non-structured mathematical problems. *IMA International Conference on Barriers and Enablers to Learning Maths: Enhancing Learning and Teaching for All Learners*.

Divjak, B., Svetec, B., & Horvat, D. (2024). How can valid and reliable automatic formative assessment predict the acquisition of learning outcomes? *Journal of Computer Assisted Learning*, *January*, 1–17. https://doi.org/10.1111/jcal.12953

Dole, S., Bloom, L., & Kowalske, K. (2015). Transforming Pedagogy: Changing Perspectives from Teacher-Centered to Learner-Centered. *Interdisciplinary Journal of Problem-Based Learning*, *10*(1). https://doi.org/10.7771/1541-5015.1538

Hand, D. J., & Till, R. J. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*, *45*, 171-186Agasisti, T., Bowers, A. J. (2017). Dat.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley. https://doi.org/10.1002/0471722146

Kabathova, J., & Drlik, M. (2021). Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. *Applied Sciences*, *11*(7), 3130. https://doi.org/10.3390/app11073130

Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, *36*(11). https://doi.org/10.18637/jss.v036.i11

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Nembrini, S., König, I. R., & Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics*, *34*(21), 3711–3718. https://doi.org/10.1093/bioinformatics/bty373

Prinsloo, P. (2020). Of 'black boxes' and algorithmic decision-making in (higher) education – A commentary. *Big Data & Society*, *7*(1), 205395172093399. https://doi.org/10.1177/2053951720933994

Rokach, L., & Maimon, O. (2014). *Data Mining with Decision Trees* (Vol. 81). WORLD SCIENTIFIC. https://doi.org/10.1142/9097

Savery, J. R. (2006). Overview of Problem-based Learning: Definitions and Distinctions. *Interdisciplinary Journal of Problem-Based Learning*, *1*(1). https://doi.org/10.7771/1541-5015.1002

Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, *28*(7), 8299–8333. https://doi.org/10.1007/s10639-022-11536-0

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: a tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, *19*(1), 12. https://doi.org/10.1186/s41239-021-00313-7

Tempelaar, D., Rienties, B., & Giesbers, B. (2015). Stability and Sensitivity of Learning Analytics based Prediction Models. *Proceedings of the 7th International Conference on Computer Supported Education*, 156–166. https://doi.org/10.5220/0005497001560166

Wang, X., Sun, D., Cheng, G., & Luo, H. (2023). Key factors predicting problem-based learning in online environments: Evidence from multimodal learning analytics. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1080294

# Prioritizing Learning Outcomes in Different Learning Design Contexts

**Blaženka Divjak, Barbi Svetec**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{blazenka.divjak, barbi.svetec}@foi.unizg.hr

**Bart Rienties, Paul Astles, Catriona Matthews**
The Open University
Walton Hall, Kents Hill, Milton Keynes MK7 6AA
{bart.rienties, paul.astles, catriona.matthews}@open.ac.uk

**Alexander Tillmann, Michael Eichhorn, Julia Schmitt**
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 1, 60629 Frankfurt am Main, Germany
{tillmann, eichhorn, schmitt}@studiumdigitale.uni-frankfurt.de

***Abstract****. To ensure constructive alignment and valid assessment within courses and study programs, it is essential to identify relevant learning outcomes (LOs). However, not all LOs are equally important, so it is valuable to prioritize the LOs based on criteria. This can be done using different, more or less complex methods, from an individual estimation to structured multi-criteria group decision-making. In this paper, we classify possible approaches to the prioritization of LOs, taking into consideration the complexity of different learning contexts and the influence of a learning program on a learning path. We use a research case study to illustrate the prioritization of LOs of a professional development course designed in an international context.*

**Keywords.** learning outcomes, weights of learning outcomes, learning design, learning analytics, group decision-making

## 1 Introduction

In the last twenty years, a range of learning design (LD) approaches and conceptualizations have been introduced across the globe (Drugova et al., 2024; Macfadyen et al., 2020; Mangaroska & Giannakos, 2019). While a lot of these LD conceptualizations and tools have been proposed and tested, very few of these approaches have been specifically developed with collaboration and co-creation in mind. Working together as educators (and perhaps students) on designing appropriate learning outcomes (LOs) and LDs is an essential element of modern and inclusive LD practices (Cumbo & Selwyn, 2022; Schmitz et al., 2022; Tsui et al., 2024).

Perhaps one notable exception of an LD approach that is collaborative-by-design is the OULDI model, which is primarily based upon designing distance learning materials together in so-called module teams at the Open University (Conole, 2013; Toetenel & Rienties, 2016). Another example of a collaborate approach is FoLA (Feedback-oriented Learning Designs and Analytics) (Schmitz et al., 2022). Initially developed as a board game, the tool, which is now available digitally (https://fola.digital), enables LDs to be developed collaboratively and synchronously. A special focus is placed on raising awareness of the possibilities of learning analytics and feedback in joint lesson planning.

With the advances of technology and being able to simultaneously edit LDs synchronously and asynchronously in state-of-the-art LD tools, some LD tools allow for a blurring of the boundaries of how to design courses together (Rienties et al., 2023). For example, the Balanced Design Planning (BDP) tool is specifically designed to allow for co-authoring and co-design of learning activities, including LOs (Divjak et al., 2022; Divjak, Rienties et al., 2023). Indeed, preliminary results of the BDP tool (Divjak et al., 2024; Divjak, Rienties et al., 2023; Rienties et al., 2023) seem to suggest that educators can effectively co-design and co-write LDs together beyond an individual course unit, an organization, or even across borders. In fact, initial evidence seems to suggest that co-creation of LD can lead to engaging and innovative LDs (Divjak, Rienties, et al., 2023). But above all, it is essential that LD is sound, which means that all the elements of LD are aligned with the intended LOs. Notably, the co-creation process should consider that not all LOs have equal relative importance, which needs to be reflected in the weights of LOs (prioritization).

While there is some preliminary evidence that both technically and organizationally educators are able to effectively (co)design together in collaborative LD tools like the BDP, to the best of our knowledge there are limited studies available on how educators across different disciplines, organizations, and/or even

countries come to a consensus in how to effectively co-design. Particularly, there is scarce research on approaches to the prioritization of LOs in different learning contexts. Therefore, we conducted research to investigate how educators from different disciplines, organizations, and countries can effectively reach consensus on how to prioritize LOs when collaboratively designing learning experiences using LD tools.

# 2 Background

To the best of our knowledge, the first research on the prioritization of LOs was done by Divjak et al. (2021). The approach was based on multi-criteria decision-making (DM), specifically the Analytic Hierarchy Process (AHP) (Saaty, 2008). The AHP was used to create a hierarchy of criteria related to the characteristics of LOs, particularly the following four criteria: the importance of a topic or a context for the future profession; the required level of the LO based on a chosen taxonomy; contribution to the development of the 21st-century generic skills, and the student workload needed to fulfill the LOs. The criteria were compared in pairs to determine their importance (weights). Based on the criteria, LOs were also pairwise compared, and the total LO priorities calculated. This was done in a group DM process, with decision-makers including course professors and teaching assistants, as well as a former student.

Another study (Divjak, Svetec et al., 2023), building on the approach explained above, developed a model to ensure the validity of assessment, based on constructive alignment with weighted LOs. The model used learning analytics to compare the planned weights of LOs with actual assessment (points) and students' results, as part of the quality assurance of teaching and learning, and the continuous improvement cycle. Furthermore, it is important to note that the prioritization of LOs should not be reflected only in summative, but also in formative assessment (Divjak et al., 2024). Appropriate prioritization of LOs is the basis for the development of sound LD, which is also fundamental for meaningful learning analytics, leading back to recommendations for further improvement of LD.

# 3 Methodology

## 3.1. Study Setting

This research has been conducted as part of the international Erasmus+ project *Innovating Learning Design in Higher Education* (iLed). The purpose of the project is to support digital readiness, resilience, and the capacity of higher education (HE) through meaningful engagement with LD. The project includes four universities from four European countries: Croatia (University of Zagreb, Faculty of Organization and Informatics and School of Medicine), Germany (Goethe University Frankfurt), Finland (University of Oulu), The United Kingdom (The Open University, London). The use and development of the Balanced Design Planning (BDP) LD tool has been central to the project. The tool was originally designed based on contemporary research findings and theories such as ABC LD from University College London and the Open University LD (Divjak et al., 2022). Foundational to the BDP tool are the weight and constructive alignment of LOs, which inform the development of learning-centred courses and learning experiences.

The project has been working towards the development of a massive open online course (MOOC). The aim of the MOOC is to create a supportive learning environment for educators within HE to enhance their competences for the development of meaningful LD based on innovative pedagogies. The MOOC can also be used by educators in K-12 and lifelong learning (LLL), to support their development. The LD of the MOOC is co-created by the project partners, based on teacher requirements collected via a survey and the lessons learnt from the earlier phases of the project. This includes previous research on user experiences with the BDP concept and tool (Divjak, Rienties et al., 2023), which showed that educators struggled with the formulation and especially prioritization of LOs, as well as constructive alignment. Moreover, the LD of the MOOC also takes into account the findings of the focus group research performed in the same project, to investigate the potential of innovative pedagogies to enhance contemporary LD.

The design of the MOOC is in the focus of this paper, as case study research in which we investigate group DM in an international context.

## 3.2. Classification of Approaches to Prioritization of Learning Outcomes

In this study, our aim was to look at the prioritization of LOs from the perspective of different contexts, taking into account the following criteria:

- **Influence of the course on learners' career: high-stake vs. low-stake.** For example, a learning program towards a formal degree (high-stake) vs. non-formal training (low-stake); an obligatory course (high-stake) vs. an elective course (low-stake).
- **Complexity of LOs and learning environment: high complexity vs. low complexity.** For example, higher levels of LOs (high complexity) vs. lower levels of LOs (low complexity); LOs of a study program and an obligatory course (high

complexity) vs. LOs of a smaller elective course or non-formal training (low complexity); institutional environment (low complexity) vs. international environment (high complexity).

The contexts also imply the diversity of decision-makers (in terms of their role, experience, number and influence) and a corresponding choice of DM methods (estimation, group DM, user requirements survey, multi-criteria DM, etc.).

These criteria are not independent of each other, but are mutually interrelated. From the perspective of these criteria, we can describe four fundamental approaches to DM in the prioritization of LOs (Fig. 1):

**Complex approach.** This approach includes *formal learning* (high-stake) and *complex LOs* (LOs of diverse levels, including higher levels). In line with that, it implies a *diverse decision-maker group*, as well as it requires more *structured and reliable DM methods*. An example is the above-mentioned study by Divjak et al. (2021). In the study, prioritization was done for high-level LOs (graduate study level) in a high-stake formal learning context (obligatory course). In a nutshell, a multi-criteria group DM method (AHP) was used to determine the weights of LOs.

**Complicated approach.** This approach includes *non-formal* or *a smaller/elective unit of formal learning* (low-stake) and *complex LOs* (LOs of diverse levels). It implies a *moderately diverse decision-maker group* and a combination of *different simpler DM approaches*. This approach was studied within this case study research done in an international context and is elaborated in the following sections.

**Standard approach.** This approach includes *formal learning* (high-stake) and *not-so-complex LOs* (no or not many high-level LOs, or a narrow range of levels). It implies *less diversity in terms decision-makers*, i.e. DM can be done by a few experts or even an individual expert. DM methods can include investigation of user requirements (standard for industry and professional training in general) or academic DM in a standard environment, based on a few criteria and therefore the use of simple DM methods (e.g., SWING - von Winterfeldt & Edwards, 1986).

**Simple approach.** This approach includes *non-formal learning* or *a smaller/elective unit of formal learning* (low-stake) and *not-so-complex LOs* (a few LOs/LOs of a similar level). It implies less diversity in terms of decision-makers, i.e. it is based on informed DM by individual decision-makers (e.g., a teacher, learning designer, curriculum developer) or a discussion of a smaller group of educators with similar expertise. The core of this approach is a direct estimation of weights or assigning equal weights to all LOs.

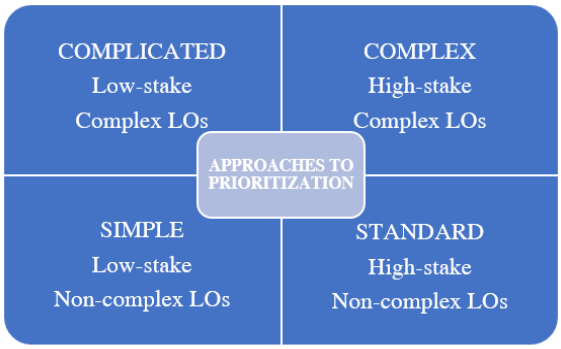



**Figure 1.** Approaches to prioritization of learning outcomes

This classification of approaches is inspired by the Cynefin framework (Snowden & Boone, 2007), which elaborated DM in four different domains: chaotic, complex, complicated and simple.

# 4 Case Study Research

We conducted case study research (Yin, 2017) which complements the existing research on LO prioritization conducted by Divjak et al. (2021) covering the *complex* approach, by providing additional insights in the *complicated* approach, as described in the Methodology section.

Our study was focused on the following research question: How to estimate the weights of LOs in a *complicated* international context?

The study was conducted as part of the above-mentioned Erasmus+ project, in relation to professional development (non-formal) activities of HE teachers. It was done in four phases, as follows.

## 4.1. First Phase: Focus Groups on Innovative Scenarios & Workshops

From July to October 2023, a total of 13 focus groups were conducted with 95 participants (63 female, 32 male) across the four institutions. The focus group participants were experienced university teachers, researchers, learning designers and student teachers. Each focus group followed the same five-step approach in the local language.

Prior to the focus group, all participants received an information package with descriptions of 11 authentic learning scenarios, a consent form and an information sheet. At least two scenarios were selected and discussed per HEI.

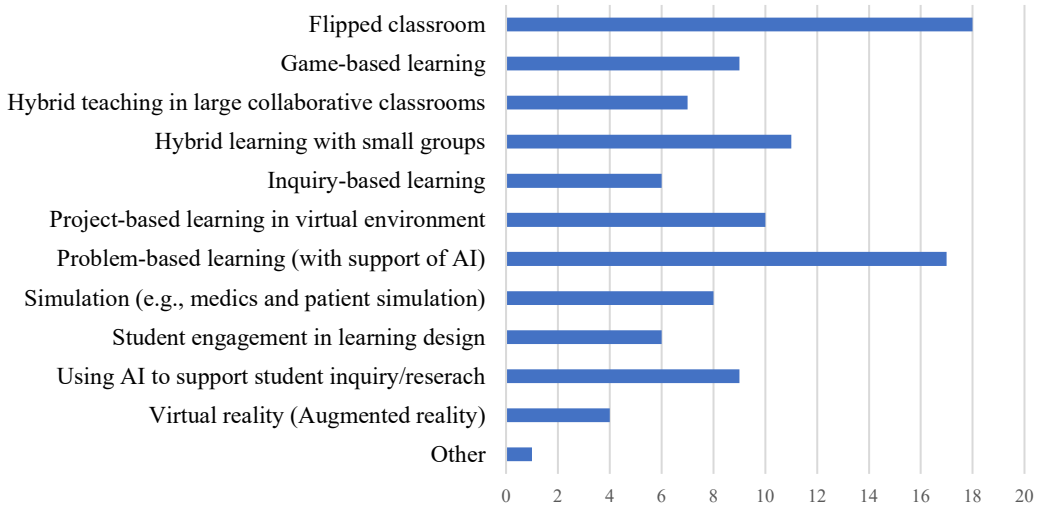



**Figure 2.** Rating of authentic learning scenarios

Based on the findings of the focus groups, a workshop was developed and implemented locally at four universities to support improvement of LD using the BDP concept and tool. The workshop LOs included understanding digital pedagogy, assessing competences, evaluating teaching and learning methods, analyzing courses, integrating digital pedagogy and using the BDP for course planning.

## 4.2. Second Phase: Survey on Teachers' Requirements & Estimation of Learning Outcome Weights

In the second phase, educators' perceptions of the importance of LOs were surveyed as part of a questionnaire covering their experiences with the workshops, as well as identifying educators' needs in terms of further training. Here we report on the results of the questionnaire in the part relevant for this study, namely, the questions referring to the sample, LO relevance, workload and innovative pedagogies.

This questionnaire was filled in by 25 participants from four HEIs (by the time of the paper submission, but data collection continues beyond this period). The majority of the participants were HE teachers (20), followed by researchers (12), learning designers (6), curriculum developers (4), educational decision-makers (3) and technical experts (1). Each participant was given the possibility to choose more than one role. The vast majority of the participants (80%) were female, and a half were mid-career experts (10 - 20 years of working experience in the respective area).

The questionnaire included a list of seven possible LOs as the basis for the project MOOC, as well as a possibility to propose additional LOs. The proposed LOs are presented in Table 1. The highest-rated LOs were those related to constructive alignment (72%), using design analytics (72%) and creating LD (64%).

The questionnaire also included a list of online or blended authentic learning scenarios, covered in the focus groups and workshops in the first phase, and subsequently in the proposed list of LOs. The participants chose the scenarios (one or more) most relevant for their (future) practice, which should be covered in the project MOOC. The highest-rated scenarios were flipped classroom and problem-based learning supported by AI, as shown in Fig. 2.

**Table 1.** Proposed learning outcomes for the MOOC

| Learning outcome (LO) | Responses |
|---|---|
| Implement innovative pedagogies in HE teaching and learning (TL) | 14 (56%) |
| Create LD of a course based on LOs | 16 (64%) |
| Use design analytics to upgrade course(s) | 18 (72%) |
| Meaningfully apply contemporary technologies and AI in TL | 8 (32%) |
| Constructively align the elements of LD (LOs, TL activities, assessment, workload) | 18 (72%) |
| Orchestrate learning design (developing content and LMS activities based on LD) | 10 (40%) |
| Self-evaluate the implementation of own course(s) | 10 (40%) |
| Other | 1 (4%) |

The participants reported on how many hours of training they would find beneficial to cover the proposed LOs, with their answers ranging from 10 to 60 hours, with the average of responses suggesting the workload of 1 ECTS. However, it should be noted that these answers were given by the participants who had already taken training in the first phase, and that for newcomers, the workload would have to be higher.

**Table 2.** Prioritization of learning outcomes

| **Learning outcomes (LO)** | LO weight: initial proposal | LO weight: group DM average | SD | Difference | **LO weight: final version** |
|---|---|---|---|---|---|
| Implement innovative pedagogies in HE teaching and learning (TL) | 18 | 18.08 | 3.26 | -0.08 | 18 |
| Create learning design of a course based on LOs and constructive alignment | 32 | 27.69 | 5.82 | 4.31 | 30 |
| Use learning analytics, especially design analytics, to upgrade course(s) | 20 | 13.85 | 3.77 | 6.15 | 18 |
| Meaningfully apply contemporary technologies and AI in TL | 10 | 15.00 | 5.22 | -5.00 | 12 |
| Orchestrate learning design (developing content and LMS activities based on learning design) | 10 | 13.46 | 6.08 | -3.46 | 12 |
| Self-evaluate the implementation of own course(s) | 10 | 11.92 | 3.34 | -1.92 | 10 |
| Total | **100** | **100** | | | **100** |

Based on the described results, the LOs were reformulated and the initial prioritization (weights) prepared by the first and second author. The reformulated LOs are presented in Table 2.

### 4.3. Third Phase: Group Decision-Making on Learning Outcome Weights

In the third phase, project experts from the four universities were presented with the proposed LOs for the MOOC and the corresponding (initial) weights, in a face-to-face project meeting. The proposed LOs for the MOOC were discussed and slightly modified. To widen the perspective and fine-tune the LO weights, group DM was done.

The group DM process included 13 project experts: HE teachers, researchers, PhD students, learning designers and technical experts. These experts were earlier engaged as trainers in the focus groups and workshops in the first phase, so they had the adequate background to estimate the importance of the proposed LOs. The experts were presented with two criteria, which they discussed and fine-tuned, and used in the process of prioritizing the agreed LOs, having in mind their insights from the focus groups and workshops:

- **Criterion 1:** Pre-knowledge and usefulness for the teaching practice
- **Criterion 2:** Workload needed to reach the desired level of LO.

### 4.4. Fourth Phase: Aggregation of Decision-Makers' Prioritization of Learning Outcomes

In the fourth phase, the group DM experts took a look at the results and discussed the differences and similarities, as presented in Table 2: between the initial proposal of the weights and the new average (Difference), and between individual DM experts' opinions (SD). In the discussion, the LO weights were further adjusted taking into account both the initial prioritization and the group DM. The final values are presented in Table 2.

## 5 Discussion

Prioritization of LOs is an important element of LD, with implications on the continuous improvement cycle of courses and study programs, as demonstrated in the study by Divjak, Svetec et al. (2023). Although valuable, it has been confirmed as challenging for educators (Divjak, Rienties et al., 2023).

To shed some more light and provide some practical guidance for educators, we classified four possible approaches to the prioritization of LOs, taking into consideration different learning contexts, with varying influence on learners' careers, and demanding to a varying extent. The classification relies on the principle of efficient use of resources (human, organizational), and therefore links DM methods and decision-maker involvement with learning contexts.

In this case study, we focused on describing a multi-stage *complicated* approach applied to the prioritization of LOs in the context of the design of a MOOC within an Erasmus+ project. The presented approach builds on the existing concepts of learner-oriented teaching and constructive alignment, integrating the perspectives of both learners and teachers. The results show that structured multi-stage prioritization of LOs, involving different stakeholder perspectives in an international context, is feasible in non-formal learning settings.

The results show that experts' opinions on LO weights were to a great extent aligned with the opinions of learners, however, with exceptions. Importantly, the use of learning analytics was considered more demanding by learners than by experts. Conversely, meaningfully applying contemporary technologies and AI in teaching and learning was considered more important by experts than by learners. (Table 2) Both might be related to the previous exposure, existing competences, and awareness of the capabilities, benefits and risks of the said tools (Rienties et al., 2023). Interestingly, both points of disagreement are related more to technology and algorithms, than pedagogy and LD.

In comparison with the *complex* approach used by Divjak et al. (2021), the *complicated* approach used within this case study is more suitable for this context, with a large group of international experts. Multi-criteria DM methods, like the AHP used in Divjak et al. (2021), are more complex and time-consuming, requiring several hours per expert to prepare and complete. Moreover, the AHP becomes almost impossible to manage with a large DM group or when there are many alternatives and criteria to consider.

We argue that the choice of appropriate methods for the prioritization of LOs strongly depends on the respective teaching-learning context. While more structured approaches like the AHP seem to be appropriate for formal, high-stake, *complex* settings, more intuitive methods may suffice for lower-stake situations. The presented approach demonstrates a promising way for holistic prioritization of LOs, involving various stakeholders. However, it is important to use the approach appropriate with respect to the context, not to burden practitioners in the preparation of LD.

The *complicated* approach used in this case study consisted of two stages: investigation of user requirements and simple group DM by experts. Each of these stages could be used as a stand-alone method within the *standard* approach, which is less demanding and time-consuming for practitioners. Additionally, another DM method suitable for this context can be the SWING method (von Winterfeldt & Edwards, 1986), in which LOs are first ranked, and the most important one given 100 points. After that, other LOs are compared with the most important one and assigned points.

In summary, the study shows that for high-stakes LOs, such as those of formal degree programs or certifications with major career implications, more time and human resources should be allocated to thoroughly prioritize the LOs. The prioritization process requires structured DM methods like AHP that allow for comprehensive analysis from multiple expert perspectives. In contrast, for a low-stake learning environment, such as non-formal professional development courses or personal development, less time is needed and more intuitive DM approaches to prioritization can be utilized. Larger studies on the prioritization of LOs in different contexts would be desirable as a future research perspective, to test the usability of the approaches and add to the toolbox of DM methods in different contexts. As regards LD concepts and tools (e.g. the BDP), guidelines and functionalities to support the prioritization of LOs would be valuable. Finally, more elaborated theoretical links of the presented approaches with the Cynefin framework can be investigated.

# 6 Conclusion

Learning outcomes (LOs) are a widely accepted instrument of ensuring learner-centred teaching and learning. Sound learning design relies on constructive alignment, which calls for meaningful prioritization of LOs, reflecting their relative importance. The prioritization of LOs can be done using four different approaches, depending on the complexity of LOs and the importance of a learning program for a learner's career. These approaches include *complex, complicated, standard* and *simple*. For each of these approaches, the prioritization of LOs can be modeled using different DM methods, from multi-criteria DM (e.g., the AHP) to a simple estimation of LO weights. In any approach, it is valuable to consider different perspectives, including teachers and learners. Finally, the complexity of the approach used to prioritize LOs should be aligned with the given teaching and learning contex.

# Acknowledgments

This study has been conducted within the Erasmus+ project *Innovating Learning Design in Higher Education* (iLed) (2022-1-HR01-KA220-HED-000085818).

# References


Conole, G. (2013). *Designing for Learning in an Open World*. Springer New York. https://doi.org/10.1007/978-1-4419-8517-0

Cumbo, B., & Selwyn, N. (2022). Using participatory design approaches in educational research. *International Journal of Research & Method in Education*, *45*(1), 60–72. https://doi.org/10.1080/1743727X.2021.1902981

Divjak, B., Grabar, D., Svetec, B., & Vondra, P. (2022). Balanced Learning Design Planning. *Journal of Information and Organizational Sciences*, *46*(2), 361–375. https://doi.org/10.31341/jios.46.2.6

Divjak, B., Kadoic, N., & Zugec, B. (2021). The Use of Decision-Making Methods to Ensure Assessment Validity. *2021 IEEE Technology & Engineering Management Conference - Europe (TEMSCON-EUR)*, 1–6. https://doi.org/10.1109/TEMSCON-EUR52034.2021.9488580

Divjak, B., Rienties, B., Bađari, J., Grabar, D., Horvat, D., & Vondra, P. (2023). Enhancing Learning Design through User Experience Research: Insights from a Survey in Four European Countries. *Proceedings of the Central European Conference on Information and Intelligent Systems*, 213–221.

Divjak, B., Svetec, B., & Horvat, D. (2024). How can valid and reliable automatic formative assessment predict the acquisition of learning outcomes? *Journal of Computer Assisted Learning*, *January*, 1–17. https://doi.org/10.1111/jcal.12953

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. https://doi.org/10.1111/bjet.13290

Drugova, E., Zhuravleva, I., Zakharova, U., & Latipov, A. (2024). Learning analytics driven improvements in learning design in higher education: A systematic literature review. *Journal of Computer Assisted Learning*, *40*(2), 510–524. https://doi.org/10.1111/jcal.12894

Macfadyen, L. P., Lockyer, L., & Rienties, B. (2020). Learning Design and Learning Analytics: Snapshot 2020. *Journal of Learning Analytics*, *7*(3), 6–12. https://doi.org/10.18608/jla.2020.73.2

Mangaroska, K., & Giannakos, M. (2019). Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516–534. https://doi.org/10.1109/TLT.2018.2868673

Rienties, B., Balaban, I., Divjak, B., Grabar, D., Svetec, B., & Vondra, P. (2023). *Applying and Translating Learning Design and Analytics Approaches Across Borders* (pp. 35–53). https://doi.org/10.1007/978-3-031-27646-0_3

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83. https://doi.org/10.1504/IJSSCI.2008.017590

Schmitz, M., Scheffel, M., Bemelmans, R., & Drachsler, H. (2022). FoLA2 — A Method for Co-creating Learning Analytics–Supported Learning Design. *Journal of Learning Analytics*, *9*(2), 265–281. https://doi.org/10.18608/jla.2022.7643

Snowden, D., Boone, M. (2007). A Leader's Framework for Decision Making. *Harvard Business Review*, *85*(11), 68–76.

Toetenel, L., & Rienties, B. (2016). Analysing 157 learning designs using learning analytic approaches as a means to evaluate the impact of pedagogical decision making. *British Journal of Educational Technology*, *47*(5), 981–992. https://doi.org/10.1111/bjet.12423

Tsui, E., Dragicevic, N., Fan, I., & Cheng, M. (2024). Co-creating curriculum with students, teachers, and practitioners in a technology-enhanced environment. *Educational Technology Research and Development*, *72*(2), 869–893. https://doi.org/10.1007/s11423-023-10301-5

von Winterfeldt, D., & Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge University Press.

Yin, R. K. (2017). *Case Study Research and Applications: Design and Methods, 6th Ed.* Sage Publications Inc.

# Predictive Factors for Completion or Dropping Out of Professional Studies

**Bojan Radišić**
Faculty of Tourism and Rural Development in Požega, Josip Juraj Strossmayer University of Osijek
Vukovarska 17, 34000 Požega, Croatia
bradisic@ftrr.hr

**Ivan Dunđer, Sanja Seljan**
Faculty of Humanities and Social Sciences, University of Zagreb, Department of Information and Communication Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
{idundjer, sseljan}@ffzg.unizg.hr

***Abstract****. The aim of this paper is to identify the predictive factors that influence the completion or dropping out of studies. It compares the data obtained through the ISVU system between two groups of students of three-year professional studies in economics. The first group consists of students who studied at a polytechnic, whereas the second group studied at the university. First, both groups are analysed for intergroup similarity. Data comparison is done through the prediction of successful completion of studies using machine learning. Finally, an analysis identifies the differences in the predictors that influence the completion or dropping out of studies.*

***Keywords.*** *Machine learning, multiple linear regression, success prediction, student dropout*

## 1 Introduction

In the Republic of Croatia, a binary higher education system was established based on the Bologna Declaration and the Act on Higher Education and Scientific Activity (Official Gazette 119/2022).

Namely, polytechnics carry out higher education and professional activities. A polytechnic can also perform scientific or artistic activities and other activities in accordance with the law and statute. A university, however, carries out higher education, scientific or artistic, and professional activities in at least two scientific areas, or a scientific and artistic area, and in at least three scientific fields. A university can, in addition, perform scientific or artistic activities and other activities in accordance with the law and statute. Polytechnics enable enrolment in professional studies that prepare students for the labour market. On the other hand, most of the universities are focused on science, but also carry out professional studies. In the academic year 2023/2024 in the Republic of Croatia, students enrolled in 116 professional studies at universities according to data on enrolment quotas in the summer enrolment period (Agency for Science and Higher Education, AZVO).

According to this data from the Agency for Science and Higher Education for the academic year 2023/2024, the number of enrolled students at the universities is satisfactory after the summer enrolment deadline. However, almost half of the enrolment places at the polytechnics in Croatia remained empty after the end of the summer enrolment period.

According to the quality standard for evaluation in the process of reaccreditation of higher educational institutions (Official Gazette 151/2022) of the Agency for Science and Higher Education, each higher educational institution must monitor and analyse the progress of students in their studies and ensure the continuity and completion of studies according to the AZVO's standard 3.3. Quality standards for evaluation in the process of reaccreditation of higher education institutions, 2023).

The conditions for enrolment and advancement of students, as well as recognition and certification, are clear, publicly announced, and consistently applied. Based on ISVU data on students from previous years, it is possible to improve the success of current and future students, which will be discussed in this article. Timely completion of studies (graduation) is important for students, their families, employers, and state authorities within the education and employment ecosystem.

The goal of this paper is to predict the success of finishing studies and to identify the factors that influence the completion or dropping out of studies.

The paper is structured in the following way. In the introductory part, the specifics of the higher education system in Croatia are presented. The second section presents related work in the field, with a special focus on suitable machine learning algorithms. The third section, i.e. the research methodology and results section, details the research workflow and findings regarding the similarities and differences between polytechnic and university students. The fourth section lists the limitations and ideas for future research, while the fifth section concludes with recommendations for university and polytechnic management.

# 2 Related Work

Data mining is becoming increasingly important in the education sector because it improves the education system and facilitates student development through predictive analysis. Educational institutions collect and store vast amounts of data, including student enrolment, class attendance, and exam results.

Different approaches to data mining are currently being used in the field of education with a particular focus on handling missing data. One research examined how four missing-value data sets affect machine learning algorithm performance (Radišić et al., 2023b). Arithmetic means, median values and geometric means were used to fill in the missing values, which in all cases had a positive effect.

However, researchers are increasingly favouring artificial intelligence and machine learning techniques for extracting information from educational datasets due to their superior reliability compared to other methods (Gupta et al., 2020).

Predicting student grades and grades based on their past academic data is a widely used and beneficial application in educational data mining. It serves as a significant source of information that can be utilised in various ways to improve the country's education quality (Yousafzai et al., 2020).

For instance, labelled student history data, consisting of 29 optimal attributes, was used to train a decision tree classifier and a regression model. The genetic algorithm-based decision tree classifier and the regression model showed remarkable results. The classification accuracy for grade prediction was 96.64%, while the regression-based grade prediction system had a root mean square error (RMSE) of 5.34.

Machine learning algorithms such as random forest, support vector machine, k-nearest neighbour and logistic regression (Yağcı, 2022) can be used to validly predict student success. The authors concluded that midterm exam results are essential to predict student success accurately.

Campanilla (2024) tried to predict study completion using a naive Bayesian algorithm. The findings show that of the 272 enrolled students, 200 (73.6%) had a higher GPA and were therefore more likely to complete their college education, while 72 (26.4%) were less likely to do so. The study suggested redesigning policies to support university initiatives aimed at reducing student dropout rates.

A multiple linear regression model was used to predict student success based on data obtained from surveys conducted among engineering students (Kumar et al., 2020). The results showed that four independent variables were significant for modelling students' academic success: gender, relationship with parents, previous success in studies (performance), and grades from the first semester.

A probabilistic neural network (PNN) achieved the best accuracy for predicting factors influencing student performance and success from student data extracted from a distance learning system compared to other classification algorithms (Nazif et al., 2020). It was tested with feature selection using neighbourhood component analysis for classification, and here feature selection is done using statistical measures, measures from information theory and interclass distance.

One study investigated the performance of undergraduate computer science students using data from 2001 to 2015 to predict student graduation using linear regression (LR), random forest regressor (RFR), multilayer perceptron regressor (MLPR), logistic regression (LoR), random forest (RF) and multilayer perceptron (MLP). Random forest had the best performance and linear regression performance is comparable (Alamgir et al., 2024).

A study of 13,696 Latin American university students from the first semester of 2008 to the second semester of 2020 analysed various influences on student dropout (Gutierrez-Pachas et al., 2023). A correlation analysis showed that the number of completed semesters has in all cases a robust negative correlation with dropping out.

A recent study using machine learning identified 887 students who were prone to dropping out and performing poorly in their studies. This was done using data on students' academic history, pre-enrolment status, socioeconomic status and psychological characteristics in order to determine their progress (Jayaprakash et al., 2020). The study showed that gender, family size, parental status, education and work can negatively affect student success. The authors used naïve Bayes, bagging classifier, logit boosting, random forest and enhanced iterative random forest.

Multiple linear regression is often used to predict student performance and select predictors that affect the chosen dependent variable (e.g. study completion). A recent study analysed the motivational factors and predicted the key elements that influence the duration of studies (Radišić et al., 2023a). Factors such as age, high school performance etc., best predict study duration in a multiple linear regression model.

There are several methods for selecting predictors. In an analysis of the Student Performance Dataset extracted from the University of California, Irvine (El Aissaoui et al., 2020), researchers used several methods for selecting variables, and the multivariate adaptive regression splines (MARS) method proved to be the best.

# 3 Research Methodology and Results

This paper deals with determining the similarities and predictive factors of success of students of professional studies at the university and polytechnic level. This is done by comparing two datasets in order to predict the success of polytechnic and university students in

completing their studies and to identify the factors that influence the completion of their studies.

Data on students were extracted from a student information system called ISVU – Information System of Higher Education Institutions (*Informacijski sustav visokih učilišta*). The data were obtained from two educational institutions that conduct three-year professional studies in the field of economics and covered the period from 2016 to 2018. This time frame was chosen because the length of the entire study, which lasted until the beginning of 2024, could be observed for the students.

The first dataset consisted of data on polytechnic students, whereas the second one covered data on university students, both from eastern Croatia.

In the first part of this research, descriptive statistics is used to show the differences and similarities between the students of both educational institutions. Here the students were analysed who have either completed their studies, dropped out of their studies or are still studying.

In the second part of this research, three machine learning algorithms for predicting the successful completion of studies are compared: random forest (RF), naïve Bayes (NB) and probabilistic neural network (PNN). This part of the research provides predictions of the success of currently enrolled students, and could enable action to be taken accordingly, in order to increase completion of studies.

In the third part of this research, the factors that influence the completion or dropping out of studies are identified, and the predictors that mostly influence the completion of studies using the multiple linear regression method. This part is essential as it can guide an institution on what kind of students to enrol, considering the influence of individual factors that affect the success of completing studies. The following hypotheses are set in this paper:

H1: Students of professional studies at the university have a shorter duration of study compared to students at the polytechnic.

H2: It is possible to identify factors that influence the completion of professional studies.

One research question was raised:

RQ1: Is it possible to predict a student's success, i.e. completion of studies?

H1 will be tested by descriptive statistics, whereas H2 will be analysed using multiple linear regression and the stepwise predictor selection method. RQ1 will be answered with the help of machine learning models.

## 3.1. Descriptive Data Analysis

In the first part of this research a descriptive analysis is conducted using on students who either completed their studies, gave up their studies, or who are still studying. According to the data extracted from the institutions' ISVU, in the academic years 2016/2017, 2017/2018 and 2018/2019, 207 students were enrolled at the polytechnic and 204 students at the university, which results in almost equal amounts of data in the two datasets. All data were taken at the end of the academic year 2022/2023.

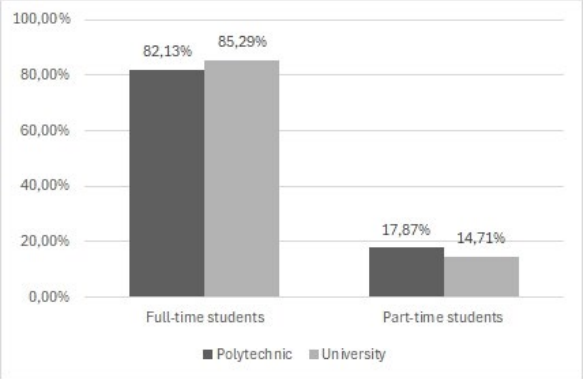



**Figure 1.** Type of study (enrolment type)

Fig. 1 shows the number of full-time and part-time students. The number is almost equal in both observed groups. Slightly more full-time students were enrolled at the university (85.29%), whereas slightly more part-time students were enrolled at the polytechnic (17.87%). These two groups are almost equal in terms of the number of women who studied: 62.80% at the polytechnic and 63.24% at the university. An almost equal number of students who graduated from one of the secondary economic schools enrolled at the polytechnic (54.59%) and the university (59.31%).

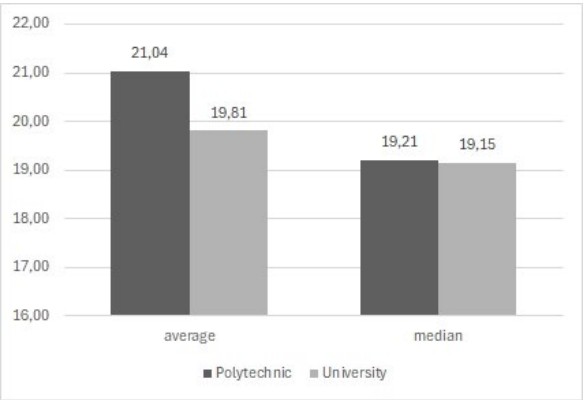



**Figure 2.** Age of students at study enrolment

Fig. 2 shows the average and median age at study enrolment. It can be seen that the median age is almost the same at the polytechnic (19.21) and university (19.15). There is, however, a difference in the average age at enrolment – namely, the average age at the polytechnic is slightly higher (21.04) than at university (19.81). Students older than 28 did not choose to study at the university but enrolled at the polytechnic. The oldest student who decided to enrol at the university was younger than 28 years old. For comparison, 17 students over the age of 28 were enrolled at the polytechnic, which resulted in a slightly higher average age of students who enrolled at the polytechnic.

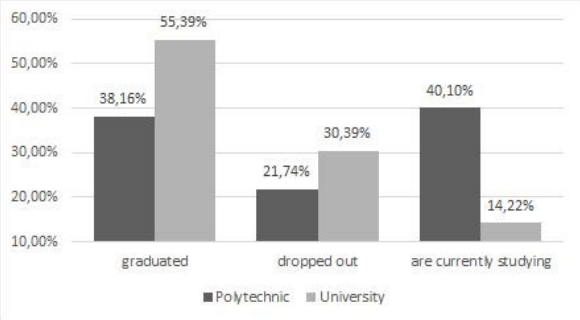



**Figure 3.** State of completion of professional studies

Fig. 3 shows the state of completion of professional studies among university and polytechnic students. Out of the total number of students enrolled, more than half (55.39%) completed their studies at the university, while slightly more than a third (38.16%) completed their studies at the polytechnic. Nevertheless, dropping out of studies is more pronounced at the university (30.39%), where almost every third of students drops out, while at the polytechnic almost every fifth student drops out (21.47%).

The most significant difference between these two groups is the number of students who are still studying. It is 40.10% at the polytechnic and 14.22% at the university, which affects the duration of studies and thus increases the projected (expected) duration of studies, which does not favour the polytechnic in this research.

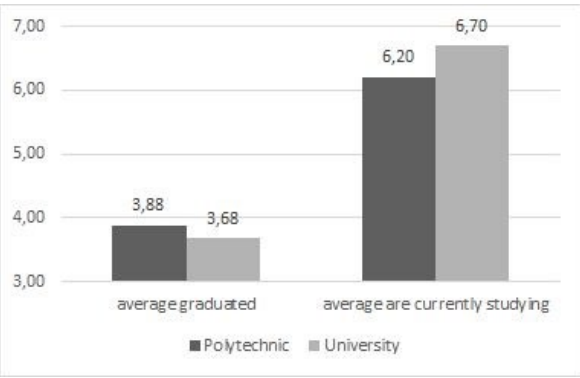



**Figure 4.** Average duration of study

Fig. 4 shows the average duration of studies, but with a note that less than 60% of the polytechnic students have completed their studies, while the rest have either dropped out or are still studying.

That part is more representative for the university and refers to slightly less than 86% of the total sample who completed their studies. The average duration of studies of students who have completed their studies (which officially lasts 3 years) is 3.88 years for the polytechnic and 3.86 years for the university. Students who are still studying study for an average of 6.20 years at the polytechnic and 6.70 years at the university, which is almost the same average.

If the weighted average of those who have graduated and those who are still studying is calculated, assuming that all those who are currently studying complete their studies, the average duration of studies is 4.81 years at the polytechnic and 4.1 years at the university. According to the weighted average, university students complete their studies in 15% less time than polytechnic students. This confirms the hypothesis (H1) that students of professional studies at the university have a shorter duration of study compared to students at the polytechnic.

Fig. 5 shows the most significant difference between students who choose the university over the polytechnic. The overall results of the state exam (state graduation/matriculation) show that the students who chose to study at the university achieved an average of 341.16 points, which is 52.34% more than the average of 223.95 points, achieved by students who chose to continue their education at the polytechnic. It should be noted here that 44 students entered the polytechnic without taking the state exam.

Furthermore, there are many similarities in the characteristics of enrolled students at the university and the polytechnic. An equal number of full-time students are enrolled; they come from the same county where they study and have approximately the same average age, and there are equally many who have graduated from one of the secondary schools of economics.

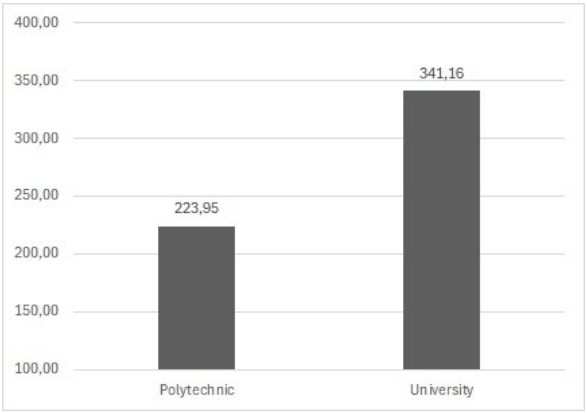



**Figure 5.** Average state exam score

On the other hand, the most significant difference is in the results they achieved on the state exam, where the students who enrolled their studies at the university achieved significantly better results. Students who had better success at the state exam enrolled at the university. This resulted in a shorter duration of studies and faster entry into the labour market or continuation of education.

## 3.2. Machine Learning

In the second part of this research, three machine learning algorithms were compared to determine their applicability in predicting student success in studies. The output variable was the completion of studies divided into three categories:

1) graduated students (**GS**),
2) students who dropped out from studies (**DO**),

3) students who are currently studying (**AS**).

Nine input variables (i.e. data features) were selected and are listed and described in Table 1.

**Table 1.** Data features

| Features | Description |
|---|---|
| Enrolment type | Part-time students, Full-time students |
| Age | Student's age at the time of study enrolment |
| Gender | Male, Female |
| County | County of residence |
| County Yes/No | Residence and place of study is in the same county or not |
| Secondary school | Type of finished secondary school |
| Economics Yes/No | Completed secondary school of economics or not |
| Score | Secondary school final score |
| Exam score | State exam score |

Machine learning (ML) methods can be used to predict study completion or dropout. Namely, the most commonly used methods include random forest, naïve Bayes, and neural networks, which are explored in this study.

### Random Forest (RF)

The random forest (RF) method is usually employed for classification and regression tasks (Kovač et al., 2022). It is a straightforward and reliable classifier that exhibits a number of beneficial attributes. During the training phase, the algorithm generates several decision trees using bootstrapping (sampling with replacement) and random feature selection. When dealing with a new data point, each decision tree makes a prediction. In the case of classification, the final output is determined by taking the mode (the most frequently predicted class). In regression, the final result is the average of all the predictions. It provides advantages in automating the handling of missing data values and efficient processing of large datasets. Certain disadvantages are associated with increased computing requirements and the use of resources needed to achieve optimal results.

### Naïve Bayes (NB)

Naïve Bayes (NB) is a classification algorithm based on Bayes' theorem (Berrar, 2019):

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

P(A) is the probability of A occurring, P(B) is the probability of B occurring, P(A|B) is the probability of A given B and P(B|A) is the probability of B given A. The naïve Bayes algorithm is very versatile and practical, making it generally preferred for categorising large datasets. In certain cases, the feature independence assumption occasionally needs to be corrected, leading to suboptimal classification results. The classifier is probabilistic, which means that it makes predictions based on the probability of an object occurring.

### Probabilistic Neural Network (PNN)

Probabilistic neural networks (PNNs) are artificial neural networks created using the Parzen approach (Mohebali et al., 2020). The purpose of PNNs is to develop a collection of estimators for the probability density function. These estimators are designed to approximate the Bayesian optimal minimisation of "expected risk", popularly known as "Bayesian strategies". Probabilistic neural networks (PNNs) have demonstrated significant potential in solving complex scientific and technical problems.

### Results

Both student datasets (containing data on polytechnic and university students, respectively) were compared using all three machine learning algorithms to compare their accuracy. 80% of the data was randomly left for training, and the remaining 20% was used for testing the algorithms.

Results for polytechnic students are presented in Fig. 6, which shows that the algorithms themselves did not achieve high accuracy. The NB and RF algorithm were almost equal, with an accuracy of 45.20% for NB and 42.86% for RF, while the PNN showed a lower accuracy of 33.33%.

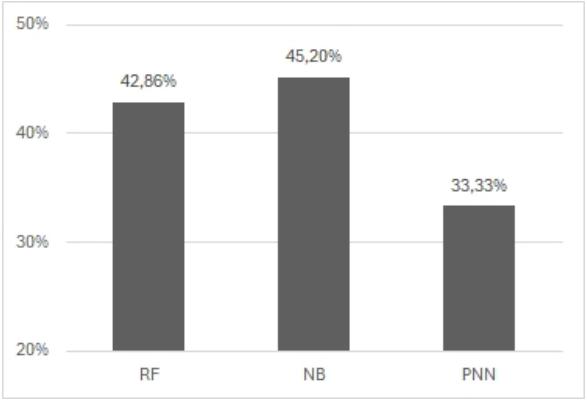



**Figure 6.** Accuracy for polytechnic students

Although the results for polytechnic students are quite low, a closer examination of the confusion matrices in Table 2 shows that NB exhibited a high accuracy of 83.30% in recognising students who are still studying. RF (66.70%) and PNN (60%) also recognised students who are still studying better than graduates or dropouts, but it was lower than for NB.

All three machine learning algorithms were least able to recognise students who dropped out of their studies; RF recognised every fifth, NB every fourth, and PNN recognised none.

Fig. 7 shows the accuracy results of the machine learning algorithms for university students. The most accurate was NB, with 63.42%, followed by less accurate RF algorithm (58.54%). Both had better accuracy compared to the same algorithms when used on the polytechnic students' dataset. As in the former case, the PNN algorithm was the least accurate, with 48.78%.

**Table 2.** Confusion matrix for polytechnic students

| | True data/Predicted data | GS | DO | AS |
|---|---|---|---|---|
| *RF* | **GS** | 5 (26.31%) | 4 | 5 |
| | **DO** | 4 | 0 (00.00%) | 2 |
| | **AS** | 10 | 4 | 9 (56.25%) |
| *NB* | **True data/Predicted data** | **GS** | **DO** | **AS** |
| | **GS** | 12 (46.15%) | 0 | 1 |
| | **DO** | 4 | 2 (20.00%) | 0 |
| | **AS** | 10 | 8 | 5 (83.33%) |
| *PNN* | **True data/Predicted data** | **GS** | **DO** | **AS** |
| | **GS** | 5 (26.31%) | 4 | 4 |
| | **DO** | 4 | 0 (00.00%) | 2 |
| | **AS** | 10 | 4 | 9 (60.00%) |

Remarks: graduated students (**GS**), students who dropped out from studies (**DO**), students who are currently studying (**AS**)

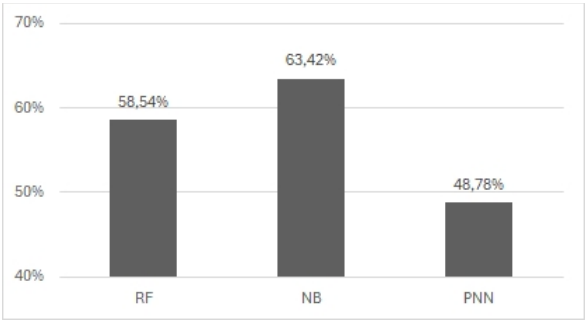



**Figure 7.** Accuracy for university students

Table 3 shows the confusion matrix of all machine learning algorithms for university students. The NB algorithm showed the highest accuracy of 76.90% in recognising university students who graduated. The RF and PNN algorithms showed almost similar accuracy results in recognising students who graduated, with 65.70% for RF and 62.50% for PNN. All three machine learning algorithms were least able to recognise students who are still studying; RF and PNN did not recognise any of the cases, whereas NB recognised 50%.

The results of this subsection can help educational institutions to identify the possible duration of studies of currently enrolled students. The use of the NB algorithm is particularly applicable and useful for the polytechnic, which achieved an accuracy of (83.30%) in recognising students who are still studying at the polytechnic.

The algorithms have the lowest accuracy when identifying students who dropped out of their studies, the NB algorithm has an accuracy of 20%, while the RF and PNN failed to recognize a single student of the polytechnic.

**Table 3.** Confusion matrix for university students

| | True data/Predicted data | GS | DO | AS |
|---|---|---|---|---|
| *RF* | **GS** | 22 (65.70%) | 1 | 1 |
| | **DO** | 9 | 1 (20.00%) | 0 |
| | **AS** | 3 | 3 | 0 (00.00%) |
| *NB* | **True data/Predicted data** | **GS** | **DO** | **AS** |
| | **GS** | 20 (76.90%) | 4 | 1 |
| | **DO** | 5 | 5 (38.50%) | 0 |
| | **AS** | 1 | 4 | 1 (50.00%) |
| *PNN* | **True data/Predicted data** | **GS** | **DO** | **AS** |
| | **GS** | 15 (62,50%) | 9 | 1 |
| | **DO** | 5 | 5 (31.25%) | 0 |
| | **AS** | 4 | 2 | 0 (00.00%) |

Remarks: graduated students (**GS**), students who dropped out from studies (**DO**), students who are currently studying (**AS**)

By timely detection of such students, it is possible to take some measures that could help students shorten their duration of study and thereby significantly shorten the overall average duration of studies at their institution. In this case, the university could use the NB algorithm (accuracy shown to be 76.90%) to identify students expected to graduate successfully. This can also identify those students who will either drop out or continue studying and focus on them to help them finish their studies.

The algorithms have the lowest accuracy in the case of identifying dropped out students, the NB algorithm has an accuracy of 38.50%, while the RF (20.00%) and PNN (31.25%) have lower accuracy for university students. This answers RQ1, since it possible to predict to a certain extent the students' success in completing their studies (e.g. 76.90% for university students with the help of NB).

## 3.3. Multiple Linear Regression

In this subsection, the authors determine what affects the completion of studies. Ten variables were selected from the ISVU database from each of the two selected institutions: completed studies (Yes/No), enrolment type (part-time or full-time students), age (student's age at the time of study enrolment), gender (male/female), county of residence, residence and place of study is in the same county (Yes/No), type of finished secondary school, completed secondary school of economics (Yes/No), secondary school final score and state exam score.

This third part of the research investigated the correlation between the completion of studies (i.e. the dependent variable), and the other mentioned

variables, also known as predictors or predictor variables.

**Multiple linear regression (MLR)** is a statistical technique for creating a model that represents the relationship between two or more independent and dependent variables by fitting a linear equation to observed data. The underlying theoretical assumption of MLR is that any unit change in the independent variable results in a consistent change in the dependent variable (El Aissaoui et al., 2020). The multiple linear regression equation is shown below:

$$Y = k_0 + k_1x_1 + k_2x_2 + \cdots + k_nx_n$$

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).

Table 4 shows a comparison of two multiple linear regression models for the polytechnic students using the stepwise method. The R-squared value of the second model is 0.099, indicating that the model explains 9.99% of the variation in graduation (i.e. the completion of studies). In the second model, the predictors are enrolment type (part-time or full-time students) and gender, and form the most efficient multiple linear regression model between the dependent variable (completion of studies) and these two mentioned predictors.

**Table 4.** Comparison of models for polytechnic students

| Model | R | R² | R² Change | F Change | df1 | df2 | p |
|---|---|---|---|---|---|---|---|
| 1 | 0.256 | 0.066 | 0.066 | 14.424 | 1 | 205 | < .001 |
| 2 | 0.315 | 0.099 | 0.034 | 7.616 | 1 | 204 | 0.006 |

1. Predictors: (Constant), Enrolment type
2. Predictors: (Constant), Enrolment type, Gender

In Table 5, the column "p" shows that both predictors make a statistically significant unique contribution (<0.05) to the final value (case of polytechnic students).

**Table 5.** Correlation between the completion of studies and other predictors for polytechnic students

| Coefficients | | | | | | |
|---|---|---|---|---|---|---|
| Model | | Unstandardized | Standard Error | Standardized | t | p |
| 2 | (Constant) | 1.368 | 0.325 | | 4.211 | < .001 |
| | Enrolment type | 0.595 | 0.152 | 0.260 | 3.913 | < .001 |
| | Gender | -0.333 | 0.121 | -0.183 | -2.760 | 0.006 |

The highest absolute value is scored by enrolment type (part-time or full-time students) which has the greatest influence of 0.260. This means that if the predictor reflecting enrolment type (part-time or full-time students) increases by one standard deviation, and other predictors remain constant, then completion of studies will increase by 0.260 of standard deviations. Also, the influence of gender has negative sign and amounts to -0.183. This means that if the predictor that reflects gender increases by one standard deviation, and other predictors remain constant, then completion of studies will decrease by 0.183 standard deviations. The model implies that full-time female students will eventually graduate.

Table 6 shows a comparison of three multiple linear regression models for the university students using the stepwise method. The R-squared value of the third model is 0.230 indicating that the model explains 23% of the variation in graduation (i.e. the completion of studies). In the third model, the predictors are gender, enrolment type (part-time or full-time students) and type of finished secondary school, and form the most efficient multiple linear regression model between the dependent variable (completion of studies) and these three predictors.

**Table 6.** Comparison of models for university students

| Model | R | R² | R² Change | F Change | df1 | df2 | p |
|---|---|---|---|---|---|---|---|
| 1 | 0.335 | 0.112 | 0.112 | 25.468 | 1 | 202 | < .001 |
| 2 | 0.433 | 0.188 | 0.076 | 18.797 | 1 | 201 | < .001 |
| 3 | 0.480 | 0.230 | 0.042 | 11.045 | 1 | 200 | 0.001 |

1. Predictors: (Constant), Gender
2. Predictors: (Constant), Gender, Enrolment type
3. Predictors: (Constant), Gender, Enrolment type, Type of finished secondary school

In Table 7 the column "p" shows that all three predictors make a statistically significant unique contribution (<0.05) to the final value (case of university students).

**Table 7.** Correlation between the completion of studies and other predictors for university students

| Coefficients | | | | | | |
|---|---|---|---|---|---|---|
| Model | | Unstandardized | Standard Error | Standardized | t | p |
| 3 | (Constant) | 2.105 | 0.288 | | 7.305 | < .001 |
| | Gender | -0.434 | 0.094 | -0.288 | -4.68 | < .001 |
| | Enrolment type | 0.617 | 0.129 | 0.301 | 4.792 | < .001 |
| | Type of finished secondary school | -0.076 | 0.023 | -0.208 | -3.323 | 0.001 |

The highest absolute value is scored by enrolment type (part-time or full-time students) which has the highest influence of 0.301. This means that if the predictor that reflects enrolment type (part-time or full-time students) increases by one standard deviation, and other predictors remain constant, then completion of studies will increase by 0.301 of standard deviations. Also, the influence of gender and type of finished secondary school have negative signs; -0.288 for gender and -0.208 for the type of finished secondary

school. This means that if the predictor that reflects gender increases by one standard deviation, and other predictors remain constant, then completion of studies will decrease by 0.288 of standard deviations and decrease by 0.208 for the type of finished secondary school. The model implies that full-time female students that finished a secondary school of economics usually graduate.

In this part, it can be concluded that gender and enrolment type (part-time or full-time students) affect the success of completing studies both at the polytechnic and the university. However, for the institutions themselves, these predictors have very little significance for attracting future students, except for increasing the enrolment quotas for full-time students.

Nevertheless, the financial benefits made by the educational institutions would make it unfavourable to reduce the number of part-time students. Another important aspect is that studying part-time suits older students because, in addition to work, it enables them to continue their education and training to acquire new knowledge that they will use in their workplace.

More significant for the university is that more successful students finished a secondary school of economics, i.e. they will be more successful and eventually graduate. The university can therefore adjust and strengthen its marketing activities there to attract as many students as possible.

Factors that significantly influence the completion of studies were determined by multiple linear regression. For polytechnic students, these are type of enrolment and gender, and for university students: gender, type of enrolment, and type of completed high school, thereby confirming H2.

# 4 Limitations and Future Research

This research has limitations regarding the number of input variables as a more significant number would allow more accurate predictions with the help of machine learning algorithms. Data on student success by semester and academic year could be included as well, and it would be beneficial to include other elements of monitoring students during their studies (attendance of classes, homework outcomes, colloquia result, academic engagement metrics etc.). For future research, it would be advisable to include a longer time period (i.e. longitudinal data and across more cohorts) in order to improve the robustness of the results with regard to temporal variations.

In addition, it would be wise to include data from more institutions from the same field for future research to obtain more representative data. It would also be valuable to compare and include educational institutions from different scientific fields. This would help to better generalize the findings and validate the models across diverse educational settings.

# 5 Conclusion and recommendations

Every higher educational institution in Croatia should monitor and analyse the progress of students, and ensure the continuity and completion of their studies. This paper showed that it is possible to predict students who will be successful in their studies and who will be able to graduate. The duration of the full study, which continued until the start of 2024, could be seen for the students, therefore that is why this time frame was selected.

The second part of this research showed that institutions can detect students who are currently enrolled and actively studying. As shown in this paper, students who are still studying and who, on average, will potentially study longer can be recognized at the polytechnic.

The polytechnic can enable such students to study for a shorter period time, possibly through various forms of assistance in learning and teaching. The university, on the other hand, can identify students who will complete their studies. However, those remaining students can be provided with different forms of assistance in learning and teaching in order to help them finish their studies more successfully.

This research has shown that students of professional studies at the university level have a shorter study time (4.1) than students at the polytechnic (4.81).

The research identified the following predictors as factors that influence the completion of studies: enrolment type (part-time or full-time students) and gender at the polytechnic; and at the university: gender, enrolment type and type of finished secondary school.

Using the naïve Bayes machine learning model, it is possible to identify polytechnic students who are still studying with an accuracy of 83.30% and with an accuracy of 76.90% the university students expected to graduate successfully.

Based on the data that higher educational institutions collect and store about their students, they can direct their marketing activities and the promotion of their institution to a precisely defined population of potential future students. Activities should be focused on students that finished a secondary school in the field of economics to attract students who are successful in their studies.

Finally, the recommendations to the polytechnic go in the direction of using machine learning to detect students who are expected to study longer and to take measures related to improving the academic efficiency of the identified students. Using machine learning, the university can identify students who will successfully complete their studies and, just like polytechnics, can take measures to improve the academic performance of other students. Furthermore, the university should work on attracting a larger number of students who completed their high school education in the field of economics.

# References


Alamgir, Z., Akram, H., Karim, S., Wali, A. (2024). Enhancing Student Performance Prediction via Educational Data Mining on Academic Data. *Informatics in Education*, 23(1), 1–24, doi:10.15388/infedu.2024.04

Agency for Science and Higher Education (AZVO) (Agencija za znanost i visoko obrazovanje). Retrieved from https://www.studij.hr/broj-prijava

Berrar, D. (2019). Bayes' Theorem and Naive Bayes Classifier. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 403–412. doi:10.1016/B978-0-12-809633-8.20473-1

Campanilla, B. S. (2024). Forecasting Degree Completion: A Naïve Bayes Predictive Model for Students' Success. In *Proceedings of the Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1–4). doi: 10.1109/ICAECT60202.2024.10469398

El Aissaoui, O., El Madani, Y. E., Oughdir, L., Dakkak, A. & El Allioui, Y. (2020). A Multiple Linear Regression-Based Approach to Predict Student Performance. In M. Ezziyyani (Ed.), Proceedings of the *Advanced Intelligent Systems for Sustainable Development Conference (AI2SD'2019)* (pp. 9–23), Advances in Intelligent Systems and Computing, vol. 1102. Springer, Cham. doi:10.1007/978-3-030-36653-7_2

Gupta, S. B, Yadav, R. K. & Gupta, S. (2020). Analysis of Popular Techniques Used in Educational Data Mining. *International Journal Of Next-Generation Computing*, 11(2), 137–162. doi:10.47164/ijngc.v11i2.178

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

Jayaprakash, S., Krishnan, S., & Jaiganesh, V. (2020). Predicting Students Academic Performance using an Improved Random Forest Classifier. In *Proceedings of the 2020 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 238–243). doi:10.1109/ESCI48226.2020.9167547

Kovač, A., Dunđer, I. & Seljan, S. (2022). An overview of machine learning algorithms for detecting phishing attacks on electronic messaging services. In *Proceedings of the 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 954–961). doi:10.23919/MIPRO55190.2022.9803517

Kumar, A., Eldhose, K. K., Sridharan R. & Panicker, V. V. (2020). Students' Academic Performance Prediction using Regression: A Case Study. In *Proceedings of the 2020 International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1–6). doi:10.1109/ICSCAN49426.2020.9262346

Mohebali, B., Tahmassebi, A., Meyer-Baese, A. & Gandomi, A. H. (2020). Probabilistic neural networks: a brief overview of theory, implementation, and application. In: P. Samui, S. Chakraborty, D. T. Bui, R. C. Deo (Eds.), *Handbook of Probabilistic Models*, Elsevier, Amsterdam (pp. 347–367). doi:10.1016/b978-0-12-816514-0.00014-x

Nazif, A. M., Hesham Sedky, A. A. & Badawy, O. M. (2020). MOOC's Student Results Classification by Comparing PNN and other Classifiers with Features Selection. In *Proceedings of the 21st International Arab Conference on Information Technology (ACIT)* (pp. 1–9). doi:10.1109/ACIT50332.2020.9300123

Official Gazette 119/2022 (NN 119/2022, Zakon o visokom obrazovanju i znanstvenoj djelatnosti). Retrieved from https://narodne-novine.nn.hr/clanci/sluzbeni/full/2022_10_119_1834.html

Official Gazette 151/2022 (NN 119/2022, Zakon o osiguravanju kvalitete u visokom obrazovanju i znanosti). Retrieved from https://narodne-novine.nn.hr/clanci/sluzbeni/2022_12_151_2330.html.

Quality standards for evaluation in the process of reaccreditation of higher education institutions (Standardi kvalitete za vrednovanje u postupku reakreditacije visokih učilišta). Retrieved from https://www.azvo.hr/wp-azvo-files/uploads/radne-grupe/12/20/Standardi-kvalitete-za-vrednovanje-u-postupku-reakreditacije-visokih-ucilista.pdf

Radišić, B., Dunđer, I. & Seljan, S. (2023a). Data Analysis of the Motivation and Factors for a Shorter Duration of Study. In *European Conference on Information and Intelligent Systems (CECIIS)* (pp. 179–187). Varaždin: University of Zagreb, Faculty of Organization and Informatics Varaždin.

Radišić, B., Seljan, S. & Dunđer, I. (2023b). Impact of missing values on the performance of machine learning algorithms. In *CEUR Workshop Proceedings: Recent Trends and Applications in Computer Science and Information Technology (RTA-CSIT 2023)* (pp. 54-62). Tirana: University

of Tirana, Faculty of Natural Sciences, Department of Informatics.

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environment*, 9(11). doi:10.1186/s40561-022-00192-z

Yousafzai, B. K., Hayat, M., & Afzal, S. (2020). Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. Education and Information Technologies, 25, 4677–4697. doi:10.1007/s10639-020-10189-1

# Student perspectives on using Artificial Intelligence tools for solving programming assignments

# Perspektiva studenata o korištenju alata umjetne inteligencije za rješavanje programskih zadataka

**Matija Novak, Marija Maček**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, Varaždin, Croatia
{matija.novak, marija.macek}@foi.unizg.hr

**Abstract.** *Students use various methods for solving their programming assignments. Notably, with the widespread adoption of artificial intelligence (AI) tools, it is unsurprising that students have incorporated them into their repertoire. The main question, however, revolves around the permissibility of such utilization. In this study a survey was conducted among university students in Croatia, aiming to clarify their perspectives on the deployment of AI tools. The primary objective is to discern the extent to which students employ these tools in solving programming assignments and investigate the ways in which the students use them.*

**Keywords.** programming, assignments, high education, Artificial Intelligence

**Sažetak.** *Učenici koriste različite metode za rješavanje programskih zadataka. S obzirom na široko prihvaćanje alata umjetne inteligencije (AI), nije iznenađujuće da su ih studenti uključili u svoj repertoar. Međutim, glavno pitanje se vrti oko dopuštenosti takvog korištenja. U ovoj studiji provedena je anketa među studentima u Hrvatskoj s ciljem razjašnjavanja njihovih perspektiva o implementaciji alata umjetne inteligencije. Primarni cilj je razlučiti u kojoj mjeri studenti koriste te alate u rješavanju programskih zadataka i istražiti načine na koje ih studenti koriste.*

**Ključne riječi.** programiranje, zadaci, visoko obrazovanje, umjetna inteligencija

## 1 Introduction

Programming is often perceived as a challenging subject to learn, as highlighted in prior research (Bain and Barnes, 2014; Guzdial, 2010). Notably, in the study by (Bain and Barnes, 2014) they "identified a major issue with problem-solving strategies: 50% had no strategy beyond 'Google it'". Studies of this nature indicate that certain students attempt to address programming assignments by seeking solutions on the Web.

Instances of students attempting to address programming assignments by seeking solutions on the Web raise ethical concerns, particularly when academic policies explicitly prohibit such practices. This behavior can result in a similarity between solutions, potentially leading to accusations of plagiarism, as outlined in the definitions provided by (Cosma and Joy, 2008) and (Novak, 2020).

Adding to the complexity, with the widespread adoption of artificial intelligence (AI) tools (e.g. ChatGPT),

## 1 Uvod

Programiranje se često doživljava kao izazovan predmet za učenje, kako je naglašeno u prethodnim istraživanjima (Bain and Barnes, 2014; Guzdial, 2010). Posebno, u istraživanju koje su proveli (Bain and Barnes, 2014), "identificiran je glavni problem sa strategijama rješavanja problema: 50% nije imalo strategiju osim 'Googlanja'". Studije ovakve vrste ukazuju na to da određeni studenti pokušavaju riješiti zadatke programiranja tražeći rješenja na Webu.

Pokušaji studenata da riješe zadatke iz programiranja tražeći rješenja na Webu izazivaju etičke zabrinutosti, posebno kada akademske politike izričito to zabranjuju. Ovakvo ponašanje može rezultirati sličnostima između rješenja, što potencijalno dovodi do optužbi za plagijat, kako je opisano u definicijama koje su navedene u radovima (Cosma and Joy, 2008) i (Novak, 2020).

Dodatno, sa sveprisutnom upotrebom alata umjetne inteligencije (eng. Artificial Intelligence - AI) (npr.

raise concerns regarding academic integrity and the preservation of originality in programming assignments.

Distinguishing between the application of AI tools and the conventional "Google it" approach lies in the capacity of such tools to generate comprehensive solutions merely by copying the problem. In contrast, conventional search engines may prove inadequate for more complex assignments, often failing to yield complete solutions. The distinctive aspect of AI tools is their ability to provide students with entire solutions, removing the need for integration or modification.

Moreover, a challenge arises when traditional similarity detection tools relying on string matching prove ineffective in scenarios where an AI tool generates varying solutions with each instance. While the importance of detection cannot be understated, first an emphasis should be placed on prevention. In order to prevent it is necessary to understand the student perspective on this matter.

In this study, the primary objective is to examine students' perspectives on AI tools. To explore this topic, a survey was administered to students in Croatia. The survey questions were predominantly open-ended to avoid influencing their opinions. Key inquiries focused on whether and how students utilized AI tools for resolving programming assignments. Furthermore, participants were asked to articulate their stance on whether these tools should be prohibited or permitted, along with providing explanations for their views. Additionally, students were prompted to explain the advantages and disadvantages they perceive in the utilization of such tools.

The rest of the paper is structured as follows. In Section 2, related work is presented. Section 3 presents the methodology and the main questions of the study. The results are presented in section 4. Section 5 contains the discussion and gives ideas for future work, and Section 6 concludes.

## 2 Related work

Various studies have been conducted to examine the usage of AI tools in the context of programming assignments. Some research focuses on detection, as demonstrated in (Ellis et al., 2024), where the authors investigated the likelihood of teachers identifying code generated by AI tools in the Python programming language. They asked teachers and used an AI detector tool (i.e. OpenAI Text Classifier) in their experiment to identify AI generated assignments. The findings from (Ellis et al., 2024) suggest "that although AI-generated results may not always be identifiable, they do not currently yield results approaching those of diligent students." Notably, using the AI detector poses a significant challenge. "Although the AI detector was largely able to identify AI-generated results accurately, it also had as many or more false positives as the human gra-

ChatGPT), javljaju se zabrinutosti u vezi s akademskom integritetom i očuvanjem originalnosti u programskim zadacima.

Razlika između primjene AI alata i konvencionalnog pristupa "Googlanju" leži u sposobnosti takvih alata da generiraju sveobuhvatna rješenja jednostavno kopiranjem problema. Za razliku od toga, konvencionalni pretraživači često se mogu pokazati neprikladnima za složenije zadatke, često ne pružajući potpuna rješenja. Distinktivna karakteristika AI alata je njihova sposobnost pružanja studentima cijelih rješenja, eliminirajući potrebu za integracijom ili modificiranjem.

Javlja se problem kod tradicionalnih alata za detekciju sličnosti, koji se oslanjaju na usporedbu nizova, da postaju neučinkoviti u scenarijima gdje AI alat generira različita rješenja svaki put. Iako važnost detekcije ne smije biti podcijenjena, prvo bismo trebali staviti naglasak na prevenciju. Kako bismo spriječili, nužno je razumjeti perspektivu studenata o ovom pitanju.

U ovom istraživanju, osnovni cilj je istražiti perspektive studenata o AI alatima. Kako bismo istražili ovu temu, provedena je anketa među studentima u Hrvatskoj. Anketna pitanja uglavnom su bila otvorenog tipa kako bi se izbjeglo utjecanje na mišljenje studenata. Ključna pitanja usmjerena su na to jesu li i kako studenti koristili AI alate za rješavanje zadataka iz programiranja. Nadalje, sudionike se potiče da izraze svoj stav o tome trebaju li ti alati biti zabranjeni ili dopušteni te da pri tome daju objašnjenja svojih stajališta. Osim toga, studenti su potaknuti da objasne prednosti i nedostatke koje percipiraju u korištenju takvih alata.

Ostatak rada je organiziran na sljedeći način. Poglavlje 2 sadrži povezane radove. Metodologija i glavna pitanja istraživanja predstavljena su u poglavlju 3. Rezultati su prikazani u poglavlju 4. Poglavlje 5 sadrži raspravu i daje ideje za buduće radove, dok poglavlje 6 zaključuje istraživanje.

## 2 Povezani radovi

Različita istraživanja provedena su kako bi se proučila upotreba AI alata u kontekstu zadataka iz programiranja. Dio istraživanja fokusira se na detekciju, kako je prikazano u (Ellis et al., 2024), gdje su autori istraživali vjerojatnost da nastavnici prepoznaju kod generiran od strane AI alata u programskom jeziku Python. Anketirali su nastavnike i koristili alat za detekciju umjetne inteligencije (npr. OpenAI Text Classifier) u svom eksperimentu kako bi identificirali zadatke generirane od strane AI alata. Zaključci iz (Ellis et al., 2024) sugeriraju "da iako AI-generirani rezultati možda nisu uvijek prepoznatljivi, trenutno ne daju rezultate koji se približavaju rezultatima marljivih studenata." Važno je napomenuti da korištenje AI detektora predstavlja značajan izazov. "Iako je AI detektor uglavnom bio sposoban točno prepoznati AI-generirane re-

ders." The authors also noted the importance of instructors being aware of how students use AI tools, regardless of whether their use is permitted in a class. In (Ellis et al., 2024), the authors also offer valuable advice to instructors on how to prevent or make it more difficult to use AI tools, such as using images instead of text in assignments to eliminate the possibility of copy-pasting.

To understand student perspectives, surveys have been undertaken among students like (Saari et al., 2022; Farhi et al., 2023) . In 2022, (Saari et al., 2022) conducted a survey involving 200 students at Tampere University, focusing on their perspective towards AI tools in the context of Java programming studies. In 2023, a research study conducted by (Farhi et al., 2023) gathered data from 388 students enrolled in two prominent universities in the United Arab Emirates (UAE), shedding light on their perceptions of ChatGPT. The survey from (Saari et al., 2022) posed two primary questions: firstly, whether the students employed AI tools in their studies, and secondly, in what situations they utilized these tools, or if they did not use them, the reasons behind such decisions. The results revealed that 45% of students used the tools infrequently, while 20% utilized them on a weekly basis, leaving 35% who did not use the AI tools. In response to the second question, approximately 50% used AI tools for learning, around 40% employed them as a search engine, roughly 38% for coding purposes, and about 25% for debugging. Among students who did not use AI tools, responses indicated that around 30% perceived no need for them, 30% expressed a desire to learn without their help, and 40% had an unspecified reason. Based on the results, the authors concluded that students do indeed utilize AI tools and highlight the need to educate students on the responsible use of such tools. Similarly in (Farhi et al., 2023) the findings unveiled that students consider ChatGPT to be a revolutionary technology, showcasing a dual impact characterized by positive advancements and simultaneous concerns regarding its effects on educational integrity. In both studies the limitations are the geographical focus but with the potential for further exploration of students' concerns in future research.

Similarly, in 2023, (Smolansky et al., 2023) conducted a survey focusing on the perspectives of both students and teachers, with a unique focus on assessment practices. The study questioned 389 students and 36 educators across two universities in Australia and the US. Among various queries, participants were asked to rate the impact of AI tools on different types of assessments, including computer code. Notably, both students and teachers concurred that computer code is significantly affected, with educators assigning higher impact ratings than students. Despite an initial anticipation of differences between educators and students, the study revealed a surprisingly high level of agreement. In contrast to (Saari et al., 2022), the authors

zultate, također je imao jednako ili više lažno pozitivnih rezultata u usporedbi s ljudskim ocjenjivačima." Autori također ističu važnost da nastavnici trebaju biti svjesni kako studenti koriste AI alate, bez obzira je li njihova upotreba dopuštena u nastavi. U (Ellis et al., 2024), autori nude korisne savjete nastavnicima o tome kako spriječiti ili otežati korištenje AI alata, poput korištenja slika umjesto teksta u zadacima kako bi se eliminirala mogućnost kopiranja.

Za dobivanje uvida u studentske perspektive, provedena su istraživanja među studentima poput (Saari et al., 2022; Farhi et al., 2023). Godine 2022., (Saari et al., 2022) proveli su anketu između 200 studenata na Sveučilištu Tampere, fokusirajući se na njihovu perspektivu prema AI alatima u kontekstu učenja programiranja u programskom jeziku Java. U 2023. godini, istraživanje provedeno od strane (Farhi et al., 2023) prikupilo je podatke od 388 studenata upisanih na dva istaknuta sveučilišta u Ujedinjenim Arapskim Emiratima (UAE), otkrivajući njihova stajališta o ChatGPT-u. Anketa iz (Saari et al., 2022) postavila je dva osnovna pitanja: prvo, jesu li studenti koristili AI alate, a drugo, u kojim situacijama su koristili te alate, ili ako ih nisu koristili, razlozi takvih odluka. Rezultati su pokazali da je 45% studenata rijetko koristilo alate, dok je 20% koristilo tjedno, ostavljajući 35% onih koji nisu koristili AI alate. Na drugo pitanje, otprilike 50% studenata koristilo je AI alate za učenje, oko 40% ih je koristilo kao tražilicu, otprilike 38% u svrhu programiranja, a oko 25% za otklanjanje grešaka. Među studentima koji nisu koristili AI alate, odgovori su pokazali da otprilike 30% ne vidi potrebu za njima, 30% želi učiti bez njihove pomoći, a 40% nisu naveli razloge. Na temelju rezultata, autori su zaključili da studenti doista koriste AI alate te su istaknuli potrebu za obrazovanjem studenata o odgovornom korištenju takvih alata. Slično u (Farhi et al., 2023), rezultati su ukazali na to da studenti smatraju da je ChatGPT revolucionarna tehnologija, s dvostrukim utjecajem koji se očituje pozitivnim napretkom, ali i istovremenom zabrinutosti u pogledu njezinog utjecaja na obrazovni integritet. U oba istraživanja, ograničenje predstavlja geografsko područje, ali postoji potencijal za daljnje proučavanje studentskih perspektiva u budućim istraživanjima.

Slično, u 2023. godini, (Smolansky et al., 2023) proveli su istraživanje usmjereno na perspektive studenata i nastavnika, s posebnim fokusom na postupak ocjenjivanja. Istraživanje je obuhvatilo 389 studenata i 36 nastavnika na dva sveučilišta u Australiji i SAD-u. Između ostalog, sudionici su trebali ocijeniti utjecaj AI alata na različite vrste ocjenjivanja, uključujući programski kod. Važno je istaknuti da su studenti i nastavnici suglasni da je programski kod značajno zahvaćen, pri čemu su nastavnici dali veći utjecaj nego studenti. Unatoč prvotnim očekivanjima da će postojati značajne razlike između nastavnika i studenata, istraživanje je pokazalo iznenađujuće visoku razinu suglasnosti. Za razliku od (Saari et al., 2022), autori tvrde da umjesto

argue that rather than enhancing students' capabilities in AI, it is more crucial to assist them in navigating the complex interplay between technology, cognition, social interaction, and values.

In 2023, (Yilmaz and Karaoglan Yilmaz, 2023) took a different approach and asked students to use ChatGPT to solve programming assignments and to provide feedback. As the main benefits they listed: "providing fast and mostly correct answers to questions, improving thinking skills, facilitating debugging, and increasing self-confidence.". Regarding the downsides, they listed: "getting students used to laziness, being unable to answer some questions or giving incomplete/incorrect answers, and causing professional anxiety in students." Based on their results, they recommend integrating the use of AI tools into courses. However, they also emphasize the importance of educating students to check and verify the results they receive. In addition, they state that "It becomes essential to equip students with the knowledge and critical thinking abilities necessary to navigate the ethical considerations and potential biases associated with AI-generated outputs."

# 3 Methodology

In this study, a survey was distributed to students in Croatia across five courses at three faculties. In an effort to avoid influencing students' responses, the decision was made to predominantly utilize open-ended questions within the survey. Subsequently, these open-ended responses required coding for result analysis, enabling the summarization of data and facilitating quantitative analysis. While the results leaned toward a qualitative analysis due to their nature, for specific responses, quantitative data was provided.

The primary objective of the survey was to ascertain whether and how students utilize AI tools for programming assignments. Secondary, the aim was to understand their perspectives on the advantages and disadvantages of using such tools. To gather insights, the survey included the following questions:

1. Have you tried AI tools such as ChatGPT, Microsoft Copilot, etc., so far? Yes/No question.
2. Please list the tools you have used. Open-end question
3. Have you used AI tools to solve assignments? Yes/No question
4. How did you use tools to solve assignments? Please describe as thoroughly as you can. Open-end question
5. Have you used AI tools to solve programming assignments? Yes/No question.

da se educiraju studenti u području korištenja AI alata, važnije je da im se pomogne u snalaženju u kompleksnoj interakciji između tehnologije, spoznaja, društvene interakcije i vrijednosti.

U 2023. godini, (Yilmaz and Karaoglan Yilmaz, 2023) su pristupili drugačije i zamolili studente da koriste ChatGPT za rješavanje zadataka iz programiranja, te pruže povratne informacije. Kao glavne prednosti naveli su: "brzo pružanje uglavnom ispravnih odgovora na pitanja, poboljšanje vještina razmišljanja, olakšavanje otklanjanja pogrešaka i povećanje samopouzdanja." Što se tiče nedostataka, istaknuli su: "navikavanje studenata na lijenost, nemogućnost odgovaranja na neka pitanja ili davanje nepotpunih/netočnih odgovora te uzrokovanje profesionalne anksioznosti kod studenata." Na temelju njihovih rezultata, preporučuju integriranje upotrebe AI alata u nastavne programe. Ističu i važnost obrazovanja studenata kako bi provjeravali i potvrđivali rezultate koje dobivaju. Također, navode da "postaje ključno unaprijediti znanje studenata i sposobnost kritičkog razmišljanja koji su potrebni za snalaženje u etičkim pitanjima i potencijalnim pristranostima povezanima s AI generiranim rezultatima."

# 3 Metodologija

U ovom istraživanju, anketa je provedena među studentima u Hrvatskoj, na pet kolegija, na tri fakulteta. S ciljem izbjegavanja utjecaja na odgovore studenata, donesena je odluka da će se u anketi uglavnom koristiti otvorena pitanja. Stoga su ovi otvoreni odgovori zahtijevali šifriranje radi analize rezultata, omogućavajući sumiranje podataka i time kvantitativnu analizu. Iako su rezultati, zbog svoje prirode, više naginjali kvalitativnoj analizi, za određene odgovore dani su i neki kvantitativni podaci.

Osnovni cilj ankete bio je utvrditi koriste li studenti i na koji način alate AI za zadatke iz programiranja. Sekundarno, cilj je bio razumjeti njihova razmišljanja o prednostima i nedostacima korištenja takvih alata. Kako bismo prikupili podatke, anketa je uključivala sljedeća pitanja:

1. Jeste li do sada isprobali alate AI poput ChatGPT-a, Microsoft Copilot-a itd.? Da/Ne pitanje.
2. Molimo navedite alate koje ste koristili. Otvoreno pitanje.
3. Jeste li koristili AI alate za rješavanje zadataka? Da/Ne pitanje.
4. Na koji način ste koristili alate za rješavanje zadataka? Molimo vas da opišete što je detaljnije moguće. Otvoreno pitanje.
5. Jeste li koristili AI alate za rješavanje zadataka iz programiranja? Da/Ne pitanje.
6. Na koji način ste koristili AI alate za rješavanje

6. How did you use tools to solve programming assignments? Please describe in as much detail as possible. Open-end question
7. Do you believe that AI tools for programming assignments should be: entirely allowed, partially allowed and partially forbidden, entirely forbidden.? One choice possible.
8. Do you believe that AI tools should be allowed or prohibited for solving programming assignments? Explain your reasoning and attempt to clarify for which types of assignments.
9. What is your overall impression of these tools? What advantages and disadvantages do you see in the context of studying?
10. If there's anything else you would like to add, feel free to write it here.

Except for the main questions there were four other questions:

- What faculty are you attending.
- Select the level of study. (graduate, undergraduate)
- What study program you are enrolled in.
- What is your current study year? (1,2,3,4,5)

The first major limitation of this survey is the uneven distribution of students between graduate (15 responses) and undergraduate (74 responses), and secondly, the fact that the majority of students come from one university (74 responses). This circumstance will limit the possibility of drawing conclusions regarding differences between groups. Nonetheless, interesting and relevant results were obtained. When appropriate, the survey described the number of students in the graduate and undergraduate categories.

# 4 Results

A total of 89 responses were received. The majority of responses (74) were from undergraduate students, while 15 responses came from graduate students. Among the undergraduate students, all were in their 3rd year of the study program. For graduate students, there were 8 in their 4th year and 7 in their 5th year.

In response to the initial question about whether students used AI tools, only 5 indicated "no", whereby 1 (i.e. 6.6%) was graduate student and 4 (i.e. 5.4%) were undergraduate students. Given that the primary focus of this research revolves around the usage of AI tools, the responses from these 5 students are excluded from the analysis but will be examined separately at the end. This leaves 84 valid responses.

Concerning the second question, the majority mentioned using ChatGPT (62%), as illustrated in Fig. 1.

zadataka iz programiranja? Molimo vas da opišete što je detaljnije moguće. Otvoreno pitanje.

7. Smatrate li da bi AI alati trebali biti: u potpunosti dopušteni, djelomično dopušteni i djelomično zabranjeni, potpuno zabranjeni? Jedan odabir je moguć.
8. Smatrate li da bi AI alati trebali biti dopušteni ili zabranjeni za rješavanje zadataka iz programiranja? Obrazložite svoje razloge i pokušajte pojasniti za koje vrste zadataka.
9. Kakav je vaš opći dojam o ovim alatima? Koje prednosti i nedostaci po vašem mišljenju postoje u kontekstu učenja?
10. Ukoliko imate još nešto što želite dodati, slobodno napišite ovdje.

Osim glavnih pitanja, postavljena su još četiri dodatna pitanja:

- Koji fakultet pohađate?
- Odaberite razinu studija. (diplomski, prijediplomski)
- Koji studijski program pohađate?
- Koja je vaša trenutna godina studija? (1,2,3,4,5)

Prvo značajno ograničenje ovog istraživanja je neujednačena distribucija studenata između diplomskog (15 odgovora) i prijediplomskog (74 odgovora) studija, te činjenica da većina studenata dolazi s jednog sveučilišta (74 odgovora). Ova okolnost ograničava mogućnost donošenja zaključaka o razlikama između grupa. Ipak, dobiveni su zanimljivi i relevantni rezultati. Kada je to bilo prikladno, prikazan je broj studenata u kategorijama diplomskog i prijediplomskog studija.

# 4 Rezultati

Prikupljeno je ukupno 89 odgovora. Većina odgovora (74) odnosi se na prijediplomske studente, dok se 15 odgovora odnosi na diplomske studente. Među prijediplomskim studentima, svi su bili na trećoj godini studijskog programa. Kod diplomskih studenata, 8 ih je bilo na četvrtoj godini i 7 na petoj godini.

Na početno pitanje o korištenju alata umjetne inteligencije, samo 5 ispitanika navelo je odgovor "ne", pri čemu je 1 (tj. 6,6%) bio student diplomskog studija, a 4 (tj. 5,4%) studenta prijediplomskog studija. S obzirom na to da je glavni fokus ovog istraživanja usmjeren na upotrebu AI alata, odgovori tih 5 studenata isključeni su iz analize, ali će se razmatrati zasebno na kraju. Time ostaju 84 valjana odgovora.

Što se tiče drugog pitanja, većina je navela korištenje ChatGPT (62%), kako je prikazano na slici 1. Mi-

Microsoft Copilot was utilized by 12%, Bing Chat by 8%, Google Bard by 4%, and BlackBox.ai by 3%. A few other tools (10%) were mentioned by three or fewer students, including PHIND, you.com, Gencraft, Dall-E, RunwayML, Liama, ForeFront.ai, Amazon CodeWhisperer, and Codeium AI.

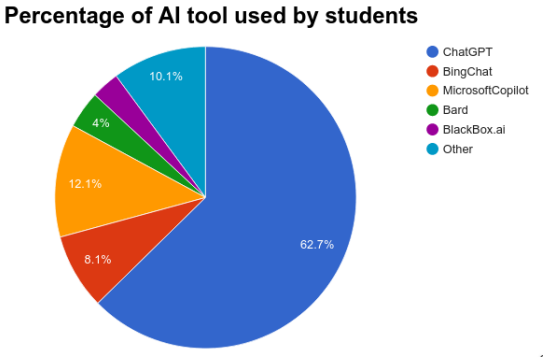


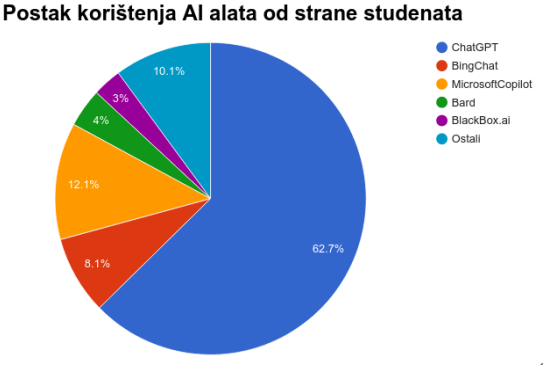

**Fig. 1:** Percentage of AI tool used by students

When they were asked if they used AI tools to solve assignments only 8 answered "no" (1 of them being graduate) and 76 answered "yes" (13 of them being graduates), while when asked specifically for programming assignments 17 answered "no" (6 of them being graduates) and 67 answered "yes" (8 of them being graduates). In other words, 10% (1 of 14) of graduates and 7% (7 of 70) of undergraduates answered "no" for using AI tool in assignments. On the other hand, 43% (6 of 14) of graduates and 16% (11 of 70) of undergraduates answered "no" for using AI tools in programming assignments. Based on this data, there is an indication that graduate students use fewer AI tools for programming assignments than undergraduates, although they exhibit similar overall usage in assignments in general. However, additional research is necessary to confirm this conclusion due to the limited number of graduate students. Moreover, it's noteworthy that nearly all undergraduate students (63 out of 70) are from a single university, whereas half of the graduate students (7 out of 14) belong to another university.

A total of 76 students declared their use of AI tools in solving assignments. Based on their essay responses, the following categories were identified and mentioned:

- Programming – 39 students (51.3%)
- Additional task explanations, including obtaining examples and visualizations – 11 students (14.5%)
- Generating ideas for solving problems – 10 students (13.2%)
- Literature search – 9 students (11.8%)

crosoft Copilot koristilo je 12%, Bing Chat 8%, Google Bard 4%, a BlackBox.ai 3%. Nekoliko drugih alata (10%) spomenuli su tri ili manje studenta, uključujući: PHIND, you.com, Gencraft, Dall-E, RunwayML, Liama, ForeFront.ai, Amazon CodeWhisperer i Codeium AI.




**Slika 1:** Postak korištenja AI alata od strane studenata

Na pitanje koriste li AI alate za rješavanje zadataka, samo 8 ih je odgovorilo "ne" (1 od njih bio je student diplomskog studija), dok je 76 odgovorilo "da" (13 od njih bilo je studenata diplomskog studija). Međutim, na pitanje o zadacima iz programiranja, 17 je odgovorilo "ne" (6 od njih bili su studenti diplomskog studija), dok je 67 odgovorilo "da" (8 od njih bili su studenti diplomskog studija). Drugim riječima, 10% (1 od 14) studenata diplomskog studija i 7% (7 od 70) prijediplomskih studenata odgovorilo je "ne" na korištenje AI alata u zadacima. S druge strane, 43% (6 od 14) studenata diplomskog studija i 16% (11 od 70) prijediplomskih studenata odgovorilo je "ne" na korištenje AI alata u zadacima programiranja. Na temelju ovih podataka, postoji indikacija da studenti diplomskog studija manje koriste AI alate za zadatke iz programiranja u usporedbi s prijediplomskim studentima, iako je slična ukupna upotreba AI alata u zadacima općenito. Međutim, potrebna su dodatna istraživanja kako bi se potvrdio ovaj zaključak zbog ograničenog broja studenata diplomskog studija. Također, važno je napomenuti da su gotovo svi prijediplomski studenti (63 od 70) s jednog sveučilišta, dok polovica studenata diplomskog studija (7 od 14) pripada drugom sveučilištu.

Ukupno 76 studenata izjavilo je da koriste AI alate u rješavanju zadataka. Na temelju njihovih esejskih odgovora, identificirane su i spomenute sljedeće kategorije:

- Programiranje – 39 studenata (51.3%)
- Dodatna pojašnjenja zadataka, uključujući dobivanje primjera i vizualizacija - 11 studenata (14.5%)
- Generiranje ideja za rješavanje problema - 10 studenata (13.2%)
- Pretraživanje literature – 9 studenata (11.8%)

- Completing assignments with the help of AI tools – 4 students (5.3%)
- Other (proofreading, sentence construction, writing summaries and reviews) – 3 students (3.9%)

Within the category of programming (51.3%), the following subcategories stand out:

- Debugging code – 16 students (21.1%) - they had an error that they could not solve and they asked the AI tool to find the error and correct the code.
- Generating partial solution – 8 students (10.5%) - asking AI tool to program for them a part of the solution like a function or class given in the assignments or generating SQL statements or database creation.
- Interpretation of program code – 8 students (10.5%) - they have a code from teachers of AI tool generated and they do not understand it and asked for explanation of the code.
- Assignment explanation – 5 students (6.6%) - asking the AI tool to explain what is meant in the assignment because the student did not understand it.
- Data generation – 2 students (2.6%) - generating data for a database.

From this it is clear that programming is the main reason why students use AI tools and that debugging and generating solutions is the main usage purpose. This is confirmed in the subsequent question, where they were specifically asked about the use of AI tools in programming assignments. For programming assignments a total of 67 students (8 of them being graduates) stated that they used AI tools, which is a larger number than in the previous question. For this question the answers of graduate and undergraduate level were analysed separately and no difference in views was seen. It is a small number of graduate students but still the percentages of students that mentioned certain category is very similar for both groups. In table 1 the number of students that mentioned using AI tools in certain way for programming assignments is presented. Again, the categories "Debugging code", "Generating partial solution" and "Interpretation of program code" where identified. In addition, two more categories where identified "For learning purposes" and "Code improvement ". Here "for learning purposes" means asking the AI tool to explain some concepts or to generate code examples for some things; in this category, only responses were counted if the students explicitly stated that the AI tool was used for learning purposes. The category "Code improvement" means optimization of the code, refactoring, clean code, etc. Next to the number of students the percentage is displayed. In all questions, some students wrote comments such as "for searching" or "for various reasons," but these answers were not counted as they did not provide expla-

- Dovršavanje zadataka uz pomoć AI alata - 4 studenta (5.3%)
- Ostalo (lektura, konstrukcija rečenice, pisanje sažetaka i recenzija) – 3 studenta (3.9%)

U okviru kategorije programiranja (51,3%), ističu se sljedeće podkategorije:

- Otklanjanje pogrešaka u kodu - 16 studenata (21,1%) - imali su pogrešku koju nisu mogli riješiti, pa su zatražili od AI alata da pronađe pogrešku i ispravi kod.
- Generiranje djelomičnog rješenja - 8 studenata (10,5%) - tražili su od AI alata da programira dio rješenja poput funkcije ili klase navedene u zadatku ili generiranja SQL naredbi ili stvaranja baze podataka.
- Objašnjenje programskog koda - 8 studenata (10,5%) - studenti imaju kod koji su dali nastavnici ili ga je generirao AI alat, ali ga ne razumiju, pa su zatražili objašnjenje koda.
- Objašnjenje zadatka - 5 studenata (6,6%) - tražili su od AI alata da objasni što se misli u zadatku jer nisu razumjeli.
- Generiranje podataka - 2 studenta (2,6%) - generiranje podataka za bazu podataka.

Iz ovoga je jasno da je programiranje glavni razlog zbog kojeg studenti koriste AI alate, a otklanjanje pogrešaka i generiranje rješenja glavna je svrha korištenja. To je potvrđeno u sljedećem pitanju, gdje su izravno pitani o korištenju AI alata u zadacima iz programiranja. Za ovakve zadatke ukupno 67 studenata (od toga 8 diplomskih studenata) navelo je da koristi AI alate, što je veći broj u odnosu na prethodno pitanje. Za ovo pitanje odgovori diplomskih i prijediplomskih studenata analizirani su zasebno i nije uočena razlika u stavovima. Iako se radi o malom broju studenata diplomskog studija, postotci studenata koji su naveli određenu kategoriju vrlo su slični za obje skupine. U Tablici 1 prikazan je broj studenata koji su naveli korištenje AI alata na određeni način za zadatke iz programiranja. Ponovno su identificirane kategorije "Otklanjanje pogrešaka u kodu", "Generiranje djelomičnog rješenja" i "Objašnjenje programskog koda". Dodatno su identificirane dvije kategorije "Za učenje" i "Poboljšanje koda". Ovdje "Za učenje" znači da studenti traže od AI alata objašnjenje pojedinih pojmova ili generiranje primjera koda za određene stvari; u ovoj kategoriji su uračunati samo odgovori ako su studenti izričito naveli da su AI alat koristili u svrhu učenja. Kategorija "Poboljšanje koda" odnosi se na optimizaciju koda, refaktoriranje, čisti kod itd. Pored broja studenata prikazan je postotak. U svim pitanjima neki studenti napisali su komentare poput "za pretraživanje" ili "iz raznih razloga", ali ti odgovori nisu uračunati jer nisu pružili objašnjenje specifične svrhe korištenja AI alata.

nations for the specific purposes of the AI tool usage.

**Table 1:** Number of mentioned usages of AI tools for programming assignments

| Category | G (8) | UG (59) | Total (67) |
|---|---|---|---|
| Debugging code | 4 (50%) | 37 (63%) | 41 (61%) |
| Generating partial solution | 4 (50%) | 31 (53%) | 35 (52%) |
| Interpretation of program code | 2 (25%) | 16 (27%) | 18 (27%) |
| For learning purposes | 2 (25%) | 12 (20%) | 14 (21%) |
| Code improvement | 1 (13%) | 2 (3%) | 3 (4%) |
| Note: G - Graduate, UG - Undergraduate | | | |

**Tablica 1:** Broj spomenutih načina korištenja AI alata za zadatke programiranja.

| Kategorija | D (8) | PD (59) | Ukupno (67) |
|---|---|---|---|
| Otklanjanje pogrešaka u kodu | 4 (50%) | 37 (63%) | 41 (61%) |
| Generiranje djelomičnog rješenja | 4 (50%) | 31 (53%) | 35 (52%) |
| Objašnjenje koda | 2 (25%) | 16 (27%) | 18 (27%) |
| Učenje | 2 (25%) | 12 (20%) | 14 (21%) |
| Poboljšanje koda | 1 (13%) | 2 (3%) | 3 (4%) |
| Napomena: D - Diplomski, PD - Prijediplomski | | | |

The next question sought their opinion on whether AI tools should be entirely allowed, partially allowed, or entirely forbidden for programming assignments. In this case, only one student answered "entirely forbidden" and provided an explanation, stating that students should seek help from teachers or other students to engage in meaningful conversation. Also, it was stated that this will then help collaboration and learn program solving as a team. Since some students on the previous question answered that they did not use the AI tools for programming assignments in Table 2 the data is presented for both groups how they responded to this question about allowing AI tools. It seems that students that did not use AI tools are more inclined to allow the tools while students that used AI tools are more inclined to partially allow. More data is needed to confirm such indications.

Sljedeće pitanje tražilo je mišljenje studenata o tome trebaju li AI alati biti u potpunosti dopušteni, djelomično dopušteni ili u potpunosti zabranjeni za zadatke iz programiranja. U ovom slučaju, samo je jedan student odgovorio "potpuno zabranjeni" i pružio objašnjenje, navodeći da bi studenti trebali tražiti pomoć od nastavnika ili drugih studenata kako bi se uključili u svrsishodan razgovor. Također je istaknuto da će to pomoći u suradnji i učenju rješavanja programa timski. Budući da su neki studenti u prethodnom pitanju odgovorili da nisu koristili AI alate za zadatke iz programiranja, u Tablici 2 prikazani su podaci za obje grupe kako su odgovorile na ovo pitanje o dopuštanju AI alata. Čini se da su studenti koji nisu koristili AI alate skloniji dopuštanju alata, dok su studenti koji su koristili AI alate skloniji djelomičnom dopuštanju. Potrebni su dodatni podaci kako bi se potvrdile takve indikacije.

**Table 2:** Student opinion about allowing or forbidding AI tools

| Group | Used AI | Not used AI |
|---|---|---|
| Allow | 41 (53%) | 7 (41%) |
| Partially Allow | 26 (39%) | 9 (53%) |
| Forbid | 0 (0%) | 1 (6%) |
| Total | 67 | 17 |

**Tablica 2:** Mišljenje studenata o dozvoli ili zabrani korištenja AI alata

| Grupa | Koristili AI | Nisu koristili AI |
|---|---|---|
| Dozvoliti | 41 (53%) | 7 (41%) |
| Djelomično dozvoliti | 26 (39%) | 9 (53%) |
| Zabraniti | 0 (0%) | 1 (6%) |
| Ukupno | 67 | 17 |

Based on the explanations that students gave for such answer the results were as follows. The students that responded completely allow reasoned in this way because:

- The AI tools can only give part of solutions - 29 students (63%)
- The AI tools are just help to program faster, nothing more - 27 students (59%)
- The code must be understood regardless, so why for-

Na temelju objašnjenja koja su studenti dali za svoj odgovor, rezultati su sljedeći. Studenti koji su za potpunu dozvolu razmišljaju tako jer:

- AI alati mogu pružiti samo dio rješenja - 29 studenata (63%)
- AI alati su samo pomoć za brže programiranje, ništa više - 27 studenata (59%)
- Programski kod treba razumjeti. Stoga, zašto zabranjivati ako ga studenti razumiju - 21 student (46%)

bid if a students understand - 21 students (46%)

- The AI tools help faster debugging - 14 students (30%)
- This is the same as searching on Google, only faster - 7 students (15%)
- Teachers can not know if a AI tool was used, why forbid it then? - 3 students (6%)
- It helps in complex assignments, but it needs to be forbidden in simple assignments in introductory courses - 3 students (6%)
- Students can learn more by using the tool - 2 students (4%)
- It is teachers job to make better evaluations - 1 student (2%)

On the other hand, the students that said partially allow, reasoned that this is so because:

- The tool is only for assistance and learning – 21 students (61.8%)
- It can be used for assistance and learning, but if usage of AI tools is regulated – 4 students (11.8%)
- The AI tools help faster debugging – 4 students (11.8%)
- It helps for simple and repetitive tasks – 3 students (8.8%)
- It helps solving complex tasks – 2 students (5.9%)

What's interesting when comparing these two groups is that students who are for entirely allowing AI tools provided more reasons to support their stance. In both cases, students perceive the tools as help for learning and problem-solving. However, it is concerning that only a small number of students recognize potential issues associated with using these tools. For instance, only one student expressed concern about the ownership of the code generated by AI tools. To further explore this matter, the next question asked students to enumerate the benefits and downsides of using AI tools.

What stands out in benefits (Table 3), both groups list the same benefits or they can be summarized in the same 3 categories, but the group that allow completely AI tools at first place has the benefit "Faster problem and assignments solving", while the group partially allow has the benefit "Learning support".

Regarding the downsides (Table 4), the results are more similar, but if one looks at the percentages can see the folowing. The group for allow the AI tools is more concerned about the tools not being reliable (answers that the tool generates are not always correct and not easy to use because of the missing context) than the

- AI alati pomažu bržem otklanjanju pogrešaka u kodu - 14 studenata(30%)
- To je isto kao i traženje na Googleu, samo brže - 7 studenata (15%)
- Nastavnici ne mogu znati je li korišten AI alat - zašto ga onda zabranjivati? - 3 studenta (6%)
- Pomaže kod složenih zadataka, ali treba biti zabranjeno kod jednostavnih zadataka - 3 studenta (6%)
- Studenti mogu više naučiti koristeći ovaj alat - 2 studenta (4%)
- Posao je nastavnika napraviti bolju evaluaciju - 1 student (2%)

S druge strane, studenti koji su za djelomičnu dozvolu, razmišljaju tako jer:

- Alat je samo za pomoć i učenje - 21 student (61.8%)
- Može se koristiti za pomoć i učenje, ali uz regulaciju korištenja AI alata - 4 studenta (11.8%)
- AI alati pomažu bržem otklanjanju pogrešaka u kodu - 4 studenta (11.8%)
- Pomažu kod jednostavnih i ponavljajućih zadataka - 3 studenta (8.8%)
- Pomaže u rješavanju složenih zadataka - 2 studenta (5.9%)

Ono što je zanimljivo prilikom usporedbe ove dvije skupine je da studenti koji su za potpunu dozvolu korištenja AI alata pružaju više razloga kako bi argumentirali svoje stajalište. U oba slučaja, studenti percipiraju AI alate kao pomoć pri učenju i rješavanju problema. Međutim, zabrinjavajuće je što samo mali broj studenata prepoznaje potencijalne probleme povezane s korištenjem tih alata. Primjerice, samo je jedan student izrazio zabrinutost u vezi vlasništva nad programskim kodom koji generiraju AI alati. Kako bi se dalje istražila ova tema, sljedeće pitanje odnosi se na prednosti i nedostatke korištenja AI alata.

Ono što se ističe u pogledu prednosti (Tablica 3) jest da obje skupine navode iste prednosti ili ih se može sažeti u iste 3 kategorije, ali skupina koja potpuno dopušta korištenje AI alata na prvom mjestu ističe prednost "Brže rješavanje problema i zadataka", dok skupina koja djelomično dopušta AI alate ističe prednost "Podrška u učenju".

Što se tiče nedostataka (Tablica 4), rezultati su sličniji, ali ako se pogledaju postoci se može vidjeti sljedeće. Grupa koja podržava korištenje AI alata više se brine o pouzdanosti tih alata (odgovori koje alat generira nisu uvijek točni i nisu jednostavni za korištenje zbog izostanka konteksta) u usporedbi s grupom koja

**Table 3:** Benefits of using AI tools for programming assignments

| Category | A | PA |
|---|---|---|
| Faster problem and assignments solving | 35 (76%) | 8 (23%) |
| Learning support | 18 (39%) | 19 (54%) |
| Quick source of information | 13 (28%) | 14 (40%) |
| Note: A - Allow (48 students) PA - Partially allow (35 students) | | |

**Tablica 3:** Prednosti korištenja AI alata za zadatke iz programiranja

| Kategorija | D | DD |
|---|---|---|
| Brže rješavanje problema i zadataka | 35 (76%) | 8 (23%) |
| Podrška u učenju | 18 (39%) | 19 (54%) |
| Brzi izvor informacija | 13 (28%) | 14 (40%) |
| Napomena: D - Dozvoliti (48 studenata) DD - Djelomično dozvoliti (35 studenata) | | |

group that is for partial allow. On the other hand, regarding learning, the group that is for partial allow is more concerned. This group is concerned that students may not truly learn, expressing worries about potential "knowledge gaps" and a lack of understanding of concepts if they become too reliant on AI tools.

podržava djelomično korištenje. S druge strane, što se tiče učenja, grupa koja podržava djelomično korištenje izražava veću zabrinutost. Ta grupa brine se da studenti neće stvarno naučiti, izražavajući zabrinutost zbog mogućih "rupa u znanju" i nedostatka razumijevanja koncepata, ako postanu previše oslonjeni na AI alate.

**Table 4:** Downsides of using AI tools for programming assignments

| Category | A | PA |
|---|---|---|
| Not reliable | 30 (65%) | 16 (46%) |
| Learning issue | 14 (30%) | 13 (37%) |
| Ethical concerns | 4 (9%) | 3 (9%) |
| Note: A - Allow (48 students) PA - Partially allow (35 students) | | |

**Tablica 4:** Nedostaci korištenja AI alata za zadatke iz programiranja

| Kategorija | D | DD |
|---|---|---|
| Nepouzdanost | 30 (65%) | 16 (46%) |
| Problematika učenja | 14 (30%) | 13 (37%) |
| Etička pitanja | 4 (9%) | 3 (9%) |
| Napomena: D - Dozvoliti (48 students) DD - Djelomično dozvoliti (35 students) | | |

From this, it can be concluded that the students who are skeptical of using AI tools are those who prioritize learning and aim to improve their knowledge. On the other hand, students who blindly trust AI tools seem to prioritize generating solutions rather than attempting to work through problems independently. To confirm these differences, further research is needed. It would be particularly interesting to examine the final grades that students received and explore whether there is any correlation between their answers and academic performance.

In the last question, only 17 students responded and they did not give any additional information that was not already mentioned in previous questions. Finally, the responses of the 5 students that were excluded at the beginning because they stated that they did not use any AI tool were analysed. All of them support partial allowance, and their reasons align with the students who used the tools and said partially allow, but one thing that was not mentioned before was as follows. "A change in the approach to creating assignments is needed because the emergence of AI tools has made solving assignments independently pointless."

Iz ovoga se može zaključiti da su studenti koji su skeptični prema korištenju AI alata oni koji stavljaju naglasak na učenje i teže unapređenju svog znanja. S druge strane, studenti koji slijepo vjeruju u AI alate čini se da daju prednost generiranju rješenja umjesto da pokušaju samostalno riješiti problem. Kako bismo potvrdili ove razlike, potrebna su daljnja istraživanja. Bilo bi posebno zanimljivo istražiti konačne ocjene koje su studenti dobili i istražiti postoji li povezanost između njihovih odgovora i akademskih rezultata.

U posljednjem pitanju, samo je 17 studenata odgovorilo. Nisu pružili nikakve dodatne informacije koje već nisu bile spomenute u prethodnim pitanjima. Konačno, analizirani su odgovori 5 studenata koji su isključeni na početku jer su naveli da ne koriste nijedan AI alat. Svi oni podržavaju djelomično dopuštanje, a njihovi razlozi podudaraju se sa studentima koji su koristili AI alate i odabrali djelomičnu dozvolu. No, jedna stvar koja nije bila spomenuta prije bila je sljedeća: "Potrebna je promjena pristupa u izradi zadataka jer je pojava AI alata učinila bespredmetnim rješavanje zadataka samostalno."

# 5 Discussion

While there is no distinction between graduate and

# 5 Diskusija

Iako nema razlike između diplomskih i prijediplom-

undergraduate students regarding the use of AI tools for assignments, a difference does emerge when considering the utilization of AI tools specifically for programming assignments among graduate and undergraduate students. It is essential to note that only 14 responses were gathered from graduate students, necessitating further research for a comprehensive understanding.

From the survey it is clear that students use the AI tool mostly for programming. The utilization of AI tools for debugging and learning is not perceived as an issue. However, the elevated reliance on these tools for generating either partial or complete solutions is a cause for concern and should be addressed appropriately. This observation is evident from the enumerated benefits and downsides. Students perceive the potential for assistance in learning and faster assignments solving. But do they really learn faster? What is concerning is that the primary issue with AI tools, as identified by students, is their occasional generation of incorrect answers and inability to solve entire assignments. Surprisingly, only 7 out of 84 students mentioned ethical issues associated with using such tools. Of the total, 27 students recognize that the improper use of tools can result in knowledge gaps, constituting approximately 30%. This leaves around 70% who did not raise this particular concern.

Based on this, it becomes evident that supporting the suggestion made by other studies, students require additional instructions and information on the correct utilization of tools. Prohibiting the use of AI tools outright is not feasible. Despite the existence and enhancement of tools designed to verify the authenticity of text or code generation, the AI tools themselves will continue to evolve. One potential scenario is the commercialization and restriction of access to many AI tools, which may potentially reduce their usage among students, although this remains uncertain.

As some students acknowledged, teachers need to design more complex assignments and enhance their assessment methods to better evaluate understanding. In this context, traditional pen-and-paper exams still play a crucial role. Efforts should be made to prevent the easy copying and pasting of assignments into AI tools. Providing students with a picture in the assignment that needs to be developed could be more effective, as it requires them to put effort into describing the problem, which might not be trivial. Our hope is that some students understand the downsides and will assist their colleagues in understanding them as well.

Compared to (Saari et al., 2022), where 35% of students reported not using the tool, our results indicate a significantly lower percentage, with only 5% of students reporting no tool usage. Additionally, while (Saari et al., 2022) reported a 25% usage for debugging, our findings show a notably higher figure at around 61%. This variance could be attributed to either a

-skih studenata u vezi korištenja AI alata za zadatke, razlika se pojavljuje kada se razmotri korištenje AI alata posebno za zadatke iz programiranja među diplomskim i prijediplomskim studentima. Važno je napomenuti da je prikupljena samo 14 odgovora diplomskih studenata, što zahtijeva daljnja istraživanja radi potpunijeg razumijevanja.

Iz ankete je jasno da studenti uglavnom koriste AI alate za programiranje. Upotreba AI alata za otklanja-nje pogrešaka i učenje ne percipira se kao problem. Me đutim, oslanjanje na te alate za generiranje djelo-mičnih ili potpunih rješenja izaziva zabrinutost i tre-balo bi ih adekvatno rješavati. Ova zapažanja proizlaze iz nabrojenih prednosti i nedostataka. Studenti perci-piraju potencijalnu pomoć u učenju i brže rješavanje zadataka. No, pitanje je postižu li zaista brže razu-mijevanje gradiva? Ono što zabrinjava jest da je pri-marni problem s AI alatima, prema studentima, njihovo povremeno generiranje netočnih odgovora i nesposob-nost rješavanja cjelovitih zadataka. Iznenađujuće je da samo 7 od 84 studenta spominje etičke probleme po-vezane s korištenjem takvih alata. Od ukupnog broja, 27 studenata prepoznaje da neprikladna upotreba AI alata može rezultirati rupama u znanju, što čini otpri-like 30%. To znači da oko 70% studenata nije izrazilo takvu zabrinutost.

Na temelju ovoga postaje očito da uvažavajući prijedloge drugih istraživanja, studentima trebaju dodatne upute i informacije o ispravnoj upotrebi AI alata. Potpuna zabrana korištenja AI alata nije izvediva. Unatoč postojanju i poboljšanju alata dizajniranih za provjeru autentičnosti teksta ili generiranja koda, sami AI alatiće se nastaviti razvijati. Jedan potencijalni scenarij je komercijalizacija i ograničavanje pristupa mnogim AI alatima, što potencijalno može smanjiti njihovu upotrebu me đu studentima, iako to nije sigurno.

Kako su neki studenti naveli, nastavnici trebaju osmis-liti složenije zadatke i poboljšati svoje metode procjene kako bi bolje procijenili razumijevanje od strane stude-nata. U tom kontekstu, tradicionalni ispiti s olovkom i papirom i dalje igraju ključnu ulogu. Trebalo bi uložiti napore kako bi se spriječilo jednostavno kopiranje za-dataka u AI alate. Davanje zadatka u obliku slike može biti učinkovitije od tekstualnog zadatka, jer je potreban trud u opisivanju problema, što nije nužno trivijalno. Nadamo se da će neki studenti shvatiti nedostatke AI alata i pomoći svojim kolegama da ih isto shvate.

U usporedbi s (Saari et al., 2022), gdje se 35% studenata izjasnilo da ne koristi alat, naši rezultati ukazuju na značajno niži postotak, sa samo 5% studenata koji nisu koristili alat. U usporedbi s (Saari et al., 2022) gdje se stopa korištenja od 25% odnosi na otklanjanje pogrešaka, naša istraživanja pokazuju primjetno veću brojku od otprilike 61%. Ova razlika može se pripisati

distinct group of students or, more likely, the evolving landscape over time, with AI tools now being more widely accepted.

What we perceive as a potential issue is that in the EU, the "Bologna Process" (Bobić et al., 2005) of teaching in higher education, which emphasizes continuous learning and regular homework assignments may now be endangered by the widespread use of AI tools. With this in mind, it becomes essential for us to revisit and reassess our evaluation models, as well as reexamine our teaching strategies.

# 6 Conclusion

The widespread use of AI tools has become a reality, and students will use them regardless of whether they are allowed or not. While it's unsurprising that ChatGPT is a prevalent choice, it's evident from the survey that many other tools are also in use. Given the inevitable use of AI tools, the most prudent approach is to educate students on their usage, outlining both the benefits and ethical concerns associated with them. Teachers should emphasize the positive aspects of these tools while taking on the responsibility of instructing students on their responsible use, including awareness of potential harm to themselves and/or others. Additionally, there is a need for concurrent changes in evaluation methods aligned with evolving educational challenges.

# References

Bain, G. and Barnes, I. (2014). Why is programming so hard to learn? In *Proceedings of the 2014 conference on Innovation & technology in computer science education*, pages 356–356.

Bobić, M. P., Bjeliš, A., Divjak, B., Mihalić, A., Grubišić, K., Franz-Štern, R., and Kramar, S. (2005). *Prvi koraci u Bolonjskom procesu*. Sveučilišna tiskara, d.o.o.

Cosma, G. and Joy, M. (2008). Towards a Definition of Source-Code Plagiarism. *IEEE Transactions on Education*, 51(2):195–200.

Ellis, M. E., Casey, K. M., and Hill, G. (2024). Chatgpt and python programming homework. *Decision Sciences Journal of Innovative Education*, n/a(n/a).

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Alshami, S. A., and Slamene, R. (2023). Analyzing the students' views, concerns, and perceived ethics about chat gpt usage. *Computers and Education: Artificial Intelligence*, 5:100180.

Guzdial, M. (2010). *Making Software: What Really Works, and Why We Believe It.*, chapter Why is it so

različitoj skupini studenata ili ono što je vjerojatnije, promjenama tijekom vremena, i sve većom prihvaćenošću AI alata.

Ono što prepoznajemo kao potencijalni problem je da bi "Bolonjski proces" (Bobić et al., 2005) visokog obrazovanja u EU, koji naglašava kontinuirano učenje i redovne domaće zadaće, sada mogao biti ugrožen zbog širokog korištenja AI alata. S obzirom na to, postaje ključno ponovno razmotriti i procijeniti naše modele vrednovanja, kao i ponovno ispitati naše metode poučavanja.

# 6 Zaključak

Sveprisutno korištenje AI alata postalo je stvarnost i studenti će ih koristiti bez obzira jesu li dopušteni ili ne. Iako nije iznenađujuće da je ChatGPT čest izbor, očito je iz ankete da se koriste i mnogi drugi alati. S obzirom na neizbježnu upotrebu AI alata, najrazboritiji pristup je educirati studente o njihovoj upotrebi, ističući prednosti, ali i etička pitanja povezana s njima. Nastavnici bi trebali naglasiti pozitivne aspekte ovih alata, preuzeti odgovornost za upućivanje studenata na njihovu odgovornu upotrebu, uključujući podizanje svijesti o mogućim štetnim posljedicama za njih i/ili druge. Konačno, potrebne su promjene u metodama vrednovanja koje su usklađene s novim izazovima u obrazovanju.

# Literatura

Bain, G. and Barnes, I. (2014). Why is programming so hard to learn? In *Proceedings of the 2014 conference on Innovation & technology in computer science education*, pages 356–356.

Bobić, M. P., Bjeliš, A., Divjak, B., Mihalić, A., Grubišić, K., Franz-Štern, R., and Kramar, S. (2005). *Prvi koraci u Bolonjskom procesu*. Sveučilišna tiskara, d.o.o.

Cosma, G. and Joy, M. (2008). Towards a Definition of Source-Code Plagiarism. *IEEE Transactions on Education*, 51(2):195–200.

Ellis, M. E., Casey, K. M., and Hill, G. (2024). Chatgpt and python programming homework. *Decision Sciences Journal of Innovative Education*, n/a(n/a).

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Alshami, S. A., and Slamene, R. (2023). Analyzing the students' views, concerns, and perceived ethics about chat gpt usage. *Computers and Education: Artificial Intelligence*, 5:100180.

Guzdial, M. (2010). *Making Software: What Really Works, and Why We Believe It.*, chapter Why is it so hard to learn to program, pages 111–124. O'Reilly Media.

hard to learn to program, pages 111–124. O'Reilly Media.

Novak, M. (2020). *Effect of source-code preprocessing techniques on plagiarism detection accuracy in student programming assignments*. PhD thesis.

Saari, M., Rantanen, P., Nurminen, M., Kilamo, T., Systä, K., and Abrahamsson, P. (2022). Survey of ai tool usage in programming course: Early observations. In *International Conference on Agile Software Development*, pages 182–191. Springer Nature Switzerland Cham.

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., and Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative ai on assessments in higher education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale*, L@S '23, page 378–382, New York, NY, USA. Association for Computing Machinery.

Yilmaz, R. and Karaoglan Yilmaz, F. G. (2023). Augmented intelligence in programming learning: Examining student views on the use of chatgpt for programming learning. *Computers in Human Behavior: Artificial Humans*, 1(2):100005.

Novak, M. (2020). *Effect of source-code preprocessing techniques on plagiarism detection accuracy in student programming assignments*. PhD thesis.

Saari, M., Rantanen, P., Nurminen, M., Kilamo, T., Systä, K., and Abrahamsson, P. (2022). Survey of ai tool usage in programming course: Early observations. In *International Conference on Agile Software Development*, pages 182–191. Springer Nature Switzerland Cham.

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., and Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative ai on assessments in higher education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale*, L@S '23, page 378–382, New York, NY, USA. Association for Computing Machinery.

Yilmaz, R. and Karaoglan Yilmaz, F. G. (2023). Augmented intelligence in programming learning: Examining student views on the use of chatgpt for programming learning. *Computers in Human Behavior: Artificial Humans*, 1(2):100005.

# Strategic Planning and Decision Making

Nikola Kadoić
**Understanding the Impact of the Voting Model at the Eurovision Song Contest: Why Croatian Baby Lasagna did not win the ESC 2024?**

Jelena Gusić Munđar, Tena Jagačić, Nikola Kadoić
**Bridging Theory and Practice: Enhancing Business Decision-Making Course through Project-Based and Work-Based Learning**

Maja Pušnik, Katja Kous, Boštjan Šumak
**Enhancing Strategic planning course with operational research methods: a case study for IT students**

# Understanding the Impact of the Voting Model at the Eurovision Song Contest: Why Croatian Baby Lasagna did not win the ESC 2024?

**Nikola Kadoić**
University of Zagreb Faculty of organisation and informatics
Department for organisation
Pavlinska 2, HR- 42000 Varaždin, Croatia
nikola.kadoic@foi.unizg.hr

***Abstract***. *This paper focuses on the voting system applied in the final of the Eurovision Song Contest (ESC) 2024. The winner was the song that achieved the maximum sum of jury and public points. Each national jury consisted of five music experts. They ranked each song based on various criteria. Then, the national jury points were calculated using the exponential weight model. In this paper, the voting system was demonstrated and critically analysed, and its good and weak points were identified. The ESC24 winner was the representative from Switzerland, who finished only the 5th according to the public votes and the 1st according to the jury votes. Croatian Baby Lasagna was the 3rd, respecting the jury votes, and the 1st, considering the audience votes. The analysis identified at least two unanswered questions. (1) How can the audience be united (by music) when it must agree with their fifth choice? (2) Would national jury points remain the same if another five experts did the jury job? While the first question is more philosophical and not primarily focused in this analysis, answering the second question requires a scientific approach, and the results, if the answer is positive, could help the organisers justify the current jury voting system; on the contrary, if the answer is negative, the organiser has to consider making additional changes to the voting system. The preliminary simulation of differences in national jury points, when four-member juries differ in only one member, indicates the negative answer to the second question. The main conclusion is that the weak points of the voting system are still high, and further changes in the system are mandatory.*

**Keywords.** Eurosong, public, jury, rankings, votes, music, exponential weight model

## 1 Introduction

Eurovision Song Contest (ESC) is one of the most popular events in Europe. Its host is the European Broadcasting Union (EBU). Its popularity is also visible in Asia, Africa, and Australia. Many researchers have studied this event in history. While some were more focused on music styles and performance, others were more focused on other aspects, such as the current political environment, marketing strategies, different social issues, inclusion of technology, sentiment analysis for the results prediction, and others. One crucial aspect is the analysis of the voting system, which is the focus of this paper. More precisely, the primary focus is analysing jury voting rules. The main research question of the paper is: *What are the strengths and weaknesses of ESC's current jury voting system?*

At the beginning of this festival, the winner was selected by national juries who gave points to songs and countries. With the development of information and telecommunication technologies (ICT), it was possible to introduce the audience to the voting system. It was and still is not possible to vote for own country. From 1997 to 2008, the winner was determined only by the audience's votes. The juries were just a backup option for the cases experiencing technical problems. A combined approach was applied starting in 2009. The aggregation of national public and jury votes changed throughout the years. At first, the ranks of both components were combined into a single set of points, and later, starting in 2016, each country gave two independent sets of jury and public points (Wikipedia authors, 2024).

The motivation for this paper came from the Croatian result at ESC 2024. Even though Croatian representative Baby Lasagna achieved the best result for Croatia ever (since Croatian independence), 2nd place, many argue about the fairness of the voting system rules, primarily related to the jury part of the voting system. Through the years, the ESC organiser, EBU, recognised the weak points of jury inclusion in selecting winners and consequently constantly applied modifications to the voting system.

Additionally, they recognised the weak points of the only-audience system, which were triggers to return the juries to the game. The official organiser says: "*Using national juries … allows each song to be*

*considered individually. It ensures the best qualitative ranking of all participants …* "(European Broadcast Union, n.d.). Further, the jury is back to avoid the influence of neighbors and diaspora voting (Michiel Vos, n.d.). However, it is questionable whether this intention has been met and whether the weak points of the only-audience voting system are higher or lower than the combined voting system.

The paper is organised as follows: Section 2 presents the short literature review related to ESC and gives the list of perspectives that were the focus of the interest of different researchers regarding the ESC. Section 3 presents the methodology for the identification of the winner of the ESC. Section 4 presents the exponential weight model applied in the jury voting process. Section 5 presents the identification of the ESC voting system's weak points and good points. Section 6, the concluding section, proposes upgrading the ESC voting systems.

# 2 A brief literature review on ESC

The idea of this section is not to do a comprehensive literature review related to the ESC but to present examples of research that are related to the ESC and focused on some particular issue. Considering the topic of this paper, the most important is to mention the literature related to the field of Computational Social Choice (COMSOC), which has two heterogeneous layers: (1) Social Choice Theory and related social disciplines (such as Economics, Political science, Philosophy, Psychology) and (2) Computer Science and related disciplines (such as Artificial Intelligence (AI), Mathematics, Data Science, Operational Research) (Dodevska et al., 2020).

The research mainly related to the first layer of COMSOC includes the following research. Cultural diplomacy was the focus of the research by Ginsburgh and Noury (2008). The European identity of ESC was analysed (Carniel, 2015; Coupe & Chaban, 2020). The political situation related to Russian participation in ESC was analysed (Alpatova, 2022). Branding using a cultural perspective was the focus of Baker (2015). Additionally, different biases that influence public voting were analysed by many authors (Budzinski & Pannicke, 2017; Clerides & Stengos, 2006; Dekker, 2007; Dogru, 2013; Mantzaris et al., 2017; Siganos & Tabner, 2020). The media exposure of the countries' representatives was investigated by (Abakoumkin, 2018; Stockemer et al., 2018; Verrier, 2012).

However, the second layer of COMSOC is more interesting for this analysis. The research in this manner is connected to predicting the results, evaluating the voting methods respecting their characteristics, creating new voting methods and participatory models, contest participation, and simulations of crowd voting. The following examples are more oriented to the second layer. Simulations of voting behaviour through the years were analysed in (Budzinski et al., 2023). Further, a simulation of how would the ESC 2021 results look like if the majority judgment method had been applied is presented in (Umair et al., 2022). For purposes of simulating the voting process by crowds, Crowd-PrefRL (Preference-Based Reward Learning from Crowds) can be possibly used, but each user in the crowd must provide a preference label for each preference query (Chhan et al., 2024). A simulation of how social influence affects voting is investigated in (Ganser & Keuschnigg, 2018), and similar simulations can be done in the ESC context.

In the time of wide AI application, it can be used to evaluate alternatives and winner selection in different competitions. The example is given in (Haqbeen et al., 2022). In the case of ESC, a similar application can be made. However, it is questionable whether the organiser and crowd will accept it. Furthermore, a machine learning tool, sentiment analysis, can predict the results. The example related to the ESC 2019 is given in (Kumpulainen et al., 2020).

A question: *Are the jury experts better quality judges than the audience?* was the subject of interest to several researchers in the scientific community, and many individuals without a scientific background were interested in ESC. In terms of scientific research, there are opposite results in different research. In the research conducted in 2005, the authors concluded that jury experts are better quality judges. They came to this conclusion after analysing the results of the ESC before the research was conducted (Haan et al., 2005). Newer research argued, "that televoters are closer to the targets than experts" (Ginsburgh & Moreno-Ternero, 2023). Additionally, some research in progress concluded that the national juries' rankings are less correlated than national audiences' rankings, which is opposite to the expectations—if juries are returned to recognise the quality of songs, how they are so uneven in recognising the quality (they have more different opinions on the quality of song among themselves, than the audience). The research from 2005 could possibly be the EBU's motivation for returning the juries to the competition, but that research should also be observed in light of the period when it was implemented, and the EBU has to rely more on novel research.

Some research proposes new ways of voting and aggregating expert and crowd opinions. A model based on expert and crowd-voting advantages is proposed in (Dodevska et al., 2020). Furthermore, a machine-learning model for the fusion of crowd and experts' opinions was developed in (Kovacevic et al., 2020). The model was also applied to ESC 2016-2018. Another proposal for a framework for the aggregation of crowd and expert opinions based on bargaining theory and optimisation was given in (Vukicevic et al., 2022). Even though the proposed models are substantial and scientifically justified, their main problem is their acceptance by organisers and the crowd. Will somebody accept the voting method if they do not understand it?

# 3 ESC 2024 (final) voting system

This section presents how the voting system was applied to the ESC 2024 final (European Broadcasting Union, 2024). At the ESC 2024, 37 countries participated in total. Six of them directly went to the finale, and others had to compete in the two semifinals first. In the final, 25 countries competed to win. Ten countries were selected for the final in the first semifinal, and ten were selected in the second semifinal. Five "big countries" participated in the final directly, without competing in the semifinals (Spain, France, United Kingdom, Germany and Italy). Additionally, the winner of the previous ESC, Sweden, went directly to the final. Due to the disqualification of the Netherlands, instead of 26, there were 25 countries in the final.

Even though all 37 countries did not compete in the final, they could participate in the voting procedure of the ESC final. The winner was to be selected using the two sets of points per country: one set is related to the jury points (1-8, 10, and 12), and the other set is related to the public points (1-8, 10, and 12). The process of voting at ESC 2024 is presented in Fig. 1.

## 3.1 Audience voting

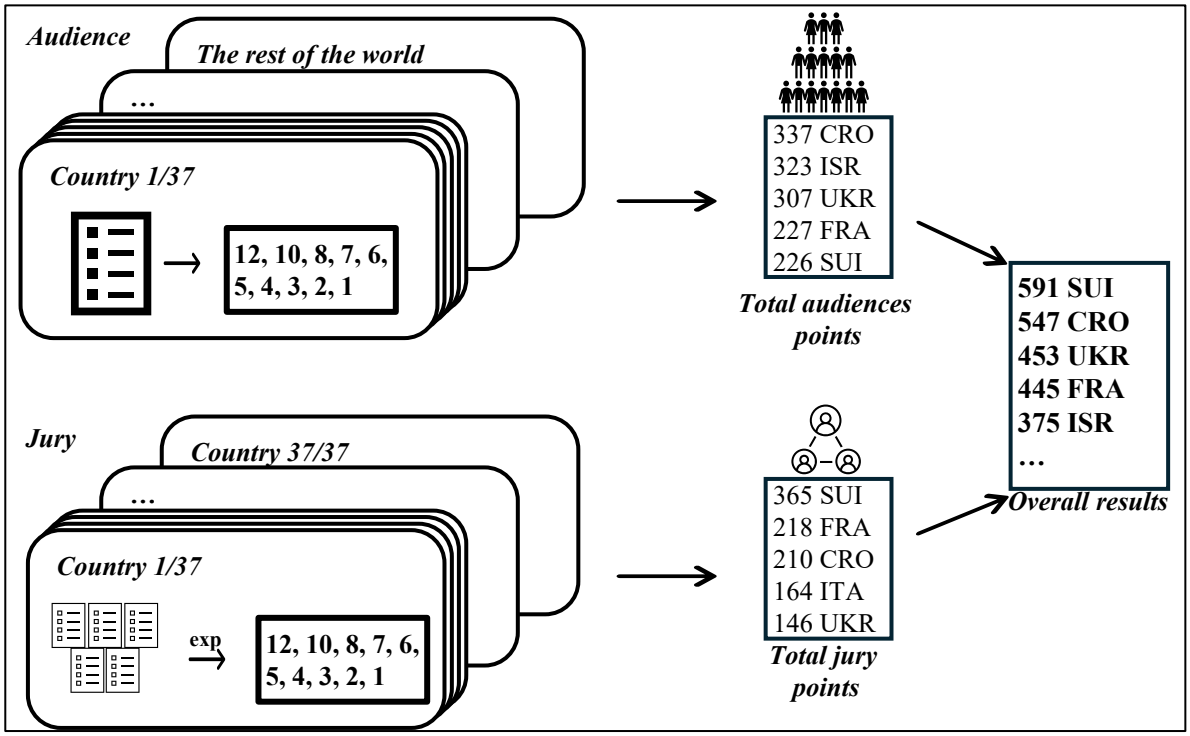
Voters (public, audience) from all 37 countries could vote for their favorites. In each country, after all votes were summed, the first-ranked song received 12 points, the second one received 10 points, the third one received 8 points, and so on, and the tenth-ranked song received 1 point, and all other songs did not receive any points. Additionally, the 38th public—the rest of the world—got one set of points. Consequently, there were 38 sets of public points. In theory, it was possible to achieve up to 444 points from the audience (if a particular country received 12 points from other 36 countries and the rest of the world).

## 3.2 Jury voting

Regarding the jury voting, the procedure is as follows:

1. Selecting the jury members by national broadcaster. Each of the 37 countries had to create a five-member jury consisting of professionals in the field (members of the national juries must pursue one of the following professions within the music industry: radio DJ, artist, composer, author of lyrics, or music producer). There should also be one backup juror for the case if one of the five members were not to participate in the voting procedure.
2. Familiarising with voting rules. National juries shall vote in all countries and in all cases. All countries shall appoint a national jury to vote in the final (even if their song is not selected for the final). Before voting, the juries had to get familiar with how to vote. The official Public Rules for ESC 2016 states, "*Each national jury shall vote in accordance with the instructions included in the so-called "Green Document"*. However, the *Green document* has never been published. It is not known if the Green Document was still valid in 2024. From what is publicly available, it can be concluded that the document contains instructions on evaluating

**Figure 1.** The process of voting at ESC 2024



the quality and voting at the ESC. The official ESC website mentioned four criteria that jury members should evaluate: (1) composition and originality of the song, (2) quality of the performance on stage, (3) vocal capacity of the performer(s), and (4) overall impression of the act.
It is not known whether the four criteria have been equally important in making the decisions on voting and whether the songs are evaluated by each juror respecting each criterion separately of having mind on all the criteria in mind at the same time during each song evaluation. Additionally, if the evaluations of songs were implemented using some predefined scale was not known.

3. Preparing for live shows. After the jury members got familiar with their job, they waited for their semifinal and final dress rehearsals. From the official website: "*When voting, Jury Members shall use all their professional skill and experience without favoring any Contestant on the account of their nationality, gender or likeliness and shall be free from bias, external influence foreign perspectives or pressure (including but not limited to betting odds or public opinions). The sole reference for their judgment must be the performance of the Acts during the second (Jury) Dress Rehearsals of the respective Shows."* Some informal ways of communication (such as forums) mention signing statements that confirm that voting will be implemented respecting the previous (and other) instructions.
4. (Voting in the semifinals. Even the focus of this section focuses on voting only in finals, but to get to the bigger picture, here it is noteworthy to mention that the juries also voted in the semifinals (only from competing countries) and that their votes did not affect the semifinals' results but were an option to use jury votes in case of technical problems in some country as a backup option).
5. Voting in the final. Voting in the final was implemented in the dress rehearsal using the instructions jurors received after they agreed to participate in the jury voting (second step). Each juror became the voting form and had to rank all competing songs.
6. Calculating the points respecting the jury votes. After each juror ranked all the competing songs, the ranks were further aggregated into the final national jury points using the exponential weight model (explained in detail in Section 4). There were 37 national juries in the final, and if a country had received 12 points from all 36 remaining national juries, it could have achieved a maximum of 432 points. The distribution of points from the national juries was familiar to the organiser before the final took place because the final dress rehearsal is usually held a day before the final. Some argue that this is a polygon for possible “influences” on the results.

## 3.3 Results announcement

In the end, after all the songs were performed and after the public voting procedure was finished, it was possible to announce the results.

- Firstly, the jury votes were announced: an announcer from each of the 37 countries was included in the live show announcing the song that received 12 points from their jury, while the other points (1-8 and 10) were presented at the table just before 12-points announcement, during the short, small talk between the announcer and ESC host.
- Secondly, the audience votes were announced in a summative way, starting with the country that was ranked last after the jury votes announcement. So, the hosts read the total number of audience points for the country ranked as last. Further, the total number of audience points were revealed for the country ranked second to last and so on up to the song that is ranked first. This way of revealing the audience's points created an intense atmosphere until the last points were revealed, which was the organiser's goal since the winner was known in some previous ESCs before all the countries presented their votes.

## 3.4 ESC 2024 specificities

Even though some cheating by juries was not registered, or at least not revealed, in the history of jury voting, there were examples of cheating by the national juries. In 2022, six national juries (Azerbaijan, Georgia, Montenegro, Poland, Romania, and San Marino) were caught cheating. They agreed to vote for each other's. The change from a linear to an exponential weight model in 2018 was also motivated by the cases of favoring or disfavouring some contestants by an individual member of some national jury.

Additionally, each ESC, including ESC 2024, was characterised by different specificities. While some argue that those specificities were just a part of the promoting strategies, others argue that they should not have a place at the ESC. The examples include Israel's participation at the ESC, the exclusion of Russia from the competition, and the scandal with the contestant representing the Netherlands (who made a "threatening move" towards one of the crew members), which was consequently disqualified. All those and other specificities caused some countries to withdraw almost 25 minutes before the final. These events are currently not the focus of this paper, but other researchers will undoubtedly analyse them, focusing on ESC's political perspective. It is possible that those events also had an impact to the final results.

## 3.5 Open questions

After the last points were announced, it was clear that the winner was Nemo from Switzerland, the leader

after the jury voting. However, he had only the fifth result of audiences. Many argue that the Swiss win is undeserved and that it is not acceptable that the winner is the song, which is the only fifth choice made by the audience. This is not the first time the winner is not the audience's favorite. Similar situations happened before (for example, just one year before, Sweden vs. Finland), but it was the first time that the difference between the winner's rank considering the jury voting (and overall rank) and the winner's rank considering the audience voting was so high.

In this direction, sentiment analysis was conducted among the ESC audience, identified many questions, and some of them were:

1. Is it justified that the juries have an equal set of points as the audience (not considering the rest of the world)? There were 185 jury members from 37 countries in total. On the other hand, there were millions and millions of viewers and voters from all over the world.
2. Is the jury voting system good and fair? Are the ESC voting system's weak points, which include the jury, lower than the weak points of the voting system, which does not include the jury? Has introducing the jury into the game solved weak points of the only-audience voting system or caused new and more significant problems?
3. How do we ensure that juries evaluate the quality of songs objectively? What is the definition of a quality song? What is the measuring instrument related to the quality of songs?
4. Is it justified to promote the ESC slogan *United by music* and at the same time conclude that the winner is only the fifth choice of the audience? How can anything and anyone be united if they have to agree on their fifth choice?
5. Is the Eurovision organised for the audience or juries?

In this paper, the previous questions were not entirely answered but at least partly analysed, and this analysis could be potentially valuable in designing a better voting system, if possible.

# 4 Exponential weight model vs. linear weight model in jury voting

In this section, a demonstration of the exponential weight model is presented. The exponential weight model was introduced in 2018. Before 2018, the linear weight model was applied.

The main difference between the two models is presented in Table 1. The table contains (possibly) real data from the French jury related to the ESC final

**Table 1.** The difference between linear and exponential weight model

| | Linear weight model | | | | | | | | Exponential weight model | | | | | | | | DIFF | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **AV** | **R** | **P** | **1** | **2** | **3** | **4** | **5** | **SUM** | **R** | **P** | **R** | **P** |
| **Malta** | 17 | 14 | 8 | 22 | 9 | 14 | 15 | | 0,5745 | 1,0157 | 3,1748 | 0,2222 | 2,6256 | 7,6128 | 15 | | 0 | 0 |
| **Albania** | 6 | 12 | 20 | 14 | 16 | 13,6 | 13 | | 4,642 | 1,4851 | 0,3249 | 1,0157 | 0,6947 | 8,1624 | 13 | | 0 | 0 |
| **Czech Republic** | 13 | 3 | 6 | 21 | 11 | 10,8 | 8 | **3** | 1,2281 | 8,2072 | 4,642 | 0,2687 | 1,7957 | 16,1417 | 8 | **3** | 0 | 0 |
| **Germany** | 16 | 23 | 11 | 7 | 22 | 15,8 | 19 | | 0,6947 | 0,1738 | 1,7957 | 3,839 | 0,2222 | 6,7254 | 17 | | 2 | 0 |
| **Russia** | 15 | 21 | 14 | 13 | 13 | 15,2 | 16 | | 0,84 | 0,2687 | 1,0157 | 1,2281 | 1,2281 | 4,5806 | 19 | | -3 | 0 |
| **Denmark** | 8 | 13 | 9 | 8 | 7 | 9 | 7 | **4** | 3,1748 | 1,2281 | 2,6256 | 3,1748 | 3,839 | 14,0423 | 10 | **1** | -3 | 3 |
| **San Marino** | 22 | 25 | 25 | 25 | 25 | 24,4 | 25 | | 0,2222 | 0,1257 | 0,1257 | 0,1257 | 0,1257 | 0,725 | 25 | | 0 | 0 |
| **North Macedonia** | 3 | 15 | 2 | 15 | 2 | 7,4 | 4 | **7** | 8,2072 | 0,84 | 9,924 | 0,84 | 9,924 | 29,7352 | 4 | **7** | 0 | 0 |
| **Sweden** | 5 | 7 | 3 | 4 | 3 | 4,4 | 2 | **10** | 5,6131 | 3,839 | 8,2072 | 6,7873 | 8,2072 | 32,6538 | 2 | **10** | 0 | 0 |
| **Slovenia** | 25 | 10 | 16 | 24 | 23 | 19,6 | 23 | | 0,1257 | 2,1714 | 0,6947 | 0,152 | 0,1738 | 3,3176 | 22 | | 1 | 0 |
| **Cyprus** | 9 | 9 | 4 | 17 | 17 | 11,2 | 10 | **1** | 2,6256 | 2,6256 | 6,7873 | 0,5745 | 0,5745 | 13,1875 | 11 | | -1 | 1 |
| **Netherlands** | 2 | 6 | 5 | 2 | 1 | 3,2 | 1 | **12** | 9,924 | 4,642 | 5,6131 | 9,924 | 12 | 42,1031 | 1 | **12** | 0 | 0 |
| **Greece** | 14 | 22 | 15 | 23 | 15 | 17,8 | 22 | | 1,0157 | 0,2222 | 0,84 | 0,1738 | 0,84 | 3,0917 | 23 | | -1 | 0 |
| **Israel** | 21 | 11 | 18 | 11 | 6 | 13,4 | 12 | | 0,2687 | 1,7957 | 0,4751 | 1,7957 | 4,642 | 8,9772 | 12 | | 0 | 0 |
| **Norway** | 23 | 20 | 24 | 18 | 14 | 19,8 | 24 | | 0,1738 | 0,3249 | 0,152 | 0,4751 | 1,0157 | 2,1415 | 24 | | 0 | 0 |
| **United Kingdom** | 20 | 18 | 23 | 6 | 20 | 17,4 | 21 | | 0,3249 | 0,4751 | 0,1738 | 4,642 | 0,3249 | 5,9407 | 18 | | 3 | 0 |
| **Iceland** | 10 | 1 | 1 | 20 | 24 | 11,2 | 10 | **1** | 2,1714 | 12 | 12 | 0,3249 | 0,152 | 26,6483 | 6 | **5** | 4 | -4 |
| **Estonia** | 12 | 17 | 19 | 12 | 18 | 15,6 | 17 | | 1,4851 | 0,5745 | 0,3929 | 1,4851 | 0,4751 | 4,4127 | 20 | | -3 | 0 |
| **Belarus** | 24 | 8 | 7 | 19 | 21 | 15,8 | 19 | | 0,152 | 3,1748 | 3,839 | 0,3929 | 0,2687 | 7,8274 | 14 | | 5 | 0 |
| **Azerbaijan** | 4 | 4 | 13 | 5 | 4 | 6 | 3 | **8** | 6,7873 | 6,7873 | 1,2281 | 5,6131 | 6,7873 | 27,2031 | 5 | **6** | -2 | 2 |
| **Italy** | 1 | 2 | 21 | 3 | 12 | 7,8 | 5 | **6** | 12 | 9,924 | 0,2687 | 8,2072 | 1,4851 | 31,885 | 3 | **8** | 2 | -2 |
| **Serbia** | 19 | 16 | 17 | 10 | 8 | 14 | 15 | | 0,3929 | 0,6947 | 0,5745 | 2,1714 | 3,1748 | 7,0083 | 16 | | -1 | 0 |
| **Switzerland** | 7 | 5 | 12 | 9 | 10 | 8,6 | 6 | **5** | 3,839 | 5,6131 | 1,4851 | 2,6256 | 2,1714 | 15,7342 | 9 | **2** | -3 | 3 |
| **Australia** | 18 | 24 | 10 | 1 | 5 | 11,6 | 11 | | 0,4751 | 0,152 | 2,1714 | 12 | 5,6131 | 20,4116 | 7 | **4** | 4 | -4 |
| **Spain** | 11 | 19 | 22 | 16 | 19 | 17,4 | 21 | | 1,7957 | 0,3929 | 0,2222 | 0,6947 | 0,3929 | 3,4984 | 21 | | 0 | 0 |
| **France** | | | | | | | | | | | | | | | | | | |

2019—it was tough to find the actual data (the mysterious *Green document* possibly forbids sharing of them). However, this dataset was found on an informal network (Reddit), and independently of the originality of the data, they will be good enough to demonstrate the exponential and linear weight models. In columns related to the linear weight model, we see how songs were ranked per each of the five French jurors. Ranks are further averaged (AV). Further, the songs were ranked (R) concerning averaged ranks (AV), and finally, the points (P) that would be given to the songs respecting the final ranks (R) if the linear model was used. The exponential weight model assigns fixed values to each position in a rank list, predefined by the EBU to follow an exponential function (e.g., first place = 12 points, second place = 9.924 points, third place = 8.207 points, fourth place = 6.787 points, and so on). Although the exact function is not found in the literature, we approximated it as $14.5591 \cdot 0.52664^{x}$.

In the case presented, ranks from linear and exponential weight models are highly correlated (the Spearman coefficient equals r=0.955). The exponential model gave Iceland and Australia 4 more points than the linear model, while Denmark lost 3 points. The EBU explained that the model change was due to significant ranking differences among jurors. Unlike the linear model, if four jurors rank a song first and one last, the exponential model could still award it 12 points. Conversely, a song ranked last by four jurors and first by one would be rated higher in the exponential model. We conducted a sensitivity analysis using data from Table 1, adjusting some votes. Iceland was ranked first by four jurors and last by one, while San Marino was ranked last by four jurors and first by one. Iceland ranked 1st in the exponential model and 2nd in the linear model; San Marino ranked 25th in the linear model and 11th in the exponential model. Iceland gained 2 points in the exponential model, and San Marino remained at 0 points.

In summary, the exponential model reduces the impact of an outlier juror when their ranking differs from the majority but increases it when they rank a song highly that others do not. However, the impact of "helping" some countries is limited as positions between 25th and 11th score 0 points.

# 5 Discussion

In this section, the goal is to identify the good sides and the weak points of the ESC voting system. The focus will be more on the weak points related to the jury voting system. The primary goal of the analysis is to emphasise aspects of the voting system that could be considered for improvement.

## 5.1 Audience voting system analysis

Regarding the audience voting system, some characteristics can be considered good and weak when observed from different perspectives. One is related to using the same set of points (1-8, 10, and 12) for each country regardless of the country size. While one can think that a bigger country has to have more points, others believe that equality in the set of points for all countries contributes to the equal importance of each nation, its culture, and its identity in the European environment. Another characteristic is related to plurality voting (voting only for a particular song(s)). While some can think that it is good to have the opportunity to vote only for some or one song, others think that each televoter should give their opinion on each song. In this direction, Ginsburgh and Moreno-Ternero (2023) endorse the use of the Shapley voting procedure for judges and tele-voters.

One of the weak points of audience voting, which will also be the case in jury voting, is related to the relative grading when distributing the points. As presented in section 3, the song that received the maximum number of votes receives 12 points, the second receives 10 points, and so on. In practice, it is possible that, for example, the first and the second song have tiny differences in terms of achieved votes, while in another case, this difference can be much higher, but in both cases, the points for the first and the second song will be 12 and 10. Probably the fairer system would be to distribute the total number of points (1+2+3+…+8+10+12=58) among the songs, respecting the relative share of votes. For example, this is a general idea in political voting in many countries. This idea will probably not be accepted, and one of the reasons is related to the ESC tradition—the system of points (1-8, 10, and 12) was introduced in 1975, which means that there are almost 50 years of tradition in this particular points-scale usage.

## 5.2 Jury voting system analysis

The good sides of the ESC jury voting system:

1. The organiser decided to deal with the only-audience voting system's problems, such as diaspora voting and voting for neighbors, by introducing juries that should objectively evaluate the songs and take care of the songs' quality,
2. After returning the juries to the ESC, the organiser is working on tracking the problems of the jury voting system and continuously upgrading the system with modifications that could result in fairer evaluation,
3. Jury members are selected among the experts in the music industry.

On the other hand, the jury voting system has some weak points.

1. The selection of the jury members is challenging, and it is questionable if only five members of the jury are enough for the demanding activity of song evaluation. Additionally, it is questionable whether the jury points from any country remain the same if some members or the complete jury are changed. It will be beneficial for the organiser to order

**Table 2.** Simulation of research that applies the control groups of jury members

| | 1 | 2 | 3 | 4 | 5 | SUM-1 | SUM-2 | SUM-3 | SUM-4 | SUM-5 | R-1 | R-2 | R-3 | R-4 | R-5 | P-1 | P-2 | P-3 | P-4 | P-5 | DIFF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Malta** | 17 | 14 | 8 | 22 | 9 | 7,038 | 6,5971 | 4,437 | 7,39055 | 4,9872 | 14 | 14 | 17 | 13 | 16 | | | | | | |
| **Albania** | 6 | 12 | 20 | 14 | 16 | 3,52 | 6,6773 | 7,8374 | 7,14667 | 7,46769 | 19 | 13 | 12 | 15 | 13 | | | | | | |
| **Czech Republic** | 13 | 3 | 6 | 21 | 11 | 14,91 | 7,9346 | 11,499 | 15,873 | 14,346 | 8 | 11 | 9 | 7 | 8 | 3 | 0 | 2 | 4 | 3 | 4 |
| **Germany** | 16 | 23 | 11 | 7 | 22 | 6,031 | 6,5516 | 4,9296 | 2,88637 | 6,50311 | 16 | 15 | 16 | 21 | 14 | | | | | | |
| **Russia** | 15 | 21 | 14 | 13 | 13 | 3,741 | 4,3119 | 3,5649 | 3,35248 | 3,35248 | 18 | 19 | 20 | 17 | 20 | | | | | | |
| **Denmark** | 8 | 13 | 9 | 8 | 7 | 10,87 | 12,814 | 11,416 | 10,8675 | 10,2034 | 10 | 8 | 10 | 10 | 11 | 1 | 3 | 1 | 1 | 0 | 2 |
| **San Marino** | 22 | 25 | 25 | 25 | 25 | 0,503 | 0,5993 | 0,5993 | 0,5993 | 0,5993 | 25 | 25 | 25 | 25 | 25 | | | | | | |
| **North Macedonia** | 3 | 15 | 2 | 15 | 2 | 21,53 | 28,895 | 19,811 | 28,8951 | 19,8111 | 4 | 2 | 5 | 2 | 6 | 7 | 10 | 6 | 10 | 5 | 5 |
| **Sweden** | 5 | 7 | 3 | 4 | 3 | 27,04 | 28,815 | 24,446 | 25,8664 | 24,4465 | 2 | 3 | 4 | 4 | 4 | 10 | 8 | 7 | 7 | 7 | 3 |
| **Slovenia** | 25 | 10 | 16 | 24 | 23 | 3,192 | 1,1461 | 2,6228 | 3,16547 | 3,14368 | 20 | 24 | 22 | 18 | 21 | | | | | | |
| **Cyprus** | 9 | 9 | 4 | 17 | 17 | 10,56 | 10,562 | 6,4001 | 12,6129 | 12,6129 | 11 | 9 | 14 | 9 | 10 | 0 | 2 | 0 | 2 | 1 | 2 |
| **Netherlands** | 2 | 6 | 5 | 2 | 1 | 32,18 | 37,461 | 36,490 | 32,1792 | 30,1032 | 1 | 1 | 1 | 1 | 2 | 12 | 12 | 12 | 12 | 10 | 2 |
| **Greece** | 14 | 22 | 15 | 23 | 15 | 2,076 | 2,8694 | 2,2516 | 2,91782 | 2,25164 | 22 | 22 | 23 | 20 | 23 | | | | | | |
| **Israel** | 21 | 11 | 18 | 11 | 6 | 8,709 | 7,1816 | 8,5021 | 7,18155 | 4,33522 | 12 | 12 | 11 | 14 | 17 | | | | | | |
| **Norway** | 23 | 20 | 24 | 18 | 14 | 1,968 | 1,8165 | 1,9891 | 1,66637 | 1,12579 | 23 | 23 | 24 | 23 | 24 | | | | | | |
| **United Kingdom** | 20 | 18 | 23 | 6 | 20 | 5,616 | 5,4657 | 5,7669 | 1,29873 | 5,61584 | 17 | 17 | 15 | 24 | 15 | | | | | | |
| **Iceland** | 10 | 1 | 1 | 20 | 24 | 24,48 | 14,648 | 14,648 | 26,3233 | 26,4963 | 3 | 7 | 7 | 3 | 3 | 8 | 4 | 4 | 8 | 8 | 4 |
| **Estonia** | 12 | 17 | 19 | 12 | 18 | 2,928 | 3,8381 | 4,0196 | 2,92751 | 3,93747 | 21 | 20 | 18 | 19 | 18 | | | | | | |
| **Belarus** | 24 | 8 | 7 | 19 | 21 | 7,675 | 4,6526 | 3,9884 | 7,4345 | 7,55869 | 13 | 18 | 19 | 12 | 12 | | | | | | |
| **Azerbaijan** | 4 | 4 | 13 | 5 | 4 | 20,42 | 20,416 | 25,975 | 21,5901 | 20,4159 | 5 | 5 | 3 | 6 | 5 | 6 | 6 | 8 | 5 | 6 | 3 |
| **Italy** | 1 | 2 | 21 | 3 | 12 | 19,88 | 21,961 | 31,616 | 23,6778 | 30,3999 | 7 | 4 | 2 | 5 | 1 | 4 | 7 | 10 | 6 | 12 | 8 |
| **Serbia** | 19 | 16 | 17 | 10 | 8 | 6,615 | 6,3136 | 6,4337 | 4,83685 | 3,83337 | 15 | 16 | 13 | 16 | 19 | | | | | | |
| **Switzerland** | 7 | 5 | 12 | 9 | 10 | 11,9 | 10,121 | 14,249 | 13,1085 | 13,5627 | 9 | 10 | 8 | 8 | 9 | 2 | 1 | 3 | 3 | 2 | 2 |
| **Australia** | 18 | 24 | 10 | 1 | 5 | 19,94 | 20,26 | 18,240 | 8,41154 | 14,7984 | 6 | 6 | 6 | 11 | 7 | 5 | 5 | 5 | 0 | 4 | 5 |
| **Spain** | 11 | 19 | 22 | 16 | 19 | 1,703 | 3,1055 | 3,276 | 2,80374 | 3,10549 | 24 | 21 | 21 | 22 | 22 | | | | | | |
| **France** | | | | | | | | | | | | | | | | | | | | | |

independent research, at least in some countries, that will simulate the voting process using one or more control groups. Achieving the results of this kind of research, that there is no difference in ranks and points between original jury and control juries, will help the organiser to argue the existence of a jury in the ESC voting model. Table 2 describes the idea of the independent research proposal: Let us assume that only four jury members are defined by the Green document (which is very close to five members, which was in reality). Since we have the ranks of all five jury members in Table 1, we can create five four-member juries that will differ in only one member. In Table 2, columns SUM-1 to SUM-5 present the sums of exponential values of five four-members (ex., SUM-1 means the sum of exponential values of a four-member jury consisting of jurors 2-5). Further, columns R-1 to R-5 present the ranks of each of the five four-member juries. Finally, columns P-1 to P-5 present the points for different countries of each of the five four-member juries. Consequently, we see maximal differences (column DIFF) in all countries that receive some points (the differences are between 2 and 8 points).

Especially this is emphasised in the case of North Macedonia, Australia, and Italy. In the cases of North Macedonia and Australia, the final jury points differ by five points. In the case of Italy, the difference is 8 points.

a. Additionally, Spearman rank correlations were calculated to examine the correlations among every two sets of ranks respecting 11 countries receiving the points, and coefficients were between 0.63 and 0.84 which are low values considering the context.
b. Furthermore, we calculated the sums of absolute differences in the number of points between each two juries. The sums vary from 16 (the differences between jury-3 and jury-5) to 26 (the differences between jury-2 and jury-5).

Even though in this simulation, we had one jury member less than in the actual case, when comparing each two juries, they differ in only one member—results indicate that significant differences in the final jury points can appear just in the case of one country—what to expect if, in the control groups, there would be the juries with entirely different members?

2. Not publishing the *Green document* contributes to the dissatisfaction and unacceptance of jury inclusion in the ESC voting system.

3. When evaluating the songs respecting the four criteria, if the jurors have to evaluate each song concerning all criteria simultaneously, then there is a potential problem with the trustworthiness of the decision-making process. The evaluation of the songs, concerning the four criteria, is a multi-criteria decision-making problem. According to Saaty and Ergu (2015), a song evaluation method could be evaluated as *low,* respecting many of the criteria defined for methods evaluation, which does not consequently ensure the used method is suitable for evaluating the song quality.
4. Additionally, when discussing the concept of quality, a complex construct, it is unknown how good, deep, and precise the concept of quality in terms of music is defined in jury members' instructions in the voting process.
5. Like in the case of the audience's voting system, the absolute differences among the songs are not considered on both the level of each juror and the level of the jury. Additionally, if certain jurors find two songs equal in terms of quality, they must give an advantage to one song, which is opposite to the concept of objective song evaluation.
    a. The song that is ranked the first for each juror will receive a value of 12 according to the exponential weight model; the second one will receive a value of 9,924, and so on, even though the absolute difference in quality between any two songs does not have to correspond to the fixed difference of exponential weight model values.
    b. At the level of the whole jury, the points are associated with countries respecting the rank of the sum of exponential values of all five jury members, not based on the relative share of country sum in all countries' sum.

    Consequently, we have a double relativisation of jury voting in the system.
6. While the goal of the exponential weight model is argued, the organiser does not argue with the selection of exponential weight model values for each rank at the level of a juror.
7. The organiser also does not argue that there is an almost equal influence of jury and public votes in the voting system. In the case of ESC 2024, there were 185 judges and millions of public voters.
8. The exponential weight model is introduced to deal with outlier opinions in the jury. The aim is "*to place the judgment of a group of jurors above the opinion of a single individual juror*". Why is someone's right to have the outlier opinion decreased? What if the "problem" is in four other judges?
9. In informal circles, many mention lobbying at the ESC. Since the number of judges is finite and not so high, it is potentially possible to influence their opinion indirectly.
10. It is interesting to analyse the similarity between jury votes among different countries. This analysis is related to the ongoing research currently in the phase of paper review (partly same authors). That research concluded that the public points from different countries are more similar than the jury points. The question that consequently arises is—if the juries are returned to the competition to ensure objectivity, how do they think differently about the quality? On the other hand, if the public cannot evaluate the quality, why do they have better similarity among themselves?

# 6 Conclusion

The answer to the question from the title is simple—when public and jury points are summed, Croatan Baby Lasagna achieved the second-best score and consequently did not win the ESC 2024. However, the purpose of the title question is not to ask simple questions but to tackle the readers as well in the scientific community (and there are many fans of Baby Lasagna) to critically reason on the voting system characteristics to get united in brainstorming the proposals that can potentially be useful for voting system improvement.

This analysis opened the discussion on several aspects of the ESC voting system, but many aspects were not even touched. However, two aspects are imposed as the most critical questions:

1. How is it possible to create a valid jury (considering the analysis presented in Table 2)? The preliminary simulation shows (too) high differences among four-member juries differing in only one member at the level of only one country.
2. How can millions be united when they must agree with their only fifth option?

Further, many other shortcomings of the jury voting system presented in this research need the organiser's attention. The main conclusion is that the voting system still needs significant upgrades. Maybe the best conclusion summarising the whole analysis is the one by Baby Lasagna: "*The audience will come to my concert, not the jury"*. Indeed, Baby Lasagna sold out many concerts after the ESC 2024, which is not the case with the ESC winner.

There are two main limitations of this research. The first relates to the subjectivity of the paper's author, who is a fan of the Baby Lasagna. However, a great effort is put into objectively arguing and elaborating all the conclusions regarding the good and weak points of the ESC voting system. The second limitation is related to the unavailability of some data related to details of the whole voting system process and motives that support the voting system being designed in the way it is designed. Consequently, some reasoning is based on the data available on different forums and discussion groups, which are not official data and might be incorrect.

## Acknowledgment

Thanks to Baby Lasagna for being a motivation for this paper and a role model for many life aspects.

# References

Abakoumkin, G. (2018). "Play it, Sam", again and again: Further instances of familiarity effects in the Eurovision Song Contest. *Hellenic Journal of Psychology*, *15*(2), 125–137.

Alpatova, O. E. (2022). 'A Song for Europe': Music or Politics? *Journal of Globalization Studies*, *13*(2), 105–118. https://doi.org/10.30884/jogs/2022.02.07

Baker, C. (2015). Introduction : gender and geopolitics in the Eurovision Song Contest. *Contemporary Southeastern Europe*, *2*(1), 74–93.

Budzinski, O., Gaenssle, S., & Weimar, D. (2023). Disentangling Individual Biases in Jury Voting: An Empirical Analysis of Voting Behavior in the Eurovision Song Contest. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4343866

Budzinski, O., & Pannicke, J. (2017). Culturally biased voting in the Eurovision Song Contest: Do national contests differ? *Journal of Cultural Economics*, *41*(4), 343–378. https://doi.org/10.1007/s10824-016-9277-6

Carniel, J. (2015). Skirting the issue: finding queer and geopolitical belonging at the Eurovision Song Contest. *Contemporary Southeastern Europe*, *2*(1), 136–154.

Chhan, D., Novoseller, E., & Lawhern, V. J. (2024). *Crowd-PrefRL: Preference-Based Reward Learning from Crowds*. https://doi.org/10.48550/arXiv.2401.10941

Clerides, S., & Stengos, T. (2006). Love thy Neighbor, Love thy Kin: Voting Biases in the Eurovision Song Contest. In *https://EconPapers.repec.org/RePEc:ucy:cypeua:1-2006*.

Coupe, T., & Chaban, N. (2020). Creating Europe through culture? The impact of the European Song Contest on European identity. *Empirica*, *47*(4), 885–908. https://doi.org/10.1007/s10663-019-09461-6

Dekker, A. H. (2007). The Eurovision Song Contest as a "Friendship" Network. *Connections*, *27*(3), 53–58.

Dodevska, Z. A., Kovacevic, A., Vukicevic, M., & Delibašić, B. (2020). Two Sides of Collective Decision Making - Votes from Crowd and Knowledge from Experts. *Lecture Notes in Business Information Processing*, *384 LNBIP*, 3–14. https://doi.org/10.1007/978-3-030-46224-6_1

Dogru, B. (2013). Modeling Voting Behavior in the Eurovision Song Contest. In *Munich Personal RePEc Archive*.

European Broadcast Union. (n.d.). *Frequently asked questions #Why allow Juries to vote in the Grand Final if their votes are no longer used in the Semi-Finals?* . Eurovosion.TV . Retrieved April 9, 2024, from https://eurovision.tv/voting-changes-2023-faq

European Broadcasting Union. (2024). *Eurovision Song Contest - Official website*. Eurivision Song Contest .

Ganser, C., & Keuschnigg, M. (2018). SOCIAL INFLUENCE STRENGTHENS CROWD WISDOM under VOTING. *Advances in Complex Systems*, *21*(6–7). https://doi.org/10.1142/S0219525918500133

Ginsburgh, V., & Moreno-Ternero, J. D. (2023). The Eurovision Song Contest: voting rules, biases and rationality. *Journal of Cultural Economics*, *47*(2), 247–277. https://doi.org/10.1007/s10824-022-09456-5

Ginsburgh, V., & Noury, A. G. (2008). The Eurovision Song Contest. Is voting political or cultural? *European Journal of Political Economy*, *24*(1), 41–52. https://doi.org/10.1016/j.ejpoleco.2007.05.004

Haan, M. A., Dijkstra, S. G., & Dijkstra, P. T. (2005). Expert Judgment Versus Public Opinion ? Evidence from the Eurovision Song Contest. *Journal of Cultural Economics*, *29*(1), 59–78. https://doi.org/10.1007/s10824-005-6830-0

HAQBEEN, J., SAHAB, S., & ITO, T. (2022). Evaluating Rank-Coherence among AI-enabled Ranking, Expert Rating and Crowd Voting for Selecting Winners in Incentivized Large-scale Idea Contest for Creative Works. *Proceedings of the Annual Conference of JSAI 200*, 1–8. https://doi.org/https://doi.org/10.11517/pjsai.JSAI2022.0_1S1IS303

Kovacevic, A., Vukicevic, M., Radovanovic, S., & Delibasic, B. (2020). CrEx-Wisdom Framework for Fusion of Crowd and Experts in Crowd Voting Environment – Machine Learning Approach. *Communications in Computer and Information Science*, *1260 CCIS*, 131–144. https://doi.org/10.1007/978-3-030-55814-7_11

Kumpulainen, I., Praks, E., Korhonen, T., Ni, A., Rissanen, V., & Vankka, J. (2020). Predicting Eurovision Song Contest Results Using Sentiment Analysis. *Communications in Computer and Information Science*, *1292 CCIS*, 87–108. https://doi.org/10.1007/978-3-030-59082-6_7

Mantzaris, A. V, Rein, S. R., & Hopkins, A. D. (2017). *Examining collusion and voting biases between countries during the Eurovision song contest since 1957*. https://doi.org/10.18564/jasss.3580

Michiel Vos. (n.d.). *Eurovision: Does the voting system with juries need to change?* A Bit of Pop Music. Retrieved April 9, 2024, from https://abitofpopmusic.com/2019/05/24/eurovision-does-the-voting-system-with-juries-need-to-change/

Saaty, T. L., & Ergu, D. (2015). When is a Decision-Making Method Trustworthy? Criteria for Evaluating Multi-Criteria Decision-Making

Methods. *International Journal of Information Technology & Decision Making*, *14*(06), 1171–1187. https://doi.org/10.1142/S021962201550025X

Siganos, A., & Tabner, I. T. (2020). Capturing the role of societal affinity in cross-border mergers with the Eurovision Song Contest. *Journal of International Business Studies*, *51*(2), 263–273. https://doi.org/10.1057/s41267-019-00271-3

Stockemer, D., Blais, A., Kostelka, F., & Chhim, C. (2018). Voting in the Eurovision Song Contest. *Politics*, *38*(4), 428–442. https://doi.org/10.1177/0263395717737887

Umair, A., Masciari, E., Madeo, G., & Habib Ullah, M. (2022). Applications of Majority Judgement for Winner Selection in Eurovision Song Contest. *ACM International Conference Proceeding Series*, 113–119. https://doi.org/10.1145/3548785.3548791

Verrier, D. B. (2012). Evidence for the influence of the mere-exposure effect on voting in the Eurovision Song Contest. *Judgment and Decision Making*, *7*(5), 639–643. https://doi.org/10.1017/S1930297500006355

Vukicevic, A., Vukicevic, M., Radovanovic, S., & Delibasic, B. (2022). BargCrEx: A System for Bargaining Based Aggregation of Crowd and Expert Opinions in Crowdsourcing. *Group Decision and Negotiation*, *31*(4), 789–818. https://doi.org/10.1007/s10726-022-09783-0

Wikipedia authors. (2024). *Voting at Eurovision Song contest*. Wikipedia. https://en.wikipedia.org/wiki/Voting_at_the_Eurovision_Song_Contest

# Bridging Theory and Practice: Enhancing Business Decision Making Course through Project-Based and Work-Based Learning

**Jelena Gusić Munđar, Tena Jagačić, Nikola Kadoić**
University of Zagreb Faculty of Organisation and Informatics
Pavlinska 2, Varaždin
{jelena.gusic, tjagacic, nkadoic}@foi.unizg.hr

***Abstract****. This paper presents the project-based (PBL) and work-based learning (WBL) implementation within Business Decision Making course over the past four years. The implementation of the activity differed in each year considering the number of organisations included, who selected the organisations and defined the problems in the activity, and the mode of course delivery. Achieved levels of generic skills were self-evaluated by students. Additionally, course grades were also analysed. Analytical thinking involving mathematical skills and the ability to work in a team were the lowest and the highest achieved competencies respecting all four years. Better results in generic competencies skills were achieved when the problems were posed by teachers. Higher self-evaluated levels of skills' achievements are not always followed by higher exams and course overall results. However, higher activity grades were positively correlated with exams and course overall results. Consequently, it can be concluded that some students overrated the achievements of self-evaluated skills. Finally, the paper proposes how to design the WBL/PBL activity in the future.*

**Keywords.** project-based learning, work-based learning, Business Decision Making, learning analytics, ANOVA, correlation, Kruskal-Wallis test

## 1 Introduction

In a world of rapid change, higher education retains its role as a catalyst for a well-educated workforce. One of the prerequisites for quality learning and teaching for future professions is the application of innovative pedagogies. Innovative pedagogies that are becoming increasingly prevalent in teaching include interactive lessons, using virtual reality technology, using AI, blended learning, project-based learning (PBL), problem-based learning, work-based learning (WBL), inquiry-based learning, flipped classroom, and experiential learning (Khamitova, 2023; Tran, 2024).

Successful collaboration with the private sector has enabled the Faculty of Organization and Informatics to actively implement PBL and WBL in the teaching process. The course Business Decision Making (BDM) is implemented in the sixth semester of the Entrepreneurship Economics study program. The course provides a comprehensive understanding of key concepts of business decision-making, decision-making theories that form the basis for decision-making processes, and the application of various decision-making approaches, preparing students for the complex challenges of decision-making in the business world. Students in seminar classes apply theoretical knowledge of methods and techniques. The seminar classes are designed for students to solve real challenges and case studies through discussions, small group work, and individual assignments. In the scope of this course, students participate in a PBL/WBL activity related to analysing different problems. This paper describes how this activity was implemented during the past four academic years (each year differently) and compares the achieved results concerning the acquired competencies and grades. The paper's goal is to evaluate different variants of PBL/WBL applied to the course. The purpose of the evaluation can be observed in the light of learning analytics (LA) because the idea is to conclude the design of PBL/WBL activities that are the most successful in practice. To track the success of those designs, two aspects are covered: (1) development of student generic competencies and (2) achieved grades.

The paper is organised as follows: Section two presents the literature review related to PBL and WBL. Section three presents the research methodology. Section 4 presents the results of the research. Finally, section 5 concludes the paper.

## 2 Literature review on PBL/WBL

Many innovative teaching and learning strategies can be used to increase the quality of the education process. Their choice depends on several factors. These factors include the subject, learning outcomes, characteristics of students, knowledge of educators to implement certain innovative teaching and learning strategies,

practical possibilities to implement certain innovative teaching and learning strategies and others. This section focuses on work-based and project-based learning.

The work-based learning (WBL) includes employers in the education process. WBL is often a very successful approach. It incorporates both theoretical and practical aspects of learning which ensures a high comprehension of the education process. There are many variants of how employers can be included in academic education. Some of them include student internship (the most advanced form), job shadowing, providing real case studies for problem-solving assignments, and participating in the evaluation and assessment phase.

The purpose of the WBL is to increase the effects of studying. The positive effects on students' achievement motivation were identified in research conducted in Indonesia. The authors concluded that WBL can improve the quality of the education process. (Sudjimat & Permadi, 2019). Additionally, WBL can improve students' soft skills. (Dogara et al., 2020)

WBL is also called workplace learning and work-integrated learning. Many scientific papers describe good and bad practices in WBL, analyse students' motivation, how the education process is organised, what the benefits of WBL are, and what are the issues. Listed topics were also the focus of the systematic literature review conducted in 2020. The issues faced by faculties and universities were related to time planning, different expectations from WBL by academia and industry, a large amount of resources needed to implement WBL, and students' skills and well-being. (Murtazin et al., 2020) In the same year, a systematic review of the literature resulted in the identification of WBL features and educational leadership features in the journal articles 2015-2020. Further, the authors analysed the types of connections between WBL and educational leadership. The highest number of articles can be associated with the collaboration as a type of connection between WBL and educational leadership indicating the need for collaboration among students and educators with industry and other relevant stakeholders. (Sudirman & Gemilang, 2020)

More recently, in 2023, another systematic literature review was conducted focusing the conducting the WBL in an online environment. The authors were dealing with the identification of WBL typologies and the effectiveness of WBL. They concluded that in some case studies, the function of technology is supportive, while in others the function of technology is delivering. Additionally, they analysed the case studies in the light of Schuster and Glavas classification. In terms of WBL effectiveness, there is a lack of robust evidence. (Rienties et al., 2023) Analysing WBL in an online environment was also analysed in the literature, especially in the light of Covid-19 pandemic. The author concluded that Covid-19 speeded up the online WBL. However, the theoretical and practical of WBL should be connected better. (Lester & Crawford–Lee, 2023)

There are two study programs types, university (scientific) and professional study programs. Professional study programs are more practice-oriented and consequently, more suitable for intensive WBL application. That was exactly the focus of the research related to the scientific nature of WBL. The authors came up with "five first principles or fundamental conditions of scientific inquiry ". They conclude if the scientific inquiry is defined as narrow, then professional programs can't be considered scientific. But, if the definition is broad, then the conducted research, which included WBL analysis, suggests that professional study programs are scientific. (Fergusson et al., 2020)

Project-based learning (PBL) is related to posing and solving different kinds of problems in different fields. Employers now do not have to be included in the education process, but it is possible to include them and witness the application of both strategies, PBL and WBL. Applying the PBL contributes to the development of both, technical and generic skills that are crucial in social science, for example in the field of marketing (Rohm et al., 2021). Additionally, PBL can enhance students' learning, increase their motivation and success (Maros et al., 2023; Ngereja et al., 2020; Rozal et al., 2021), improve students' self-efficacy (Krsmanovic, 2021) and boost their higher-order thinking skills (Eliyasni et al., 2019).

Similarly to WBL, PBL played an important role during the Covid-19 pandemic. This is presented in the research that analysed the interviews with teachers in the United States. Results show that PBL improved individual learning experiences, tackled students to take responsibility for their learning, upgraded communication and collaboration between students and educators, and promoted technology (Hira & Anderson, 2021). Another research among teachers resulted in a structural equation model that confirmed that PBL improved student engagement by enabling discussions (Almulla, 2020). A student perspective on PBL is also analysed in the literature. One of the common student-related problems in PBL is collaboration among students during the project implementation. (Hussein, 2021).

There are different approaches available for students' knowledge and produced artifact assessment. Assessment is one of three components of constructive alignment. Both PBL and WBL present the polygon for diverse (formative and summative) assessment approach applications. Different authors presented sets of assessment activities implemented with PBL in engineering courses (Cifrian et al., 2020) and business decision making (Kadoic & Slibar, 2020). Further, PBL can be combined with another innovative pedagogy and there is some evidence of its combination with flipped classrooms (Chis et al., 2018; Hossein-Mohand et al., 2021; Shih & Tsai, 2016).

A systematic literature review on PBL research resulted in 76 papers that were focused on cognitive aspects of teaching and learning, affective aspects, behavioural aspects, and artefact performance. The authors also gave recommendations for how future studies in PBL should be designed. Some of them are related to presenting a more comprehensive project evaluation, giving more details on methodologies applied, and enabling more details on reporting and validity of developed instruments. (Guo et al., 2020)

The success of different WBL and PBL variants varies depending on different aspects. Our research evaluates the success of WBL/PBL variants by considering the level of generic competencies students achieve and their grades.

# 3 Methodology

This section describes the application of PBL and WBL in the BDM course, the research questions, and the methods applied to answer the research question.

## 3.1 Application of PBL and WBL in BDM

As a part of the BDM course, a team activity titled "Development of Project Idea" has been applied for over ten years. This paper focuses on analysing the implementation of activities in the last four academic years. The implementation of the activity differed in each of the four years.

In this activity, students practically address the business challenges of public organisations, non-profit organisations, and private companies from their immediate environment. In the first observed year, all assigned challenges were related to an association involving counselling, education, promotion, and improvement related to individuals with various impairments and disabilities. In collaboration with responsible persons from the association, seven topics were prepared for the team members to work on, each being addressed by two teams of students.

The elaboration of the topic involved the prescribed elements of the PrOACT approach (Hammond et al., 1999), including:

1. analysis of the business case problem,
2. identification of criteria for evaluating created alternatives,
3. creation of alternative solutions to the problem,
4. identification of the consequences of individual alternative solutions and selection of the best alternative solution,
5. financial and time plan for the implementation of the chosen alternative solution,
6. analysis of the risks of applying the chosen alternative solution,
7. ethical background of the analysed problem and proposed solution, and
8. marketing aspects of the chosen solution.

**Table 1.** The comparison of the activity implementation over four years

| | Number of organisations | Who selected organisations in PBL/WBL | Who defined topics for projects | Teaching process |
|---|---|---|---|---|
| 2021 | 1 | Teachers | Organisations | Online |
| 2022 | 5 | Teachers | Teachers and organisations | F2f |
| 2023 | 31 | Students | Students | F2f |
| 2024 | 1 | Teachers | Teachers | F2f |

During the practical handling of the assigned topic, students actively communicated with responsible persons from the association. The final solutions were presented online, as epidemiological measures due to COVID-19 were in effect during that period. The project evaluation was done by teachers and carried the highest percentage in the overall grade, followed by evaluations from the person responsible for the association and evaluations from student colleagues. In the second year of the observed topic, challenges included family agricultural business, home crafts, transportation company, private clinic, and cultural arts society, totalling eight topics. For the needs of project activities in the third observed year, students themselves researched and negotiated challenges, so there were no recurring topics. The challenges included organisations from the public and private sectors, totalling 31. In the fourth year, the challenges were related to the needs and projects of the faculty and the banking sector. Solutions were physically presented at the university in the second, third, and fourth years. The mentioned course of project activity was the same during all observed years (FOI, n.d.).

## 3.2 Research questions and methods

After the "Development of Project Idea" activity had been implemented, the students participated in a survey related to self-evaluation on the level at which they achieved ten generic competencies. The competencies are presented in Table 2. The survey consisted of 10 closed questions related to the generic competencies and two open questions related to the positive and negative aspects of the activity implementation. The generic competencies used in this research were previously used in different research by researchers in the Center for Student Support and Career Development (University of Zagreb Faculty of Organization and Informatics), and in (Anicic & Buselic, 2021; Pažur Aničić et al., 2023). The closed questions were quantitatively defined, describing the level at which particular competence was achieved. Five levels (from 0 to 4) describe each competence

achievement. The students participated in the research voluntarily.

The answers to the survey from the past four years, together with the grades, are analysed in this paper to answer the following research questions:

- RQ1: What were the average levels of achievement of each competence per year? Is there a significant difference among the competencies' achievements in different years?
- RQ2: Do those who are defining the project topic influence the improvement of students' generic competencies?
- RQ3: What were the average grades of students per year? Is there a significant difference among the students in different years?
- RQ4: In the years when they improve their generic competencies, do the students notice improvements in their project and exam scores?
- RQ5: Are the project scores and midterm exams scores correlated, and how has the correlation changed through time?

For comparing the improvement of students' generic competencies over time, the Kruskal-Wallis test was performed because the assumptions for the one-way analysis of variance (ANOVA) were not satisfied. If there were a significant difference in the Kruskal-Wallis test, then post hoc tests were done to find where the difference exists. For this purpose, the Dunn test with multiple testing corrections was performed (p-values adjusted with the Holm method). Pearson correlations between project and midterm exam scores were calculated and compared over four years. For comparing students' project and exam scores over four years, ANOVA with post hoc tests (Tukey HSD test with adjusted p-values) were performed because the assumptions for using ANOVA were satisfied.

# 4 Results

The answers to the RQ1 can be found in Table 2.

**Table 2.** Comparison of generic competencies over four years

| Ability to quickly acquire new knowledge (Brzo_OCJ) | | | | |
|---|---|---|---|---|
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 2.88 | 0.98 | 0.4785 |
| 2022 | 21 | 2.90 | 1.04 | |
| 2023 | 38 | 2.61 | 0.95 | |
| 2024 | 50 | 2.80 | 0.99 | |
| Ability to identify and solve problems (IRP_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.22 | 0.822 | 0.0081* |
| 2022 | 21 | 3.24 | 0.889 | |
| 2023 | 38 | 2.66 | 0.909 | |
| 2024 | 50 | 2.98 | 0.869 | |
| Ability to apply knowledge to practical problems (PRIMJENA_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.14 | 0.862 | 0.0997 |
| 2022 | 21 | 3.00 | 0.894 | |
| 2023 | 38 | 2.68 | 0.989 | |
| 2024 | 50 | 2.94 | 0.890 | |
| Ability to work in a team (TIM_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.33 | 0.894 | 0.1688 |
| 2022 | 21 | 3.48 | 0.512 | |
| 2023 | 38 | 2.95 | 1.06 | |
| 2024 | 50 | 3.34 | 0.848 | |
| Analytical thinking involving mathematical skills (AR_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 2.73 | 0.949 | 0.0002* |
| 2022 | 21 | 2.62 | 1.02 | |
| 2023 | 38 | 2.21 | 1.02 | |
| 2024 | 50 | 1.9 | 1.22 | |
| Responsibility in work and evaluation of the quality of your own work (OCG_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.33 | 0.822 | 0.0711 |
| 2022 | 21 | 3.33 | 0.730 | |
| 2023 | 38 | 2.97 | 0.915 | |
| 2024 | 50 | 2.96 | 1.03 | |
| Ability to adapt and act in new situations (PRIL_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.17 | 0.755 | 0.0026* |
| 2022 | 21 | 3.19 | 0.814 | |
| 2023 | 38 | 2.53 | 1.06 | |
| 2024 | 50 | 2.70 | 1.02 | |
| Capacities for generating new ideas (GEN_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.19 | 0.823 | 0.0983 |
| 2022 | 21 | 3.14 | 0.910 | |
| 2023 | 38 | 2.68 | 1.07 | |
| 2024 | 50 | 3.08 | 0.877 | |
| Ability to work under pressure (PRIT_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 2.79 | 1.01 | 0.0275* |
| 2022 | 21 | 3.05 | 0.92 | |
| 2023 | 38 | 2.29 | 1.14 | |
| 2024 | 50 | 2.48 | 1.11 | |
| Independence in work (SAM_OCJ) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.28 | 0.84 | 6.291e-05* |
| 2022 | 21 | 3.05 | 1.12 | |
| 2023 | 38 | 2.58 | 1.13 | |
| 2024 | 50 | 2.40 | 1.20 | |
| Average grade | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 81 | 3.11 | 0.6 | 0.0008* |
| 2022 | 21 | 3.1 | 0.688 | |
| 2023 | 38 | 2.62 | 0.752 | |
| 2024 | 50 | 2.76 | 0.672 | |

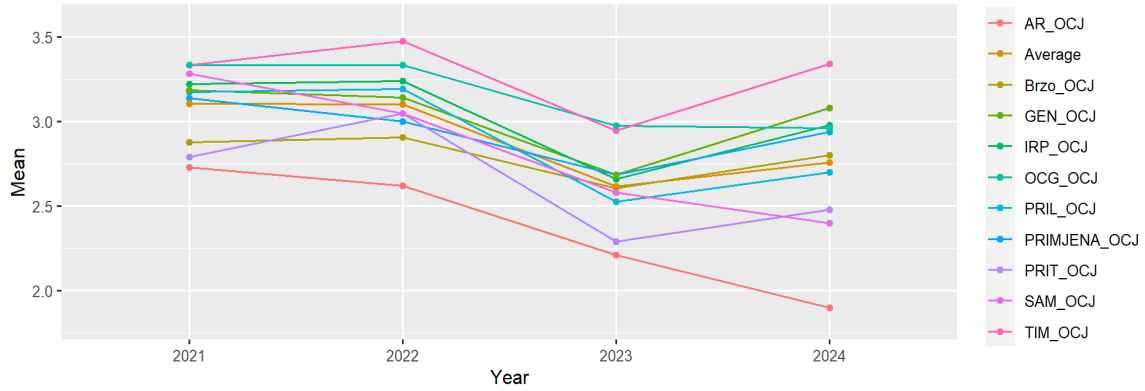



**Figure 1.** Comparison of generic competencies over four years over four years

On average, the levels of competencies' achievements vary among both years and competencies.

The lowest average level of competence achievement was 1.9, which was achieved in 2024 in relation to analytical thinking involving mathematical skills. The highest level of competence achievement was 3.48, which was achieved in 2022, related to the ability to work in a team. Additionally, analytical thinking involving mathematical skills and the ability to work in a team were the lowest and the highest achieved competencies respecting all four years. A detailed analysis of the competencies' achievement per year and each competence is presented in Table 2.

In terms of statistically significant differences among the competencies' achievements in different years, they are identified in several competencies. They are the ability to identify and solve problems (2021-2023 p.adj=0.008, 2022-2023 p.adj=0.070), analytical thinking involving mathematical skills (2021-2023 p.adj= 0.051, 2021-2024 p.adj=0.00019), the ability to adapt and act in new situations (2021-2023 p.adj=0.0102, 2021-2024 p.adj=0.042), the ability to work under pressure (p-value= 0.02745<0.05; 2022-2023 p.adj= 0.074189), and independence in work (2021-2023 p.adj=0.006778, 2021-2024 p.adj= 0.00013). A similar conclusion is for the average improvement after WBL/PBL activity (2021-2023 p.adj= 0.00314982, 2021-2024 p.adj=0.02086040). The average improvement for all generic competencies over four years is represented in Fig. 1.

Now, RQ2 can be answered. From Table 1 and Fig. 1, it is clear that in 2023, when students defined the topics of their projects, the lowest levels of competencies achievements were achieved in the majority of competencies.

To answer the RQ3, Table 3 is prepared. According to it, 84 students enrolled in the course in 2021, 92 in 2022, 83 in 2023, and 56 in 2024. Also, for all categories, students' results statistically significantly changed over time, and students had the best grades in most categories in 2024 (except PROC). Points for assessment (PROC) that students achieved statistically significantly differ for 2023 and 2022 (p.adj= 0.0168424), and on average, students had the best results in 2023 for this grading category. Achieved points for report (RAD) were statistically significantly greater in 2024 than in other years (2024-2021 p.adj=0, 2024-2022 p.adj=0.0000001, 2024-2023 p.adj=0) and also lower in 2021 than in other observed years (2022-2021 p.adj=0.0000076, 2023-2021 p.adj=0.0059596).

**Table 3.** Comparison of grading components

| Points for assessment (PROC) (4) | | | | |
|---|---|---|---|---|
| | n | $\bar{x}$ | SD | p value |
| 2021 | 84 | 2.27 | 0.33 | 0.0156 * |
| 2022 | 92 | 2.24 | 0.376 | |
| 2023 | 83 | 2.40 | 0.427 | |
| 2024 | 56 | 2.25 | 0.294 | |
| Points for report (RAD) (16) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 84 | 12.3 | 1.67 | <2e-16 * |
| 2022 | 92 | 13.6 | 1.82 | |
| 2023 | 83 | 13.2 | 1.54 | |
| 2024 | 56 | 15.2 | 1.63 | |
| Midterm exam 1 (K1) (35) | | | | |
| | N | $\bar{x}$ | SD | p value |
| 2021 | 84 | 15.3 | 5.07 | 2.53e-08 * |
| 2022 | 92 | 16.7 | 5.27 | |
| 2023 | 83 | 17.0 | 3.84 | |
| 2024 | 56 | 20.5 | 5.07 | |
| Midterm exam 2 (K2) (35) | | | | |
| | N | $\bar{x}$ | SD | p value |
| 2021 | 84 | 16.9 | 5.29 | 6.82e-12 * |
| 2022 | 92 | 13.5 | 4.28 | |
| 2023 | 83 | 16.4 | 5.36 | |
| 2024 | 56 | 20.4 | 6.54 | |
| K1+K2 (70) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 84 | 32.2 | 9.37 | 1.01e-11 * |
| 2022 | 92 | 30.2 | 7.25 | |
| 2023 | 83 | 34.7 | 7.26 | |
| 2024 | 56 | 40.9 | 10.6 | |
| Total number of points (Bodovi) (100) | | | | |
| | n | $\bar{x}$ | SD | p value |
| 2021 | 84 | 59.0 | 10.5 | 3.3e-05 * |
| 2022 | 92 | 59.1 | 10.2 | |
| 2023 | 83 | 54.0 | 8.4 | |
| 2024 | 56 | 62.2 | 11.6 | |

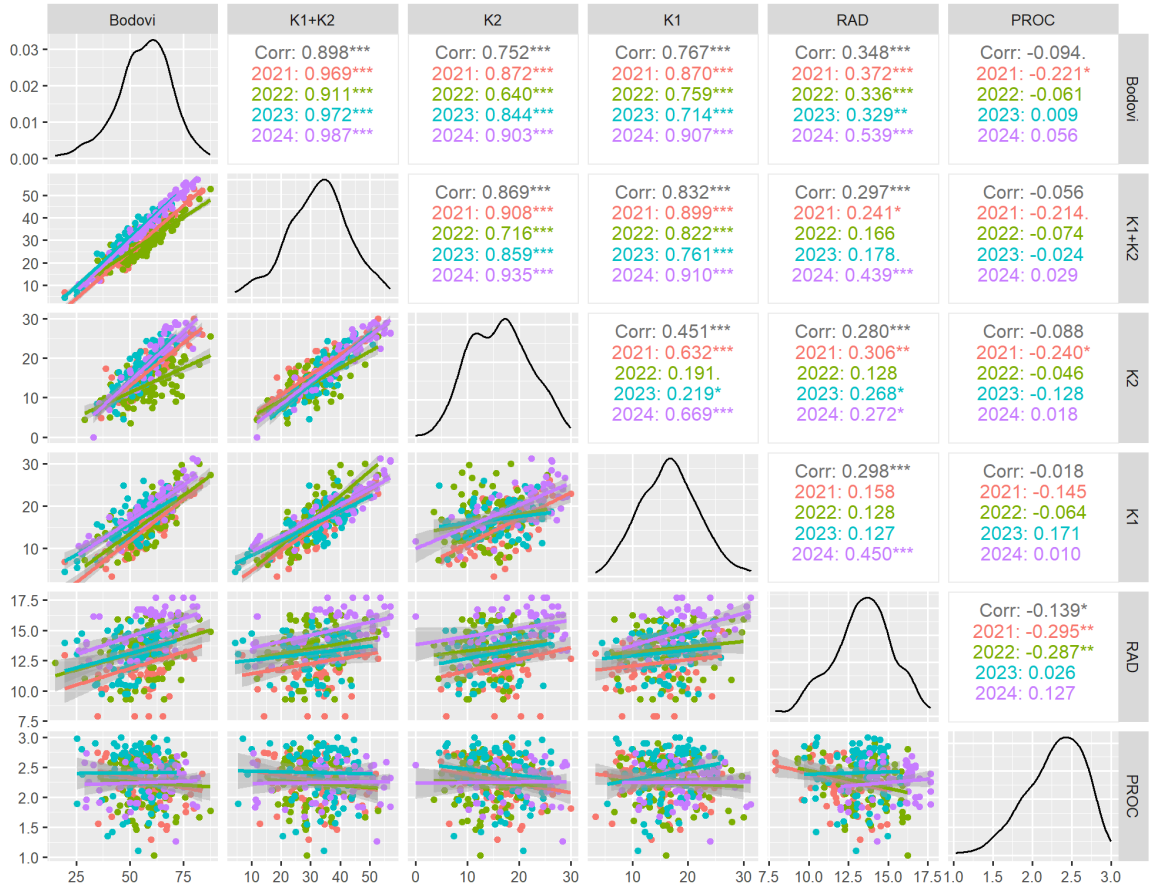



**Figure 2.** Correlations between grading components over four years

Also, the number of points that students achieved on midterm exam 1 (K1) (2024-2021 p.adj=0, 2024-2022 p.adj= 0.0000279, 2024-2023 p.adj=0.0001921), midterm exam 2 (K2) (2024-2021 p.adj=0.0011726, 2024-2022 p.adj=0, 2024-2023 p.adj=0.0001276), and total number of points on midterm exams (K1+ K2) (2024-2021 p.adj=0.0000001, 2024-2022 p.adj=0, 2024-2023 p.adj=0.0001276), were statistically significant greater in 2024 than in other observed years. Achieved points for K2 were statistically significantly lower in 2022 than in other years (2022-2021 p.adj=0.0001277, 2023-2022 p.adj=0.0017179), but the total number of points on midterm exams was higher in 2022 than in 2023 (2023-2022 p.adj=0.0027084). The total number of achieved points (Bodovi) statistically significantly differs for the following years: 2021-2023 p.adj=0.0076733, 2022-2023 p.adj=0.0054387, and 2023-2024 p.adj=0.0000254.

To answer the RQ4, the results from Table 1 and Table 2 have to be combined. In the years when they improve their generic competencies, students notice improvements in their project and exam scores in 2021-2023, but not in 2024.

Correlation analysis was implemented to answer the RQ5. Fig. 2 shows that the total number of points exhibited statistically significant positive Pearson's correlation with all other grading components except points for assessment (PROC) over four years. The total number of points and points for assessment are negatively correlated, especially in 2021. Also, PROC was negatively correlated with the midterm exams results and also exhibited statistically significant negative Pearson's correlation with achieved points for the report (RAD) (especially for 2021 and 2022), but achieved points for the report (RAD) and results of exams are statistically significant positively correlated (RAD and K1 for 2024, RAD and K2 2021,2022, and 2024).

# 5 Conclusion

This paper presents the analysis of the implementation of PBL/WBL activity in the BDM course, considering the grades achieved and the achievements of self-evaluated levels of competencies.

Analytical thinking and mathematical skills were the least developed competencies over all four years, while teamwork skills were the most developed. Students performed better in generic competencies when teachers assigned problems. However, higher self-assessments of skill achievements did not always correlate with better exam and course results. On the other hand, higher activity grades were positively correlated with exam and overall course results. This suggests that some students overestimated their self-assessed skill achievements.

Based on the review of competencies development results by year, it can be concluded that students developed their teamwork skills to a lesser extent in the year when they chose the topics of WBL/PBL activity themselves (2023). One reason could be groupthink and the dominance of the team member whose idea was selected and analysed in the WBL/PBL activity. Additionally, in the same year (2023), their poor selection of decision-making topics for WBL/PBL activity negatively impacted their ability to identify and solve problems, as confirmed by the overall points achieved. One possible reason for the lower development of idea-generation skills in that year (2023) is the subjectivity in problem analysis, as they chose the problems themselves, and teachers evaluated the solutions as less creative. Due to subjectivity and the anchoring trap during that year, their ability to adapt and act in new situations decreased.

When choosing topics independently (2023), students had a less developed ability to work under pressure because they did not have many opportunities for consultations with teachers and organisations—which were not included in the topic definition. On the contrary, students had a more developed ability to work under pressure when teachers and organisations defined topics for WBL/PBL activities because, in those situations, students had the opportunity to consult them and were not left to themselves.

Students were less independent when teachers defined the topics for WBL/PBL activities (2024) because they had more opportunities for consultations with teachers, which they used. During the same year (2024), analytical thinking skills decreased due to the increased use of artificial intelligence tools, reducing their reasoning and problem analysis.

In light of the foregoing LA analysis, revising the WBL/PBL activity considering the previous conclusions is necessary. In the new activity model, students should not choose topics independently; teachers should define the topics in collaboration with organisations. Students should have more frequent consultations with the organisations presenting the problems, and experts from the organisation should be included in the evaluation process of students' projects. WBL/PBL activity needs to be adjusted to the challenges and opportunities offered by artificial intelligence (AI) because AI can be used to create solutions efficiently. Additionally, it is necessary to train students to evaluate AI-generated solutions critically. Finally, the WBL/PBL activity revision should include an evaluation of each team member's contribution. Those are valuable LA conclusions for teachers at the course.

Consequently, considering the above conclusions, it is proposed that the redesigned activity be implemented for the next academic year. Further, it is recommended that a survey related to competencies development and grade analysis be conducted, which will enable evaluation of the impact of the revised WBL/PBL activity (did the proposals for activity revision achieve desired results).

The limitations of this research include the voluntary nature of the survey, meaning that not all students who enrolled on the course participated in the survey. Further, the surveys were anonymous, so matching survey results with individual student performance was not possible. Finally, competency achievements were assessed through self-assessment, a subjective evaluation method.

## Acknowledgements

This work has been fully supported by the Croatian Science Foundation under the project IP-2020-02-5071.

# References


Almulla, M. A. (2020). The Effectiveness of the Project-Based Learning (PBL) Approach as a Way to Engage Students in Learning. *SAGE Open*, 10(3). https://doi.org/10.1177/2158244020938702

Chis, A. E., Moldovan, A.-N., Murphy, L., & Muntean, C. H. (2018). International Forum of Educational Technology & Society Investigating Flipped Classroom and Problem-based Learning in a Programming Module for Computing Conversion Course. Source: *Journal of Educational Technology & Society*, 21(4), 232–247. https://doi.org/10.2307/26511551

Cifrian, E., Andrés, A., Galán, B., & Viguri, J. R. (2020). Integration of different assessment approaches: application to a project-based learning engineering course. *Education for Chemical Engineers*, 31, 62–75. https://doi.org/10.1016/j.ece.2020.04.006

Dogara, G., Saud, M. S. Bin, & Kamin, Y. Bin. (2020). Work-Based Learning Conceptual Framework for Effective Incorporation of Soft Skills among Students of Vocational and Technical Institutions. *IEEE Access*, 8, 211642–211652. https://doi.org/10.1109/ACCESS.2020.3040043

Eliyasni, R., Kenedi, A. K., & Sayer, I. M. (2019). Blended Learning and Project Based Learning: The Method to Improve Students' Higher Order Thinking Skill (HOTS). *Jurnal Iqra' : Kajian Ilmu Pendidikan*, 4(2), 231–248. https://doi.org/10.25217/ji.v4i2.549

Fergusson, L., Shallies, B., & Meijer, G. (2020). The scientific nature of work-based learning and research: An introduction to first principles. *Higher Education, Skills and Work-Based Learning*, 10(1), 171–186. https://doi.org/10.1108/HESWBL-05-2019-0060

FOI. (n.d.). FOI Nastava - Poslovno odlučivanje. Retrieved from https://nastava.foi.hr/course/128683

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research*, 102. https://doi.org/10.1016/j.ijer.2020.101586

Hammond, J., Keeney, R., & Raiffa, H. (1999). Smart Choices: A Practical Guide to Making Better Decisions.

Hira, A., & Anderson, E. (2021). Motivating Online Learning through Project-Based Learning During the 2020 COVID-19 Pandemic. *IAFOR Journal of Education*, 9(2), 93–110. https://doi.org/10.22492/ije.9.2.06

Hossein-Mohand, H., Trujillo-Torres, J.-M., Gómez-García, M., Hossein-Mohand, H., & Campos-Soto, A. (2021). Analysis of the Use and Integration of the Flipped Learning Model, Project-Based Learning, and Gamification Methodologies by Secondary School Mathematics Teachers. *Sustainability*, 13(5), 2606. https://doi.org/10.3390/su13052606

Hussein, B. (2021). Addressing collaboration challenges in project-based learning: The student's perspective. *Education Sciences*, 11(8). https://doi.org/10.3390/educsci11080434

Kadoic, N., & Slibar, B. (2020). Increasing the Learning Efficiency in Decision-Making Field using the Workshop Activity in Moodle. 2020 *43rd International Convention on Information, Communication and Electronic Technology (MIPRO),* 592–597. https://doi.org/10.23919/MIPRO48935.2020.9245132

Khamitova, A. (2023). Innovative Learning Spaces of Higher Education: a Systematic Mapping Review of Themes. *TechTrends*, 67(5), 830–842. https://doi.org/10.1007/s11528-023-00892-4

Krsmanovic, M. (2021). Course Redesign: Implementing Project-Based Learning to Improve Students' Self-Efficacy. *Journal of the Scholarship of Teaching and Learning*, 21(2). https://doi.org/10.14434/josotl.v21i2.28723

Lester, S., & Crawford–Lee, M. (2023). Learning from digital adaptations to the pandemic: enhancing work-based higher education. *Higher Education, Skills and Work-Based Learning*, 13(4), 786–799. https://doi.org/10.1108/HESWBL-01-2022-0008

Maros, M., Korenkova, M., Fila, M., Levicky, M., & Schoberova, M. (2023). Project-based learning and its effectiveness: evidence from Slovakia. *Interactive Learning Environments*, 31(7), 4147–4155. https://doi.org/10.1080/10494820.2021.1954036

Murtazin, K., Shvets, O., & Piho, G. (2020). Literature Review on Work-Based Learning. 2020 *IEEE Frontiers in Education Conference (FIE)*, 1–8. https://doi.org/10.1109/FIE44824.2020.9274264

Ngereja, B., Hussein, B., & Andersen, B. (2020). Does project-based learning (PBL) promote student learning? a performance evaluation. *Education Sciences*, 10(11), 1–15. https://doi.org/10.3390/educsci10110330

Pažur Anicic, K., & Buselic, V. (2021). Importance of Generic Skills of ICT Graduates - Employers, Teaching Staff, and Students Perspective. IEEE Transactions on Education, 64(3), 245–252. https://doi.org/10.1109/TE.2020.3034958

Pažur Aničić, K., Gusić Munđar, J., & Šimić, D. (2023). Generic and digital competences for employability — results of a Croatian national graduates survey. Higher Education, 86(2), 407–427. https://doi.org/10.1007/s10734-022-00940-7

Rienties, B., Divjak, B., Iniesto, F., Pažur Aničić, K., & Žižak, M. (2023). Online work-based learning: A systematic literature review. *International Review of Education*, 69(4), 551–570. https://doi.org/10.1007/s11159-023-10008-y

Rohm, A. J., Stefl, M., & Ward, N. (2021). Future Proof and Real-World Ready: The Role of Live Project-Based Learning in Students' Skill Development. *Journal of Marketing Education*, 43(2), 204–215. https://doi.org/10.1177/02734753211001409

Rozal, E., Ananda, R., Zb, A., Fauziddin, M., & Sulman, F. (2021). The Effect of Project-Based Learning through YouTube Presentations on English Learning Outcomes in Physics. AL-ISHLAH: *Jurnal Pendidikan*, 13(3), 1924–1933. https://doi.org/10.35445/alishlah.v13i3.1241

Shih, W.-L., & Tsai, C.-Y. (2016). Students' perception of a flipped classroom approach to facilitating online project-based learning in marketing research courses. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.2884

Sudirman, A., & Gemilang, A. V. (2020). Promoting work-based learning as a praxis of educational leadership in higher education. *International Journal of Learning, Teaching and Educational Research*, 19(3), 149–173. https://doi.org/10.26803/ijlter.19.3.9

Sudjimat, D. A., & Permadi, L. C. (2019). Effect of Work-Based Learning Model on Students' Achievement Motivation. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 25(2), 204–212. https://doi.org/10.21831/jptk.v25i2.24416

Tran, E. (2024). 15 Innovative Teaching Methods with Guide and Examples | Best in 2024. AhaSlides. Retrieved from https://ahaslides.com/blog/15-innovative-teaching-methods/

# Enhancing Strategic planning course with operational research methods: a case study for IT students

**Maja Pušnik, Katja Kous, Boštjan Šumak**
University of Maribor
Faculty of Electrical engineering and Computer Science
Koroška cesta 46, 2000 Maribor
{maja.pusnik, katja.kous, bostjan.sumak}@um.si

***Abstract***. *This paper addresses the need to enhance the interdisciplinary skills of information technology students, who traditionally, within their studies, focus mainly on technical courses such as software development, coding, databases, security, and similar. As students advance in their studies, they encounter some interdisciplinary courses, designed to prepare them for future careers that will probably extend beyond coding. The interdisciplinary courses include customer relations, project management, strategic planning, and others. This paper specifically focuses on improving students' strategic thinking abilities, a critical skill often underdeveloped in information technology (IT) curricula. Recognizing the importance of long-term planning for their professional success, the curriculum has been revised to integrate the mathematically oriented Operational Research into the Strategic Planning course, making the two courses more engaging and relevant to market needs. Although the work is still in progress, and the research methodology was limited to standard documentation of attendance, grades, and student feedback, early findings suggest that while improvements are still needed, the changes are moving in the right direction toward better preparing students for the complex challenges they will face in their careers. Therefore, the aim of the paper is to present the modification of the Strategic Planning course with Operational Research, to better prepare students for the needs of the market and motivate them to think more interdisciplinary and out of the box.*

**Keywords.** Strategic planning, Operations research, Operational research

## 1 Introduction

Information technology (IT) students typically follow curriculum that prioritizes acquiring technical knowledge. This approach is driven by the need to equip students with the essential technical skills required in a rapidly evolving computer science field. As a result, courses that enhance soft skills, such as strategic planning, communication, and critical thinking, are not considered as primary, therefore less courses are dedicated to these fields. The mainly technical focus aims to ensure that IT students are proficient in the core areas of computer science (programming, data structures, algorithms, systems architecture, and similar), which are fundamental competencies that are crucial for their employment and technical problem-solving. However, the increasing complexity of the IT landscape and technology integration into all parts of business and society demands a more holistic approach to education.

In the contemporary job market, employers seek well-rounded professionals who not only possess strong technical skills but also the ability to navigate organizational challenges, work effectively in teams, and contribute to strategic decision-making processes. This shift in employer expectations emphasises the importance of soft skills, which are crucial for leadership roles and adapting to the tech industry's dynamic demands.

Despite the traditional focus on technical skills, there is a growing recognition of the need to balance curricula with courses that develop these essential soft skills. Addressing this gap is important for preparing IT students to meet the economy's and society's current and future demands. By integrating strategic planning and other soft skills into the IT curriculum, educational institutions can produce graduates who are not only technically proficient but also capable of leading and innovating in a digitally transformed world.

An additional motivation to reorganize the curricula resulted from several projects, introduced in the last four years, addressing the customer and end-user aspect of software development, how important it is to include all possible participants, and what additional activities should be conducted. For example, several innovative projects are being undertaken within higher education to enhance inclusivity and accessibility, leveraging technology and strategic curriculum development. These projects aim to address various educational challenges and prepare students for

the evolving demands of the modern economy and society. Some of these projects are:

- *Accessibility in Digital Communication Higher Education Curricula* is a project focusing on educating university teaching and educational staff in the field of communication on how to publish digitally accessible content and create a more inclusive teaching environment. By establishing a unique partnership between higher education institutions, educational organizations, and the private sector, the project aims to incorporate accessibility skills into communication curricula. This initiative prepares students to meet current and future economic and societal needs.
- *Introducing Training on User Testing with People with Disabilities into UX Design and Related Higher Education Programs* is a project, which aims to integrate training on accessible and inclusive user testing involving people with disabilities into UX design and related higher education courses. This project enhances the curriculum and empowers individuals with disabilities by raising awareness of their rights and encouraging active participation in usability testing. It promotes equal partnership between people with disabilities and organizations in both the public and private sectors.
- *Self-Evaluation Tools for e-Inclusion in Higher Education Institutions* is a project which recognizes the dual impact of digitalization on education, both positive and potentially exclusionary. It seeks to improve inclusive digital capabilities within higher education.
- *Enhancing Inclusive Education in Higher Education Institutions with Artificial Intelligence* explores the potential of Artificial Intelligence to advance education for students with diverse learning needs, including those with special needs and disabilities, aiming to address the current gaps in teachers' knowledge, skills, and resources regarding the effective integration of AI in educational processes. By leveraging AI tools, the project seeks to create more inclusive and productive educational environments in higher education institutions.

These projects collectively aim to transform higher education by promoting accessibility, inclusivity, and the integration of advanced technological tools, ultimately preparing students for a diverse and dynamic world and were inspirations for creating a new course connected to user experience. However, in addition to the lack of knowledge in the mentioned topics, there are several other areas in which IT students' education could be enhanced. Therfore the secundary objective of this paper is to evaluate the success of enhancing interdisciplinary knowledge among IT students, primarily focusing on strategic planning course.

One of most significant areas in the interdisciplinary IT domain is strategic decision-making, both short and long term. Strategic thinking skills, which are crucial for effective planning and execution of projects, often need further development. The introduction of decision-making frameworks, as well as operational research into the curriculum, is an important part of our efforts. Although some of the topics are covered in previous years, the students need encouragement to use the formerly gained knowledge in their everyday activities just before their studies are completed. These areas are critical for equipping students with the ability to analyse complex situations, make informed choices, and optimize processes within their future workplaces.

By addressing identified gaps, educational institutions can better prepare IT students to meet the complex demands of the modern workforce. This holistic approach to education will not only improve their technical capabilities but also their strategic and operational competencies, making them more versatile and effective professionals. Accordingly, the paper presents an example of the integration of Operational Research into the Strategic Planning course, with the goal of better preparing students for market demands and fostering interdisciplinary thinking. Since this is an ongoing process, the methodology has been limited to standard documentation of attendance, grades, and student feedback. The initial findings indicate that while significant work remains, the changes implemented are progressing in the right direction.

The paper is organized in the following way: after the introduction, the course design and implementation are presented, followed by the longitudinal case study for IT students from 2020 to 2024 in section 3. The results and short analysis are provided in chapter 4, followed by the discussion in section 5 and conclusion section 6, which includes planned future activities.

# 2 Course Design and Implementation

Strategic Planning is an elective course in the study program Informatics and Data Technologies at the University of Maribor, Faculty of Electrical Engineering and Computer Science (UM FERI), 2nd year of master’s study, available for students of modules IT Security And Security Management, Digital Innovation, and Information Systems Management. The course's primary goals include the knowledge from the field of strategic planning in informatics, focusing on their ability to demonstrate, apply, and critically evaluate methodologies, concepts, and best practices of strategic planning in IT.

Strategic planning is generally a systematic approach that organizations use to define their long-term goals and the means to achieve them, encompassing various approaches such as strategic business planning, which focuses on overall organizational objectives and market analysis, and strategic IT planning, which aligns technology initiatives with business strategies to promote efficiency and innovation (Grünig and Kühn, 2015). Key characteristics of strategic planning include its long-term orientation, adaptability, stakeholder involvement, and an iterative nature that encourages continuous improvement (Dixit and Nalebuff, 2010). Understanding the phases of strategic planning involves assessment and analysis, strategy formulation, implementation, monitoring and evaluation, and periodic review and renewal, each of which is essential for ensuring that the strategic plan effectively guides the organization toward its desired future (Dhillon, 2014). All the listed topics were included in the course by default, however the concept of strategic planning has evolved significantly over the past few years, integrating deeper insights and more sophisticated tools to address the dynamic challenges of modern organizations.

In this section, we provide an overview of the Strategic Planning course as it was introduced and developed over the years (from 2020 to 2024), encompassing its definition, types, characteristics, and lifecycle phases. The most significant change was connecting the course Strategic Planning with topics from the field of operational research (a course that existed on its own until 2022 but was reorganized and its topics were integrated into the courses of Optimisation of Business Processes and Strategic Planning). Because operational research plays a critical role in strategic planning by providing quantitative analysis and decision-making frameworks that help organizations optimize their resources and achieve their long-term objectives, such as applying statistical and mathematical models to analyse complex data, enabling organizations to make informed decisions based on empirical evidence rather than intuition alone. Using operational research techniques, organizations can identify the most efficient ways to allocate resources, such as time, money, and personnel. This ensures that strategic initiatives are executed effectively and in a sustainable manner. Moreover, operational research methods allow organizations to simulate scenarios and assess potential outcomes. This capability is crucial for strategic planning, as it enables decision-makers to evaluate risks and opportunities associated with various strategies before implementation. Operational research equips organizations with tools and methodologies, such as linear programming, simulation, and queuing theory, to address complex problems systematically. This structured approach facilitates the identification of optimal solutions in strategic planning. Operational research also provides metrics and analytical techniques to evaluate the effectiveness of strategic initiatives. By monitoring key performance indicators, organizations can adjust their strategies based on real-time data, ensuring alignment with overall objectives. In summary, operational research is essential in strategic planning, allowing organizations to make data-driven decisions, optimize resources, analyze scenarios, and solve complex problems. By integrating operational research into their strategic frameworks, organizations can enhance their agility, improve performance, and achieve sustainable success in a competitive landscape (Hillier and Lieberman, 2010).

As the course Strategic Planning did not initially incorporate operational research methods, they were gradually incorporated into the course over four years. The following list documents topics that were covered each year for the course Strategic Planning (the topics were based on various literature from the field of strategic planning (Cassidy, 2016), (Peppard and Ward, 2016) and (Dobbs and Dobbs, 2016) and (Rumelt, 2012) as well from the field of operational research (Hillier and Lieberman, 2010)):

- *Introduction to Strategic Planning* (2020, 2021, 2022, and 2023), including the main definition of strategic planning, specifics of strategic business planning and strategic planning in information technology, presentation of the characteristics of strategic planning, and understanding the phases/lifecycle of strategic planning.
- *The importance of strategic elements* (Mission, Vision, Values, Goals) (2020, 2021, 2022, and 2023), including the definition of the organization's principal goal, fundamental values, and motives for product existence.
- *Strategic Planning in Informatics* (2020, 2021, 2022, and 2023), including analysis of IT strategic planning and connection to the global strategic plan, linking factors of strategic planning within the organizational and information domain, emphasizing the importance of adapting the strategy to information systems development.
- *Environmental Analysis* (2020, 2021, 2022, and 2023), including understanding the importance of environmental analysis, use of different methods and approaches to strategic planning, and the evaluation of the suitability of methods in specific contexts.
- *Management of strategic planning* (2021, 2022, and 2023), including performance monitoring, performance measurement, evaluation, operational research analytics, scenario planning, and risk management, as well as continuous improvement.
- *Enhancing strategic plans with operations research* (2022, 2023), including application of linear programming for strategic planning activities, optimal resource distribution approaches, and forecasting and prediction.

- *Strategic decision-making technique* (2022, 2023), including what-if analysis of strategic decisions, sensitivity analysis within strategic scenarios and holistic decision-making and optimal strategy selection with operational research.
- *Game theory and strategic games* (2023), including understanding the application of the mathematical framework to analyze the interactions of rational individuals in strategic situations, Application of Nash equilibrium, and negotiation in strategic decision-making.
- *Case studies and practical applications* (2020, 2021, 2022, and 2023), including case study analysis and critical evaluation, lessons learned from strategic failures, and best practices in strategic planning.
- *Emerging trends in strategic planning* (2020, 2021, 2022, and 2023) depend on each year, including agile and adaptive strategies, purpose-driven strategies, Gartner and Innovations overviews.

As additional topics were added, students' success and their attitude towards the course were monitored over four years based on their grade, attendance and feedback. We should note that the elective course does not include large number of students, therefore the small groups within courses enabled a personalized approach to adapting the course according to the needs and interests of students.

# 3 Case Study: Implementation of the Strategic Planning course for IT Students from 2020 to 2024

In addition to evolving course content, the teaching and learning approaches underwent significant transformations over the four years. By implementing a variety of instructional strategies, the program improved student engagement and learning outcomes. The overall assessment method was simple and combined participation in labs and lectures with a final grade, allowing for a general evaluation of student performance. The teaching approach evolved significantly from student year 2020/2021 to 2023/2024, adapting to enhance student engagement and deepen understanding of strategic planning concepts.

In 2020/21, the educational framework was primarily centered on a standard curriculum that included traditional lectures and laboratory sessions, providing foundational knowledge but offering limited opportunities for interactive learning. This model laid the groundwork for more dynamic pedagogical strategies. However, the approach was upgraded the following year.

In 2021/22, the introduction of the flipped classroom model created a shift in teaching methodology. Students were encouraged to engage with instructional content outside the classroom through reading materials, allowing class time to be devoted to discussions, problem-solving, and collaborative activities. This approach fostered a more active learning environment, empowering students to take greater ownership of their education.

The following year, in 2022/23, the curriculum further advanced by integrating active strategic planning development within the classroom. Students participated in real-time strategic planning exercises, working collaboratively to apply theoretical concepts to practical scenarios. This experimental learning approach not only enhanced their understanding of strategic planning but also developed skills such as teamwork, critical thinking, and effective communication.

In 2023/24, the focus expanded to include specialized topics related to decision models and operational research. This year emphasized the importance of analytical frameworks and quantitative methods in strategic planning, equipping students with tools to make informed decisions in complex scenarios. By incorporating these elements, the learning approach not only addressed theoretical aspects of strategic planning but also provided practical insights into operational efficiency and effective decision-making processes, ultimately preparing students for the complex challenges of the modern business landscape, requiring ability of interdisciplinary and interdisciplinary thinking and advanced soft skills.

# 4 Results and analysis

As the course content gradually changed during the 4 years, the teaching methods also changed to get as much collaboration and focus from the students as possible. As no additional survey was conducted, we documented the average grade, attendance rate, and students' general evaluation. The results are presented in Table 1. By focusing on attendance as a key metric, the analysis highlighted the relationship between student participation and academic success, informing future decisions regarding curriculum enhancements.

**Table 1.** Student's statistics

| Year | Number of students | Average attendance | Average grade |
|---|---|---|---|
| 2020/2021 | 25 | 72% | 83% |
| 2021/2022 | 27 | 73% | 84% |
| 2022/2023 | 24 | 64% | 73% |
| 2023/2024 | 24 | 78% | 80% |

With the exception of the generation of the year 2022/2023, the average attendance slowly increased, while the grade slowly decreased over the 4 years.

Although the additional topics did not create more learning material (as existing topics were shortened and the overall amount of work remained the same), the introduction of operational research resulted in higher attendance, but a lower grade (Fig. 1), documenting the increasing difficulty of the topic or lack of knowledge in fields, which should already be understood by the students, however the existing knowledge was not always well incorporated into new topics.

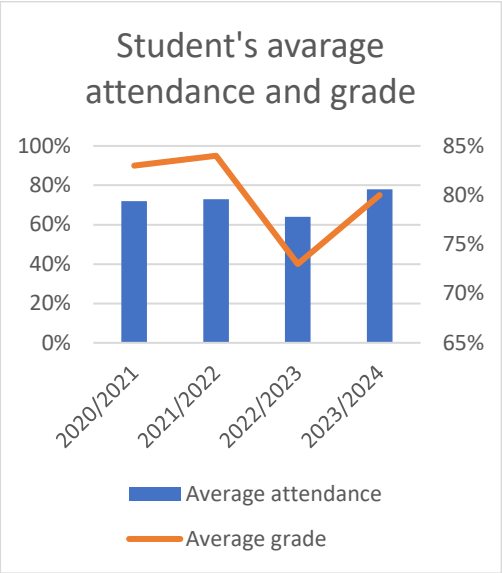



**Figure 1. Relationship between student's average attendance and grade.**

There was a positive 0.69 correlation between attendance and grades, indicating the higher the attendance, the higher the grade. An additional note is that not all students completed the course:

- In 2020/2021, 6 students (24%) were unsuccessful,
- In 2021/2022, 2 students (7%) were unsuccessful,
- In 2022/2023, 4 students (17%) were unsuccessful,
- In 2023/2024, 4 students (17%) were unsuccessful.

Each year, after all obligations within a course are completed, the students evaluate the course including the professor and the assistant (each study module separately). The following grades were provided on a scale from -2 to +2:

- 2020/2021 - Students grade 1.6 and 1.71 for different modules,
- 2021/2022 - Students grade 1.8, 1.88 and 1.5 for different modules,
- 2022/2023 - Students grade 1.5 for all modules,
- 2023/2024 - Students grade N/A, as the results are available by the end of the year.

Based on the student survey, the course grade decreased, having a peek in 2021/2022, when the flipped classroom technique was introduced for the first time. Note that the typical average grade at the University of Maribor level for all courses was 1.45 in 2020/2021, 1.48 in 2021/2022, and 1.46 in 2022/2023.

# 5 Discussion

The results highlight the success of students in the Strategic Planning course following the COVID situation. The aim was to create a more engaging subject that appeals to IT students and fills gaps in the knowledge they may have missed in previous years of study. However, since operational research played a significant role in the previous years of study, we aimed to integrate this existing knowledge into the revised program. The main included themes of operational research, which were included, are optimizing cost/profit using linear programming, solving transportation and production problems to maximize resource efficiency and minimize waste, conducting sensitivity and what-if analyses, applying the critical path method to find the most optimal solution and pay attention to critical ones, and exploring game theory. These topics were incorporated to establish a stronger connection with strategic planning.

As mentioned in the beginning of the paper, these topics, along with others, were part of a different course and some were not included in Strategic Planning. However, following curriculum changes that incorporated Operational Research into Strategic Planning, several of existing topics were integrated, promoting interdisciplinary thinking and connecting diverse areas of knowledge. Although student's grades did not increase significantly and even dropped slightly in the 2022/23 academic year, both interest and comments in the form of unorganized qualitative data displayed progress. Students were more engaged and felt that the strategic plans were more relevant to their activities. Student satisfaction with the inclusion of Operational Research in the course increased, particularly as they gained a better understanding of how soft skills and their work in the economic field are connected to mathematical knowledge and algorithmic thinking.

# 6 Conclusion

The paper addresses the systematic reorganization process of the course Strategic Planning, adapted for IT students. The data reveals a general trend of variability in both average attendance and grades. Attendance peaked in the 2023/2024 academic year at 78%, while the highest average grade was achieved in 2021/2022 with 84%. The students' grades in relation to their

individual performance metrics show some variation across the years. The data highlights the need for consistent monitoring of both attendance and grades to better understand and enhance student outcomes in the course.

The future work will address the challenges in strategic planning in higher education institutions and bridge the gap between current computer science programs and the evolving needs of the private sector. The methodology will be further refined in future research. For now, the findings are primarily relevant to one faculty with a specific set of students. Only after conducting additional research in the coming academic years will enable us to offer broader guidelines applicable to similar programs. The overarching objective is to incorporate strategic planning skills into higher education curricula, equipping future developers with the necessary knowledge to meet economic (Carvalho, 2024) and societal (Pacios, 2024) demands. One of the future aims will be identification of the gap by conducting an assessment to pinpoint the disparity between current computer science programs and future requirements related to strategic planning and optimization activities in software development. Among other activities, we plan to investigate existing tools as well as develop recommendations and best practices in the field of strategic planning for IT professionals. Other similar case studies will be taken into consideration as well (Arevalo-Ascanio et. all 2024).

The research is in its early stages, with limited data currently available. This paper serves as a preliminary investigation into how to improve the long-term study program, beginning with the course of strategic planning. Data collection will be expanded in the coming year through an extended survey including both qualitative and quantitative data.

## Acknowledgments

The authors acknowledge the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0057). This research was conducted within the activities of the Self-Evaluation Tools for e-Inclusion in HEI (SET4Inclusion) project (KA220-HED-000088368) and Introducing training on user Testing with people with disabilities into UX design and related higher education programmes (INTUX) project (KA220-HED-28BE6544) , which are projects co-funded by the Erasmus+ Programme of the European Union.

## References

Arevalo-Ascanio, R., De Meyer, A., Gevaers, R., Guisson, R., & Dewulf, W. (2024). *From operational to strategic modelling: A continuous multi-scale approach for last-mile analysis*. Transportation Research Part E: Logistics and Transportation Review, 191, 103738. https://doi.org/https://doi.org/10.1016/j.tre.2024.103738

Carvalho, J.M.S. (2024). *Strategic Planning. Research on Economics*, Management, and Information Technologies, Universidade Portucalense, Porto, Portugal. https://doi.org/10.1016/B978-0-443-13701-3.00058-X.

Cassidy, A. (2016). *A Practical Guide to Information Systems Strategic Planning*, Second Edition, AUERBACH.

Dhillon, G.S. (2014). *Strategic Information Systems Planning: Readings and Cases*, Semantic Books.

Dixit, A.K. and Nalebuff, B.J. (2010). The Art of Strategy: A Game Theorist's Guide to Success in Business and Life. W. W. Norton & Company.

Dobbs, J. H., & Dobbs, J. F. (2016). *Strategic planning: a pragmatic guide.*

Grünig, R., & Kühn, R. (2015). *The Strategy Planning Process*. Springer Berlin Heidelberg.

Hillier, F. S., & Lieberman, G. J. (2010). *Introduction to Operations Research* (9th edition). McGraw Hill. Higher Education.

Pacios, A. R. (2024). *Strategic Management and Planning. In Reference Module in Social Sciences*. Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-323-95689-5.00041-9

Peppard, J., & Ward, J. (2016). *The strategic management of information systems: building a digital strategy* (Fourth Edition). Hoboken Wiley.

Rumelt, R. P. (2012). *Good Strategy/Bad Strategy: The Difference and Why It Matters*. Strategic Direction, 28(8).

# Trends in Software Engineering

Víctor-Pablo Prado-Sánchez, Adrián Domínguez-Díaz, Daniel Rodríguez, José-Javier Martínez-Herráiz, Luis de-Marcos
**How can ChatGPT help humans in Dark Web content classification? Assessing GPT models reliability and effects of explanations on human decisions**

Juraj Dončević, Kristina Ćaleta, Marko Horvat, Ivica Botički
**Mine Action Data Observatory: Tackling Uncertainty and System Longevity with Evolutionary Architecture**

Miro Zdilar
**Attribute Based Access Control Metamodel for Spreadsheet Programs**

Sara Kijanović, Darko Stefanović, Dušanka Dakić, Teodora Vučković
**Using AI on Open Data: A Systematic Literature Review**

Michaela Pauriková, Michal Varga, Norbert Adamko
**Reevaluation of intentions of dynamic agent utilizing BDI reasoning in ABAsim architecture**

Lea Masnec, Zlatko Stapic, Maja Benkus
**Bridging Academia and Industry: A Collaborative Approach to Software Development Education**

# How can ChatGPT help humans in Dark Web content classification? Assessing GPT models reliability and effects of explanations on human decisions

**Víctor-Pablo Prado-Sánchez, Adrián Domínguez-Díaz, Daniel Rodríguez, José-Javier Martínez-Herráiz, Luis de-Marcos**

Department of Computer Science, University of Alcalá, Spain

Alcalá de Henares, España

{victor.prado, adrian.dominguez, daniel.rodriguezg, josej.martinez, luis.demarcos}@uah.es

***Abstract***. *This study evaluated the reliability of ChatGPT for Dark Web content classification and the effect of its explanations on human classification decisions. The study objectives were to determine whether ChatGPT can be considered a reliable classifier, identify the content categories where it performs differently than humans, and assess if its explanations can enhance human understanding of these contents. Results show that while ChatGPT is outperformed by supervised models, its reliability is similar to human classifiers, outperforming them in technology-related categories. Additionally, while ChatGPT's explanations allowed humans to better understand its decisions, their effect was not consistent across different people.*

**Keywords.** Dark Web, Darknet, Cybersecurity, Language models, LLM, NLP, ChatGPT, GPT-3.5.

## 1 Introduction

The increased volume and complexity of activities conducted on the Dark Web pose challenges for the classification of its contents, especially in the context of illegal and potentially harmful activities (Avarikioti et al., 2018). Consequently, there has been a growing interest in applying machine learning techniques to address this issue (Al Nabki et al., 2017). In this regard, various approaches based on supervised machine learning models have been developed to detect and classify illegal activities on the Dark Web (Jin et al., 2022).

Given the linguistic nature of the contents to be classified, the performance of pre-trained language models in this task has been explored. For example, the use of models from the BERT (Bidirectional Encoder Representations from Transformers) family has proven effective in classifying text across various domains, including Dark Web content (Devlin et al., 2019; Jin et al., 2022). The need to evaluate the performance of different models has led to the creation of datasets with manually labeled texts from Dark Web sites, such as DUTA (Darknet Usage Text Addresses) and CoDA (Comprehensive Darkweb Annotations) (Al Nabki et al., 2017; Jin et al., 2022). Despite the achieved performance, these models require a prior phase of supervised learning or fine-tuning of the language model on a portion of the dataset to effectively classify them (Arslan et al., 2021). This limitation suggests that the classifiers' performance in real-world environments would be limited by the availability and quality of labeled data, as well as the possibility of overfitting to the training data, making it difficult to adapt to the dynamic and heterogeneous nature of the Dark Web (Al Nabki et al., 2017).

An alternative that would reduce the need for labeled data and the effort of human classifiers is the use of large language models, such as GPT-3 and its derivatives, through zero-shot or few-shot prompting (Kalyan, 2024; Roumeliotis & Tselikas, 2023). When using these techniques, the language model is not specifically fine-tunned for the required task, and just receives prompt messages with the description of the task to perform (zero-shot) or the description and a few classification examples (few-shot). The model must rely on its pre-existing knowledge and understanding, which it has acquired during training on a broad range of data, to perform the classification task (Reiss, 2023). While this approach has been tested for financial text classification (Loukas et al., 2023) or sentiment analysis tasks (L. Li et al., 2023), it performance and reliability remains to be tested for Dark Web content classification, where different categories of content have specific and somewhat obscure terminology which could make both humans and artificial classifiers struggle to determine the correct category of content. Understanding how ChatGPT performs compared to other classification models and to humans

would facilitate its integration in real-work environments.

In addition to its usage as an isolated classifier, the capability of ChatGPT to explain its classification decisions in natural language could also be used to help humans better understand Dark Web contents. In other contexts, research such as (Huang et al., 2023) explores ChatGPT ability to explain implicit hate speech, offering a valuable tool for understanding and addressing this sensitive issue. Similar capabilities in the Dark Web context could help law enforcement agents to better identify and understand illegal or potentially harmful contents.

This study aims to address several key objectives: firstly, to determine whether ChatGPT can be considered reliable for Dark Web content classification tasks; secondly, to identify the types of Dark Web contents where ChatGPT classifier performance differs more from humans; and thirdly, to assess the extent to which ChatGPT explanations can help improve human understanding of Dark Web contents for classification tasks. To achieve these objectives, the study provides a comprehensive evaluation of ChatGPT's ability to classify Dark Web content by comparing its results with other models and with humans. Furthermore, it investigates how ChatGPT's explanations impact human classification of Dark Web contents. This paper provides a comprehensive view of the potential role of ChatGPT and humans in collaborative Dark Web content classification. To that end, the following research questions have been defined:

- RQ1: Can ChatGPT be considered reliable for Dark Web content classification tasks?
- RQ2: In what content categories does ChatGPT classification performance differ from classification performed by human?
- RQ3: To what extent can ChatGPT explanations help improve human understanding of Dark Web contents for classification tasks?

The remainder of this document is structured as follows: Section 2 provides an overview of related work on text content classification and on GPT models explanatory capabilities. Section 3 details the methodology employed to evaluate ChatGPT reliability and the effects of its explanations on human classification decisions. Section 4 presents the results and discusses the findings considering the research questions. Finally, Section 5 outlines the conclusions and discusses future lines of work.

# 2 Related Work

## 2.1 ChatGPT classification metrics and reliability in zero-shot prompting

The natural language processing capabilities of the latest LLMs, such as GPT models, enable tackling text classification tasks without the need for specific prior training for the classification categories.

In the financial domain, (Loukas et al., 2023) proposes using conversational GPT models, such as GPT-3.5 and GPT-4, to classify financial texts with minimal examples and compares their performance with other pre-trained and fine-tuned models. GPT-3.5 and GPT-4 outperform other models in classification with 1 or 3 examples, achieving F1 scores of 75.2% and 83.1%, respectively, but trail behind fine-tuned language models. The study highlights the ease of implementing effective classifiers through services like OpenAI, though it notes the high cost for small organizations as a drawback. Reliability is assessed using a confidence metric, where values close to 1 indicate high reliability and values close to 0 indicate low confidence. These confidence values are crucial for evaluating the certainty of model predictions across different configurations in the financial domain.

In addressing the challenge of detecting harmful content on social networks, (L. Li et al., 2023) employs ChatGPT as a detection model, comparing its results with human annotations. Achieving an 80% precision in identifying harmful content, the study emphasizes the consistency of classifications while acknowledging the impact of the prompt used on its performance. ChatGPT's reliability is assessed by comparing its results with annotations from MTurkers, reporting approximate values of around 80% for Precision, around 70% for Recall, and around 75% for F1-score. These results indicate ChatGPT's ability to accurately identify harmful comments compared to human annotations.

Finally, in (Reiss, 2023) examines the reliability of ChatGPT for annotation and text classification tasks in various application contexts. While its potential is recognized, concerns are raised about its non-deterministic nature, which can lead to variable results even with small changes in inputs. Caution is advised when using ChatGPT for these tasks without additional validation, such as comparisons with human-annotated data. In the study, a Krippendorff's Alpha value above 0.8 is considered to indicate reasonable reliability in ChatGPT outputs for text classification tasks. Values below this threshold raise questions about the reliability of the results. It is recommended to validate ChatGPT outputs with human-annotated data to ensure reliable results.

### 2.2 Dark Web Language and existing datasets and classifiers

In the realm of Dark Web research and online illegal activity detection, several studies have been conducted exploring different machine learning algorithms and language models for text classification through supervised learning. This is due to the need to address the inherent challenges in analyzing the specific and diverse language that characterizes the Dark Web. Furthermore, the scarcity of public data and the anonymity and cryptography techniques used in this clandestine digital environment have driven the creation of specialized datasets. For instance, (Al Nabki et al., 2017) introduces the DUTA dataset for active domains on the Dark Web, comprising 6,831 documents classified into 26 categories. It was found that the combination of TF-IDF and Logistic Regression achieves a weighted F1 score of 97% in classifying a subset of the dataset into 9 categories.

In an attempt to overcome some limitations of the DUTA dataset, (Jin et al., 2022) introduces the CoDA dataset, consisting of 10,000 web documents obtained from the Dark Web for text analysis. Linguistic differences between the Dark Web and the Surface Web are examined, and differences between the CoDA and DUTA datasets are analyzed. The performance of various Dark Web page classification methods is also evaluated, achieving a weighted F1 score of 92.49% using the pre-trained BERT language model, following fine-tuning for document classification. In (Jin et al., 2023), an enhanced classifier based on a pre-trained BERT adaptation with texts obtained from the Dark Web is presented, achieving an F1 score of 94.25% in classifying the contents of the CoDA dataset.

### 2.3 ChatGPT natural language explanations

The ability of ChatGPT to provide natural language explanations of its decisions is mentioned. We explore how these explanations can be beneficial to individuals by providing a clear and easily accessible understanding of the reasoning behind the responses generated by the model.

In the context of recommender systems (RSs), ChatGPT can generate explanations that justify the recommendation and contextualize it in a way that is relevant to the user (Silva et al., 2024). Research indicates that personalized explanations are more effective than generic ones, especially when the recommended item is unfamiliar to the user, highlighting the importance of tailoring explanations to each individual.

Multiple studies have examined ChatGPT's ability to provide natural language explanations in the context of sentiment analysis (Huang et al., 2023, 2024). These studies show that GPT models can generate effective explanations regarding hate speech detection and have compared its explanations with those generated by humans, evaluating their quality and alignment with human standards.

The role of ChatGPT in generating interpretable explanations in mental health analysis is also noted, with promising results suggesting its utility in this area. In (Yang et al., 2023), the limitations of previous research are addressed through a comprehensive evaluation of the models' mental health analysis and emotional reasoning abilities of ChatGPT across 11 datasets and 5 tasks. Results show that ChatGPT generates explanations that approach human performance, showing great potential in explainable mental health analysis.

Finally, (B. Li et al., 2023) assesses ChatGPT's overall capability across seven information extraction tasks. The performance, explainability, calibration, and fidelity of ChatGPT are analyzed. While the performance varies greatly depending on the information extraction setting, the study shows that ChatGPT provides high-quality and trustworthy explanations for its decisions.

## 3 Methodology

### 3.1 CoDA dataset selection

The CoDA dataset (Comprehensive Darkweb Annotations) was utilized, representing a valuable public resource for Dark Web analysis, comprising a collection of 10,000 web documents intended for text-based research in this environment. These documents cover a wide range of topics and were classified into ten distinct thematic categories. Primarily in English, the documents were sourced from onion services on Tor, i.e., the Dark Web, providing significant insight into this relatively unexplored digital space.

The documents were categorized into ten thematic categories, including Drugs, Financial, Gambling, Cryptocurrency (Crypto), Hacking, Arms/Weapons (Arms), Violence, Electronics, as well as Pornography and Others. This wide variety of categories allows for an assessment of various aspects of activity on the Dark Web. It is worth noting, however, that it was necessary to exclude web documents categorized as Pornography due to conflicts with OpenAI's content policies.

### 3.2 Zero-shot classification with GPT 3.5 Turbo model

For the classification of documents from the Dark Web, the decision was made to employ OpenAI's GPT-3.5 language model, particularly the variant known as GPT-3.5 Turbo. This choice was based on the demonstrated ability of this model to efficiently process text and understand the complex and varied context characteristic of the Dark Web (Ye et al., 2023).

The selection of this model variant was justified by its better balance between capabilities and cost compared to other models such as GPT-4, which was 20 times more expensive at the time of the study, and its better suitability to the problem's needs, with an input context of 16,385 tokens that is suitable for the size of documents to be classified.

The model received the necessary data to perform the classification task under a zero-shot prompting model (Chen et al., 2023). For each Dark Web site in the dataset, ChatGPT was sent a prompt that explained the classification task to perform and included the site text content. The task explanation included the name and descriptions of the possible categories, as defined in the CoDA dataset (Jin et al., 2022), and requested an answer that included the name of the category that best described the content, as well as to briefly explain the reasons why it has chosen that category. The language used for the prompt was English, as the bulk of documents included in CoDA are in this language. The parameters of the GPT model call were set by the system's defaults.

### 3.3 Reliability assessment and human-GPT disagreement analysis

Based on classification results, standard classification performance metrics and inter-coder reliability measures were calculated, comparing ChatGPT classification with the original human-made classification included in the dataset. Then, for those documents in which ChatGPT and the human classification disagreed, a comprehensive analysis and tie-breaking process was conducted to determine up to what point ChatGPT, or human assigned classes were right. The tie-breaking process was performed by three independent reviewers over a random sample of 180 documents, distributed in 20 documents per category, based on the original dataset classification.

To carry out the tie-breaking process, a double iteration approach was implemented. In the first iteration, the disagreements were reviewed without accessing the explanation that ChatGPT provided to justify the class it selected. In this initial phase, reviewers determined if they agreed with the class originally assigned by humans (which was included in the dataset) or with the class assigned by ChatGPT relying solely on the analysis of the content of the site in question. Reviewers could answer if they thought the original classification was right (0), the ChatGPT classification was right (1), or if they were not sure (2). Subsequently, in the second iteration, researchers were exposed to the explanation provided by ChatGPT, so the evaluators had the opportunity to consider the reason behind ChatGPT decision before making a final decision. Again, they were allowed to provide one of the three possible answers. Based on the results of the second iteration, each document was assigned to a final category, assigning the original dataset category or the one selected by ChatGPT when at least two or the three reviewers agreed with either of them, or a "Not sure" value when that condition was not met.

The average percentage of agreement with humans and with ChatGPT was calculated for the whole sample after each iteration. Average percentage of agreement was also calculated for each category, grouping documents by the final category they were assigned to at the end of the tie-breaking process. Finally, inter-coder agreement was calculated to determine the level of agreement between the three reviewers for the whole sample and for each category.

The main objective of this process was first to assess whether human or ChatGPT performed better in difficult to classify documents, and to identify in which categories human or the GPT model performed differently. In addition, the two iterations approach was adopted to determine whether the explanation provided by ChatGPT influenced evaluators' decision making, leading them to change their initial classification decision or keeping it unchanged.

## 4 Results and discussion

### 4.1 ChatGPT classification reliability

As shown in (Table 1), the performance of supervised classifiers (Jin et al., 2022) reveals several important considerations. First, it is observed that supervised classifiers such as SVM, CNN and BERT, show generally superior performance to ChatGPT with the GPT-3.5 Turbo model in terms of accuracy, recall and F1-score. This suggests that supervised approaches may benefit from a larger labelled dataset specific to the classification task compared to the zero-shot learning model used by ChatGPT. The result is in line with studies showing that GPT models without specific training offer good performances, but always below other algorithms prepared specifically for the problem to be addressed (Kocoń et al., 2023).

**Table 1.** Performance comparison (weighted average) with supervised classifiers

| Model | Precision | Recall | F1 |
|---|---|---|---|
| SVM | 91.59% | 91.17% | 91.19% |
| CNN | 88.08% | 87.30% | 87.23% |
| BERT | 92.51% | 92.50% | 92.49% |
| **ChatGPT** | **85.63%** | **82.63%** | **83.14%** |

It is important to consider the way supervised classifiers are trained and the possibility of overfitting. Supervised models, when trained on labelled data, can be overfitted to the specific details of the training set. Since the datasets are drawn from a limited number of sources, the performance of these classifiers in production environments may suffer when classifying documents from other sources. On the other hand,

ChatGPT is based on a zero-shot approach, which means that it lacks prior training on the specific documents to be classified. This could lead to a more generalizable performance, maintaining similar performance values in real environments to those obtained on the dataset.

To evaluate the reliability of ChatGPT as a potential substitute for a human in Dark Web classification tasks, intercoder reliability values such as Cohens Kappa and Krippendorf's Alpha were examined. These values provide a measure of the consistency between ChatGPT's and the original human-made classification. Intercoder reliability metrics (Table 2) indicate a high level of agreement between ChatGPT and the original human CoDA classifiers. With a Cohen's Kappa of 0.8, a Weighted Cohen's Kappa of 0.82 and a Krippendorff's Alpha of 0.79, these metrics exceed (Cohen's Kappa) or get near (Krippendorffs' Alpha) the thresholds considered indicative of excellent agreement and high reliability.

In clinical studies, a Kappa value of 0.8 is considered indicative of excellent agreement. Furthermore, it is mentioned that when both sensitivity and specificity are less than 0.9, they can never produce a Kappa of 0.8 or higher, even with a raw agreement approaching maximum sensitivity and specificity (Feuerman & Miller, 2008). On the other hand, an Alpha value close to 1 indicates high reliability in measuring interobserver agreement (Krippendorff, s. f.). High Alpha values, such as 0.8 or higher, suggest good reliability in the data.

These metrics indicate that ChatGPT could be used effectively to complement the classification performed by humans on the dataset. Its high level of agreement with the original human classifiers suggests that the model can provide accurate and reliable classification of the data, making it a valuable tool for improving and streamlining classification processes in different contexts.

**Table 2.** Intercoder reliability metrics between human and ChatGPT classifiers

| | |
|---|---|
| Cohen's Kappa | 0.8 |
| Weighted Cohen's Kappa | 0.82 |
| Krippendorff's Alpha | 0.79 |

## 4.2 Human vs ChatGPT reliability

The results of the tie-breaking process between human classifiers and ChatGPT (Table 3) reveal valuable insights into the performance and reliability of ChatGPT and human classifiers across different content categories. For each category, the level of agreement with human classification and with ChatGPT classification was calculated by dividing the number of documents that reviewers agreed to assign to the category (with the agreement of 2 or more reviewers) by the total number of documents that the dataset or ChatGPT, respectively, originally assigned to the same category. In addition, the average level of agreement between the three reviewers was measured for the whole sample and for each category by calculating the Fleiss' Kappa intercoder reliability metric. Fleiss' Kappa is related to Cohen's Kappa but applicable to more than 2 raters (Powers, 2012)

**Table 3.** Comparison of reviewers' agreement with humans and with ChatGPT across diverse categories

| Final Class | Agree with human | Agree with ChatGPT | Fleiss' Kappa |
|---|---|---|---|
| Gambling | 10 / 20 (.5) | 1 / 1 (1) | -0.18 |
| Arms/Weapons | 8 / 20 (.4) | 2 / 4 (.5) | -0.09 |
| Electronics | 11 / 20 (.55) | 4 / 4 (1) | 0.61 |
| Hacking | 9 / 20 (.45) | 18 / 18 (1) | 0.41 |
| Drugs | 3 / 20 (.15) | 8 / 12 (.67) | 0.15 |
| Crypto | 5 / 20 (.25) | 15 / 20 (.75) | 0.74 |
| Financial | 3 / 20 (.15) | 13 / 40 (.325) | 0.38 |
| Violence | 1 / 20 (.05) | 5 / 6 (.83) | 0.32 |
| Others | 4 / 20 (.2) | 44 / 66 (.66) | 0.14 |
| **Total** | **54 / 180 (.3)** | **110 / 180 (.61)** | **0.32** |

Overall, reviewers show higher degree of agreement with ChatGPT that with humans across the sample, although the intercoder reliability, considering Cohen's Kappa reference values (Altman, 1990) is just fair (> 0.2) when considering the whole sample. This probably indicates that the sampled documents are generally difficult to classify both for ChatGPT and humans alike. There could be different reasons, such as difficult to understand language, ambiguous content that could fit multiple categories or the lack of a well stablished classification protocol that unequivocally determines a document class when multiple values are possible. An in-depth analysis of the reviewers' decisions during the tie-breaking process would be required to understand low intercoder reliability in each category properly.

Tie-breaking results show that reviewers had higher level of agreement with ChatGPT than with the original human-made classification, with most categories left with 10 or less documents assigned to their original category (from the original 20 documents per category) and an overall agreement of 30% across the sample, clearly below 50%. These values suggests that ChatGPT, with an overall agreement of 61%, could be a more reliable classifier than humans for Dark Web content classification tasks, performing

better in extensive document classification and with difficult to classify documents. This conclusion must be taken with caution though, as low intercoder reliability in the tie-breaking process means that agreement percentages are not strongly grounded. However, it shows a tendency that could be confirmed by additional studies.

Focusing on specific categories, high differences in agreement percentage between humans and ChatGPT, accompanied by a moderate or good intercoder reliability (Fleiss' Kappa > 0.4) can be found in the "Electronics" (0.55 for human vs 1.0 for ChatGPT), "Hacking" (0.45 for human vs 1.0 for ChatGPT) or "Cryptocurreny" (0.25 for human vs 0.75 for ChatGPT) categories. These results suggest that the original human classifiers struggled to understand advanced technological terminology related to electronic devices, hacking and cryptocurrency products and services offered through the Dark Web, while ChatGPT may have a stronger performance due to its capabilities to process and understand complex and dynamic technical language.

Even with insufficient or fair intercoder reliability during the tie-break, high differences in agreement levels in "Drugs", "Violence" and "Others" categories should also be highlighted. According to the reviewers, the original dataset contained many documents misclassified in these categories. Although an in-depth analysis of the documents is out of the scope of this paper, as an example, reviewers reported that many documents originally classified as "Violence" where just describing historical events or contained political discussions which did not fit the category description of "Human trafficking, hitman, kidnapping, poisoning, torture, extortion, sextortion, sex slavery, blackmail, etc.", and were classified as "Others" by ChatGPT. This case could be pointing towards the importance of reporting the classification protocol followed during the creation of a dataset, so ambiguous cases could be resolved unequivocally in future classification studies. That classification protocol is not included in the CoDA dataset description, and thus it's not easy to determine classification errors for certain documents in the original dataset.

## 4.3 Effect of ChatGPT explanations on human classification

Before and after explanation tie-breaking results (Table 4) show that reviewer's level of agreement with ChatGPT increased 20 percentage points after the exposition to its explanations, reducing almost equally the agreement with the original dataset and the agreement uncertainty by approximately 10 percentage points each, while Fleiss' Kappa intercoder reliability sightly increased by 0.05 units. This shows that while the explanations provided by ChatGPT had an impact on the average levels of agreements, convincing some reviewers to agree with ChatGPT decisions; the effect was not consistent in a per-document basis for the three reviewers, becoming insufficient to increase intercoder reliability from low (> 0.2) to moderate (> 0.4) level. The brevity or the insufficient quality of the explanations could be limiting factors that prevent reviewers to agree in the classification. Future studies should measure the quality of explanations and acknowledge possible differences between reviewers during the classification process to better understand the influence of explanations in classification decisions.

**Table 4.** Before and after explanations tie-breaking results showing average level of agreement and intercoder reliability.

| | Before | After | Diff |
|---|---|---|---|
| Fleiss' Kappa | 0.27 | 0.32 | +0.05 |
| Level of agreement | | | |
| With human | 73 / 180 (.4) | 54 / 180 (.3) | -0.1 |
| With ChatGPT | 74 / 180 (.41) | 110 / 180 (.61) | +0.2 |
| Not sure | 33 / 180 (.18) | 16 / 180 (.09) | -0.09 |

As an example of the effect of explanations, in Table 5 we can observe a record from the CoDA dataset which is initially categorized as Others, but categorized as Drugs by ChatGPT. While one of the reviewers initially agreed with the original classification, after evaluating the reason behind ChatGPT's categorization, she reassessed her tie-break decision and changed it, finally agreeing with ChatGPT response.

**Table 5.** Tie-break decision altered by ChatGPT explanation.

| | |
|---|---|
| Content | urlaub / vacation corona hat uns im griff … bis auf weiteres nur notprogramm. corona got us … no shipping for some time to abroad … unser notprogramm für euch auf darkmarket, empire und cannazon… greensupreme ID_EMAIL ——begin pgp public key block—— version: gnupg ID_VERSION gafo3nirma== =gfqo ——end pgp public key block—— |
| Before explanation | Others |
| ChatGPT Category | Drugs |
| ChatGPT categorization reason | The content mentions 'greensupreme' and provides a PGP public key block, which are typically associated with the sale |

| | |
|---|---|
| | of drugs on darknet markets. Additionally, phrases like 'no shipping for some time to abroad' suggest that the website is offering drug products for sale. |
| After explanation | Drugs |

# 5 Conclusions

This study investigated the reliability and performance of ChatGPT, specifically the GPT-3.5 model, in classifying Dark Web content using a zero-shot learning approach. In addressing the first research question, it was found that although supervised classifiers such as SVM, CNN, and fine-tuned BERT models outperform ChatGPT in precision, recall, and F1-score, ChatGPT still shows significant reliability, comparable to human classifiers. The review of discrepancies between ChatGPT and the original human classification showed improved ChatGPT performance in technology-related categories and revealed difficulties in certain ambiguous cases both for humans and models alike. Lastly, the research demonstrated that ChatGPT's brief natural language explanations influence human understanding and decision-making, but they are not sufficient to drastically increase intercoder agreement as measured by Fleiss' Kappa metric.

The promising results suggest several future research directions. These include refining classification protocols to improve consistency and reliability, especially for ambiguous cases, and investigating the fine-tuning of GPT models on domain-specific datasets to enhance accuracy and reliability. Finally, conducting longitudinal studies to evaluate ChatGPT's performance in real-world Dark Web monitoring and law enforcement scenarios will be crucial for understanding its impact on operational efficiency and accuracy over time.

Finally, some limitations that could put the generalizability and accuracy of the study should be noted. First, the small sample size for human-GPT disagreement analysis and potential reviewer bias complicates the assessment of performance discrepancies. Additionally, the two steps iterative review process may not fully isolate the impact of ChatGPT's explanations, as reviewers could be influenced by their initial decisions or the mere presence of an AI-generated explanation. Finally, the quality and clarity of ChatGPT's explanations themselves are not systematically evaluated, which could affect their usefulness in improving human understanding. These factors underscore the need for cautious interpretation and highlight areas for future research refinement to enhance the study's robustness.

# Acknowledgments

This work has been funded by the Spanish Ministry of Science and Innovation, under the "Project for the analysis and recovery of criminal evidence associated with hidden networks" (PARCHE), with reference PID2021-125645OB-I00 of the 2021 call for "Knowledge Generation Projects".

# References

Al Nabki, M. W., Fidalgo, E., Alegre, E., & de Paz, I. (2017). Classifying Illegal Activities on Tor Network Based on Web Textual Contents. En M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 35-43). Association for Computational Linguistics. https://aclanthology.org/E17-1004

Altman, D. G. (1990). *Practical Statistics for Medical Research*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429258589

Arslan, Y., Allix, K., Veiber, L., Lothritz, C., Bissyandé, T. F., Klein, J., & Goujon, A. (2021). A Comparison of Pre-Trained Language Models for Multi-Class Text Classification in the Financial Domain. *Companion Proceedings of the Web Conference 2021*, 260-268. https://doi.org/10.1145/3442442.3451375

Avarikioti, G., Brunner, R., Kiayias, A., Wattenhofer, R., & Zindros, D. (2018). *Structure and Content of the Visible Darknet* (arXiv:1811.01348; Número arXiv:1811.01348). arXiv. https://doi.org/10.48550/arXiv.1811.01348

Buldin, I. D., & Ivanov, N. S. (2020). Text Classification of Illegal Activities on Onion Sites. *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 245-247. https://doi.org/10.1109/EIConRus49466.2020.9039341

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2023). *Unleashing the potential of prompt engineering in Large Language Models: A comprehensive review* (arXiv:2310.14735). arXiv. https://doi.org/10.48550/arXiv.2310.14735

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. En J. Burstein, C. Doran, & T. Solorio (Eds.), *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)* (pp. 4171-4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Feuerman, M., & Miller, A. R. (2008). Relationships between statistical measures of agreement: Sensitivity, specificity and kappa. *Journal of Evaluation in Clinical Practice*, *14*(5), 930-933. https://doi.org/10.1111/j.1365-2753.2008.00984.x

Graczyk, M., & Kinningham, K. (s. f.). *Automatic Product Categorization for Anonymous Marketplaces*.

Huang, F., Kwak, H., & An, J. (2023). Is ChatGPT better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech. *Companion Proceedings of the ACM Web Conference 2023*, 294-297. https://doi.org/10.1145/3543873.3587368

Huang, F., Kwak, H., Park, K., & An, J. (2024). *ChatGPT Rates Natural Language Explanation Quality Like Humans: But on Which Scales?* (arXiv:2403.17368). arXiv. https://doi.org/10.48550/arXiv.2403.17368

Jin, Y., Jang, E., Cui, J., Chung, J.-W., Lee, Y., & Shin, S. (2023). DarkBERT: A Language Model for the Dark Side of the Internet. En A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 7515-7533). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.415

Jin, Y., Jang, E., Lee, Y., Shin, S., & Chung, J.-W. (2022). *Shedding New Light on the Language of the Dark Web* (arXiv:2204.06885; Número arXiv:2204.06885). arXiv. https://doi.org/10.48550/arXiv.2204.06885

Kalyan, K. S. (2024). A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal*, *6*, 100048. https://doi.org/10.1016/j.nlp.2023.100048

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., Bielaniewicz, J., Gruza, M., Janz, A., Kanclerz, K., Kocoń, A., Koptyra, B., Mieleszczenko-Kowszewicz, W., Miłkowski, P., Oleksy, M., Piasecki, M., Radliński, Ł., Wojtasik, K., Woźniak, S., & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, *99*, 101861. https://doi.org/10.1016/j.inffus.2023.101861

Krippendorff, K. (s. f.). *Computing Krippendorff's Alpha-Reliability*.

Li, B., Fang, G., Yang, Y., Wang, Q., Ye, W., Zhao, W., & Zhang, S. (2023). *Evaluating ChatGPT's Information Extraction Capabilities: An Assessment of Performance, Explainability, Calibration, and Faithfulness* (arXiv:2304.11633). arXiv. https://doi.org/10.48550/arXiv.2304.11633

Li, L., Fan, L., Atreja, S., & Hemphill, L. (2023). *«HOT» ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media* (arXiv:2304.10619; Número arXiv:2304.10619). arXiv. https://doi.org/10.48550/arXiv.2304.10619

Loukas, L., Stogiannidis, I., Malakasiotis, P., & Vassos, S. (2023). *Breaking the Bank with ChatGPT: Few-Shot Text Classification for Finance* (arXiv:2308.14634; Número arXiv:2308.14634). arXiv. https://doi.org/10.48550/arXiv.2308.14634

Powers, D. M. W. (2012). The Problem with Kappa. En W. Daelemans (Ed.), *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 345-355). Association for Computational Linguistics. https://aclanthology.org/E12-1035

Reiss, M. V. (2023). *Testing the Reliability of ChatGPT for Text Annotation and Classification: A Cautionary Remark* (arXiv:2304.11085; Número arXiv:2304.11085). arXiv. https://doi.org/10.48550/arXiv.2304.11085

Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet*, *15*(6), Article 6. https://doi.org/10.3390/fi15060192

Silva, Í., Marinho, L., Said, A., & Willemsen, M. C. (2024). Leveraging ChatGPT for Automated Human-centered Explanations in Recommender Systems. *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 597-608. https://doi.org/10.1145/3640543.3645171

Yang, K., Ji, S., Zhang, T., Xie, Q., Kuang, Z., & Ananiadou, S. (2023). *Towards Interpretable Mental Health Analysis with Large Language Models* (arXiv:2304.03347). arXiv. https://doi.org/10.48550/arXiv.2304.03347

Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., Cui, Y., Zhou, Z., Gong, C., Shen, Y., Zhou, J., Chen, S., Gui, T., Zhang, Q., & Huang, X. (2023). *A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models* (arXiv:2303.10420). arXiv. https://doi.org/10.48550/arXiv.2303.10420

# Mine Action Data Observatory: Tackling Uncertainty and System Longevity with Evolutionary Architecture

**Juraj Dončević, Kristina Ćaleta, Marko Horvat, Ivica Botički**
Faculty of Electrical Engineering and Computing, University of Zagreb
Department of Applied Computing
Unska 3, 10000 Zagreb
{juraj.doncevic, kristina.caleta, marko.horvat3, ivica.boticki}@fer.hr

**Abstract.** *This paper demonstrates the applicability of the concept of evolutionary architectures in the design process for a Mine Action Data Portal. Mine action refers to a range of activities aimed at reducing the threat posed by landmines, unexploded ordnance (UXO), and other explosive remnants of war (ERW). These activities typically include mine clearance, and raising awareness of the general public. The primary design concern for such a system is for it to be long-lasting. The design approach aims to future-proof the system under study by determining possible changes that could affect the system in the future. This is tackled by determining exact design choices as responses to the inferred changes. To avoid the danger of analysis-paralysis during the design process, the design approach acknowledges the existence of uncertainties and circumvents them by introducing adaptivity to the architectural design.*
**Keywords.** software architecture, evolutionary architecture, data platform, software design, mine action

## 1 Introduction

Mine action (MA) represents a multifaceted endeavor intertwining humanitarian assistance and developmental studies, focusing on eradicating landmines and mitigating the socio-economic and environmental repercussions of landmines and explosive remnants of war (ERW) (Maathuis, 2003). The overarching objective of mine action is to identify and mitigate the effects and risks posed by explosive hazards to ensure the safety of individuals. Beyond mere demining operations, mine action encompasses a spectrum of measures designed to safeguard communities, aid victims in achieving self-sufficiency and integration, and foster conditions conducive to stability and sustainable progress. Humanitarian demining, inclusive of activities such as mine and ERW surveys, land release procedures, mapping, marking, and clearance, constitutes a vital subset within the broader ambit of mine action endeavors (Bajić et al., 2011) (Matic et al., 2014).

A Mine Action Data Observatory is envisioned as a data portal supporting public information access, MA at operational and strategic levels, and informed decision-making. Its primary task is to manage and analyze vast amounts of data collected from mine-affected regions. The data portal could serve as a critical tool in not only storing and organizing data but also in processing it through advanced methods like optical character recognition (OCR), handwriting recognition (HWR), and information discovery-oriented artificial intelligence (AI) to extract actionable insights that guide mine clearance operations and policy-making. Previous research has already investigated the extent to which modern technological solutions can be used to support MA. It has been determined that ontologies (Horvat et al., 2024) (Horvat et al., 2022) and data lakes (Horvat et al., 2023b) are workable warehousing solutions for MA data observatories. Introducing blockchain technology to MA data observatories has also been proposed to offer tamper-proof data, trust, and transparent data access (Horvat et al., 2023a).

The Mine Action Data Observatory addresses the critical need for integrating diverse and heterogeneous data sources to enhance the efficiency and accuracy of humanitarian demining efforts. In mine action, the complexity and diversity of data gathered through different remote sensing methods, such as multispectral and hyperspectral imagery, ground-penetrating radar data (GPR), magnetic sensor readings, and unstructured documents like hand-drawn minefield maps and accident reports, present significant challenges in terms of data standardization and integration (Krtalić and Bajić, 2019)(Ibrahim et al., 2021). By utilizing formal knowledge representation methods such as the MINEONT+ ontology (Horvat et al., 2022), the proposed observatory semantically integrates these diverse data sets, transforming them into a structured and interoperable format. This approach simplifies the process of data consolidation and enables more efficient retrieval and analysis of relevant information, resulting in more efficient and streamlined decision-making in demining operations in terms of cognitive load and overall operating effort (Bajic, 2010) (Meurer et al., 2010) (Krtalić et al., 2018). To ensure the longevity and adaptability of the Mine Action Data Observatory, an evolution-

ary architecture is employed, emphasizing flexibility, scalability, and resilience, allowing the system to adapt to new technologies and changing requirements over time. This paper explores how evolutionary architecture can address these challenges, enhancing the observatory's adaptability, scalability, and resilience, thus ensuring its effectiveness and longevity in the dynamic field of humanitarian demining (Horvat et al., 2024).

The principles of evolutionary architecture (Ford et al., 2017) form the basis for the approach; allowing the isolation of changes to specific areas of the system without affecting others, facilitating smoother updates and enhancements.

The remainder of the paper is organized as follows; Section 2 offers the scope and context of an MA data observatory that will be the system under study in the design process. The Observatory's architecture is designed by focusing on system evolvability in Section 3. Future vision and future work are discussed in Section 4. Section 5 provides the conclusion to the paper's proposal.

## 2 Observatory scope and context

The Observatory's domain is centered around mine suspected areas (MSA) and the assortment of data gathered about them through mine records, interviews, monographs, MAs, MSA indicators, MSA interpretations by experts, and monographs (Fig.1). The presented entities are laden with geographical data, containing precise positioning of mission-critical objects and areas. The domain doesn't contain any structural complexities or elaborate branchings from the central MSA entity, therefore no detachments are envisioned in its physical implementation.

The proposed Observatory is envisioned as a data portal to facilitate both past and future data. The use cases for the Observatory outline two general purposes for the system; as an archival system handling historical data and an operational system handling recent data. The data is expected to be ingested into the observatory through a data portal which is implemented to support heterogeneous data ingestion.

The context of the system is presented in Fig 2., according to the C4 design model (Brown, 2010) (Richards and Ford, 2020). CMAC is expected to enter existing MA data into the observatory. This data is expected to predominantly consist of scanned physical documents of hand-filled forms such as mine action reports, interview minutes, testimonies, monographic documents, and GIS digital files. The greatest challenge in supporting this ingress data flow is combining object detection, optical character recognition (OCR), and handwriting recognition (HWR) to enable semi-autonomous document entry. These methods are required to ease the archival document entry process for system operators.

Data is also expected to ingress the observatory as current operational MA reports are entered by demining companies. The ingressing data should be entered through forms, enforcing standardization of both the data structure and format.

Consumed ingress data must be standardized regardless of its archival or operational nature - it is distilled into a unified model to support multiple consumption cases. The standardized data is expected to be used by MSA interpreters possessing expert domain knowledge. Interpreters are tasked with enriching the existing data and providing further insight into MA data. This includes expanding or correcting MSAs, or providing additional semantics to raw data to guide future MAs.

The general public will be served MA data through a read-only data portal that supports GIS visualization. Due to the expected sensitivity of some data, it should be anonymized and curated. If required, documents should be redacted by blackouts and the accompanying data anonymized.

To facilitate future information and knowledge discovery, the consumed and standardized observatory data will be used to train various AI models. These could be used in future decision support systems for minefield navigation or for guidance in MSA discovery where data is sparse or incomplete.

## 3 System capabilities and architectural design

Richards and Ford (Richards and Ford, 2020) claim that prioritizing system architecture is an effective way to tackle some requirements placed on a piece of software. Requirements are allowed to freely project themselves onto the architectural design of the Observatory.

The key requirement impacting the Observatory's architectural design is the need for the system to be long-lived due to its archival nature. This does not just encompass the deployability of the executable code, but also the maintainability of the Observatory's codebase. The primary goal is to implement a system that can be kept updateable, modifiable, and deployable not just in terms of years, but of decades. The Observatory must be able to evolve during its lifespan by tackling technological and domain changes; marked with the architectural meta-characteristic of evolvability. Evolvability is achieved by aligning a system architecture with the concept of evolutionary architecture (Ford et al., 2017).

Our research has determined that the dimensions of change that might impact the Observatory are:

- Changes in web or client-side application technology stacks;
- Changes in framework technology;
- Changes in data storage technologies;

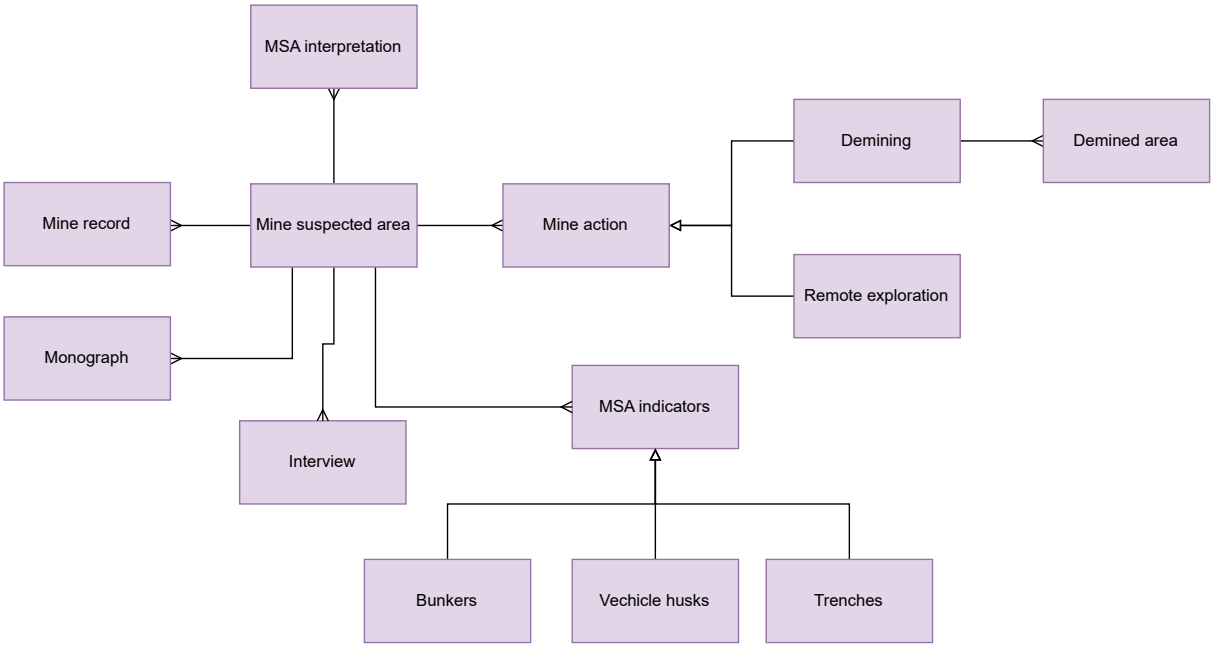



**Figure 1:** A simplified conceptual domain model for MA

- Changes in the document formats that need to be processed by the system;
- Changes in AI technology stacks and tools;
- Changes in data warehousing paradigms;
- Changes in load switching between operational and archival system context parts.

Determining impactful changes has provided a unique opportunity to push ahead the system architecture design in the overall design process. Along with the general requirements, the only specific detail known at the time was the conceptual domain model and that it didn't require the separation of the domain into separate bounded contexts - necessitating the decoupling within the domain model itself (e.g. via vertical slicing or microservices).

The architectural design was conducted earlier in the project life-cycle than usual but with an initiative to accommodate any changes in requirements without constraining the project. This allows managers, researchers, and developers to determine both certainties and uncertainties regarding the shape of the future system.

The certainties were represented by the information that the system would contain a separately implementable and deployable core service, a database, a file storage solution, and a previously implemented ontology knowledge database.

The uncertainties were tackled by separating the operational concerns of the Observatory into individual components. A general decision was made to treat the unspecified components as implementation details. The impact and propagation of their changes throughout the system would be minimized by interfacing. The interfaces are not determined ahead of the implementation process but are inferred only when implementing a specific functional point requires their actualization. Some notable uncertainties faced during the design were the technology for the client application, the technology of the file storage solution, and the specific technological implementation and time requirements of the HWR/OCR and AI model services.

The following design decisions were made to tackle the aforementioned changes that the system might have to evolve around:

- The client-side application was extracted into a separate executable component to support changes without impacting the core service.
- A core library and domain were implemented without 3rd party dependencies; effectively decoupling the domain from the framework used to implement the core service. The core domain is propped up by concepts of domain-driven design (DDD) (Evans, 2003) and functional programming (FP). DDD principles were used to design the domain and to characterize models as entities, aggregates or value objects. DDD effectively provided a systematization of the design process and detached domain considerations from technological specifics. FP was used to enable monadic chain calls, the control of exceptions and errors and the expression of nullability. In this way, technological features were detached from the implementation of core workflows, data exchange, framework usage, and error handling.
- All technological changes (including changes in data storage technology) are supported by the ports-and-adapters pattern (Cockburn, 2005). An interface rep-

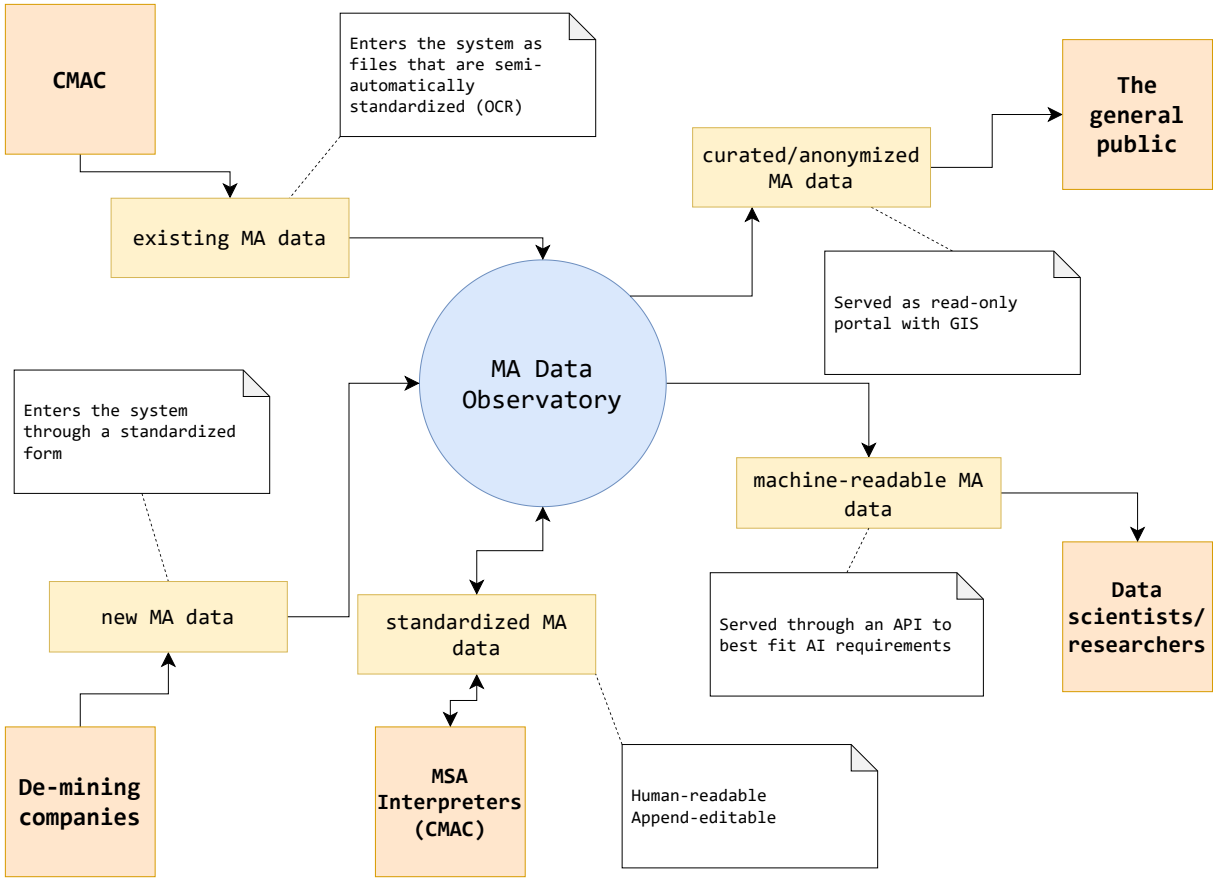



**Figure 2:** Observatory system context and data flow

resents a port, while its implementations represent adapters that can be swapped out by dependency injection.

- Changes in document processing are primarily supported by the extraction of OCR and HWR services from the core service; allowing the use of strategies for specialized physical document analysis.

- All AI-oriented solutions are extracted from the core service, allowing them to be deployed on specialized hardware optimized for machine learning (e.g. FPGA boards, GPUs, specialized CPUs).

- Data warehousing functionalities are extracted from the core service to allow technological flexibility of the warehousing solution. An unstructured data lake in formats like Avro or a structured star schema warehouse is expected.

- Since it was impossible to determine the computational load required from the core service at design time, the core service was split into two vertical slices (Newman, 2019). The split was not influenced by the domain, but by the dual system use cases - archival or operational system. Due to the possibility of the slices being split into separate microservice components, the WET (write everything twice) principle (Pai and Xavier, 2017) was adopted for everything except the core library.

Although evolutionary architecture proposes the use of event-driven architecture, this was not followed in this case to simplify orchestration due to the relatively small deployment scale (e.g., no need for service replication or redundancies). Conversely, an orchestration or choreography pattern would introduce architectural complexity in terms of an additional mediator component for the former, or multiple message queues in the latter case. Due to the context of the system, we found this trade-off to be unjustifiable, as the added complexity would not provide a significant benefit over the current design. Despite this, message queues are still utilized to provide asynchronicity for time-exhaustive operations. The case-specific message queues are also introduced to the core service through the ports-and-adapters pattern.

Fig 3. represents the container part of the C4 model and illustrates the described architectural design decisions. Although the system may be adapted to the cloud due to its containerization, its initial deployment might not be on the cloud. A simpler goal has been set for a proof of concept - the ability to deploy the system on a single via a Docker compose script. The system design even future-proofs the containerization technology by providing the ability to configure the system

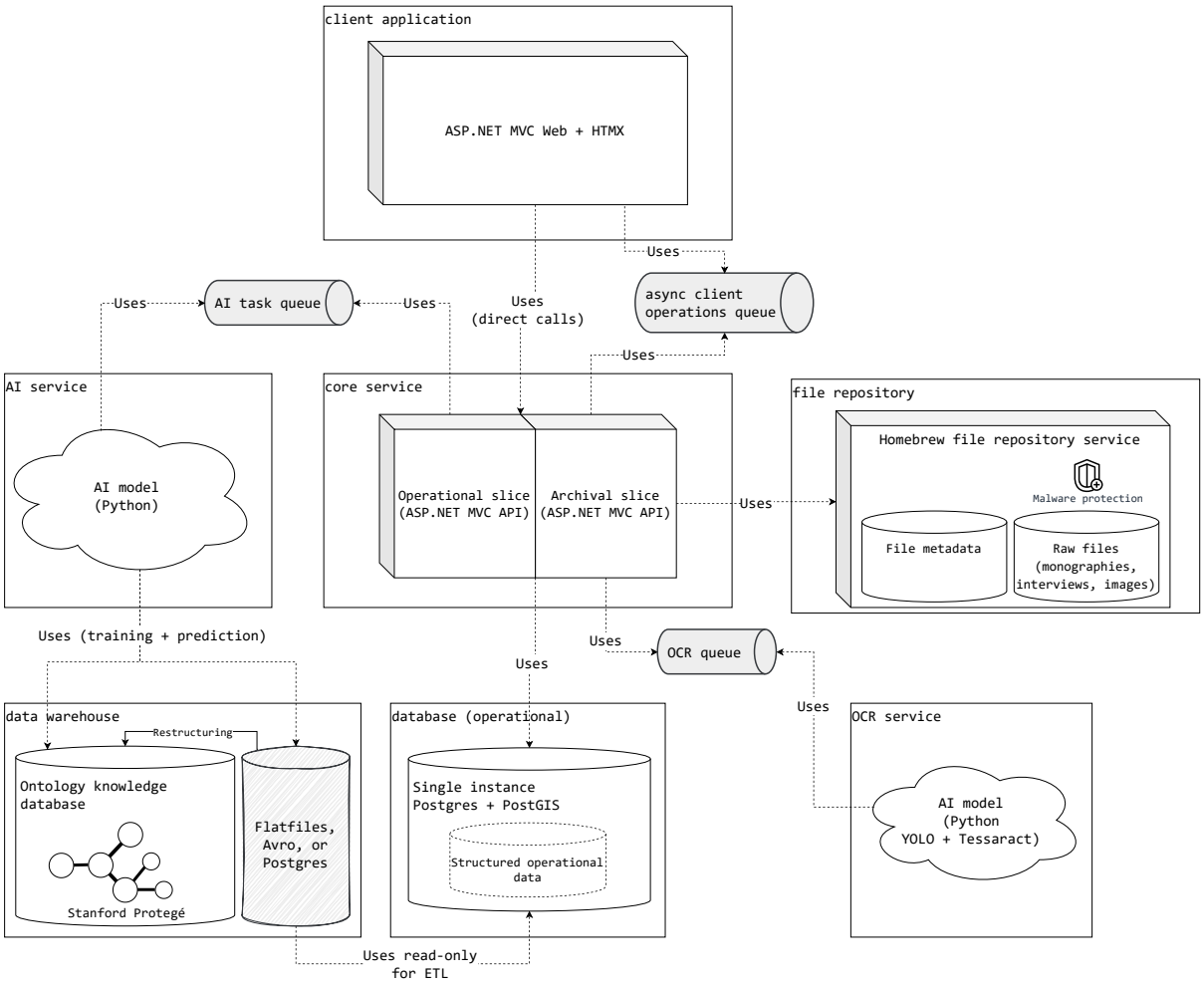



**Figure 3:** Observatory containers and component outline (and proposed technologies)

components through configuration files.

# 4 Future vision and future work

The Observatory system is designed to be robust, adaptable, and responsive to continuous change through evolvability. Future work will focus on the continuous observation of new technologies and trends in frameworks, data storage, AI, and OCR/HWR technologies. The Observatory offers a playground to test emerging technologies, as well as the ability to test its own evolvability. The assessment of these results through future research will provide more insight into appropriate strategies and techniques for future-proofing software systems.

It is expected that the Observatory, if successfully implemented, will not only meet the archival challenges in Croatia, but also in other countries with unsupervised MSAs or unfortunate countries with ongoing conflicts. The latter will be the real test of the Observatory's operational viability and longevity.

# 5 Conclusion

The development of the Mine Action Data Observatory represents a significant achievement in the realm of mine action initiatives, marking a decisive step towards a safer and more sustainable future. Through strategic design decisions following evolutionary architecture principles, a system can be crafted that not only addresses the immediate challenges of mine action but also embodies resilience and adaptability in the face of uncertainty. By designing a long-term solution resistant to change, we have fortified the observatory against unpredictable shifts in operational landscapes, technological advancements, and user requirements, ensuring its enduring relevance and effectiveness.

Furthermore, the Observatory's comprehensive scope underscores its pivotal role in addressing the multifaceted dimensions of mine action data. By standardizing data about demining operations the observatory empowers stakeholders with invaluable insights and resources to make informed decisions. Its capacity to facilitate transparent and accountable mine action efforts sets a new standard for the field, promot-

ing best practices and driving positive outcomes for affected communities worldwide.

The Observatory stands as a testament to the power of technology in support of lasting humanitarian endeavors. As challenges and opportunities emerge, the Observatory will continue to be incrementally designed and implemented. Work focusing on refining the system design in regard to deployment capabilities and orchestration is expected as it gets deployed in different varieties and physical locations. The Observatory will also provide a playground for experimenting with state-of-the-art technologies and principles.

## Acknowledgments

This work was supported by the Ministry of Science and Education of the Republic of Croatia grant: NPOO.C3.2.R3-I1.01.0268.

## References

Bajic, M. (2010). The Advanced Intelligence Decision Support System for the Assessment of Mine-suspected Areas. *The Journal of Conventional Weapons Destruction*, 14(3).

Bajić, M., Matić, , Krtalić, A., and Vuletić, D. (2011). Research of the Mine Suspected Area. *Research of the Mine Suspected Area*.

Brown, S. (2010). The C4 model for visualising software architecture.

Cockburn, A. (2005). Hexagonal architecture.

Evans, E. (2003). *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley Professional, Boston, 1st edition edition.

Ford, N., Parsons, R., and Kua, P. (2017). *Building Evolutionary Architectures: Support Constant Change*. O'Reilly Media, Beijing, 1st edition edition.

Horvat, M., Krmpotić, K., Krtalić, A., and Akagić, A. (2023a). Bridging Blockchain Technology and Humanitarian Demining: A Novel Concept for Decentralized Storage of Landmine and UXO Locations. pages 369–375. ISSN: 1847-2001.

Horvat, M., Krtalić, A., Akagić, A., Krmpotić, K., and Skender, S. (2023b). Humanitarian Demining Using Data Observatories and Data Lakes. *19th International Symposium "Mine Action 2023", Book of Papers*, page 31.

Horvat, M., Krtalić, A., Akagić, A., and Mekterović, I. (2024). Ontology-Based Data Observatory for Formal Knowledge Representation of UXO Using Advanced Semantic Web Technologies. *Electronics*, 13(5):814. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Horvat, M., Krtalić, A., Bajić, M., Muštra, M., Laura, D., and Gold, H. (2022). MINEONT: A proposal for a core ontology in the aerial non-technical survey domain. pages 47–51. ISSN: 1849-3718.

Ibrahim, N., Fahs, S., and Alzoubi, A. (2021). *Land cover analysis using satellite imagery for humanitarian mine action and ERW survey*.

Krtalić, A. and Bajić, M. (2019). Development of the TIRAMISU Advanced Intelligence Decision Support System. *European Journal of Remote Sensing*, 52:40–55.

Krtalić, A., Racetin, I., and Gajski, D. (2018). The Indicators of Mine Presence and Absence in Airborne and Satellite Non-technical Survey. pages 57–61.

Maathuis, B. H. (2003). Remote Sensing Based Detection of Minefields. *Geocarto International*, 18(1):51–60. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10106040308542263.

Matic, , Laura, D., Turšić, R., and Krtalić, A. (2014). *Analytical assessment for the process of collecting additional data on a suspected hazardous area in humanitarian demining*.

Meurer, H., Wehner, M., Schillberg, S., Raven, N., Wirtz, T., and Ime, F. (2010). An Emerging Remote Sensing Technology and its Potential Impact on Mine Action.

Newman, S. (2019). *Monolith to Microservices: Evolutionary Patterns to Transform Your Monolith*. O'Reilly Media, Beijing China ; Sebastopol, CA, 1st edition edition.

Pai, P. and Xavier, S. (2017). *.NET Design Patterns*. Packt Publishing Ltd. Google-Books-ID: t08oDwAAQBAJ.

Richards, M. and Ford, N. (2020). *Fundamentals of Software Architecture: An Engineering Approach*. O'Reilly Media, Sebastopol, CA, 1st edition edition.

# Attribute Based Access Control Metamodel for Spreadsheet Programs

**Miro Zdilar**
Faculty of Organization and Informatics,
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
mizdilar@student.foi.hr

***Abstract***. *Spreadsheets are one of the most popular tools used for various purposes in organizations. With recent technological advancement and new features added, spreadsheets have become powerful tool for complex analysis and modelling. However, spreadsheets are associated with high incidence of errors and unauthorized changes in multiuser environments causing companies significant losses. The goal of this work-in-progress paper is to present Attribute-Based Access Control metamodel for spreadsheet programs that empowers organizations to control unauthorized activities with spreadsheets in multiuser environments.*

**Keywords.** Authorizations, Attribute Based Access Control, Metamodel, Spreadsheets, Spreadsheet Error Detection, Spreadsheet Programs

## 1 Introduction

Spreadsheets are one of the most successful end-user programming systems used in business and academia (Panko, 2008). End-user programming systems empowers end-users to build and execute powerful computer programs without the use of traditional programming languages. In U.S. alone, it has been estimated that the number of end-user programmers outnumber traditional software programmers (Scaffidi et al., 2005). Spreadsheets are traditionally used in almost all companies in the U.S and Europe for accounting and financial reporting purposes (Panko & Ordway, 2008). Several researchers have studied spreadsheet usage patterns (Reschenhofer & Matthes, 2015) and they identified following business processes supported by spreadsheets; strategic and capacity planning, financial reporting, stakeholder analysis, risk management, performance calculation, data transformation, cash-flows analysis, time-series transformations, and simulations. Since the first introduction of electronic spreadsheets for personal computers in 1979, they have been developed as single user application targeted for personal computers. With recent technological advancement and new features added, spreadsheets have evolved to cloud-powered computational platform.

However, spreadsheets still lack effective access control functionality capable to adapt to technical and organizational dynamics of enterprises and growing complexity of spreadsheet programs. Even though spreadsheet allows Discretionary Access Control (DAC) at the level of cell and worksheet, this type of access control does not allow granularity for different roles, centralised administration with access policy and monitoring data flow becomes almost impossible as the spreadsheet program grows in complexity.

Uncontrolled access to spreadsheet programs and data have been linked with many spreadsheet errors causing companies and organizations significant financial losses and negative publicity. Non-profit and voluntary organization European Spreadsheet Risk Interest Group (EuSpRIG) maintains list of horror stories with details of spreadsheet errors and uncontrolled access to spreadsheet programs and data (European Spreadsheet Risk Interest Group, 2023).

This work-in-progress article presents Attribute Based Access Control (ABAC) metamodel for spreadsheet programs which can describe authorizations and its enforcement (access control) to prevalent spreadsheets, including dynamics of multiuser organizations. Therefore, the work presented in this paper answers following research questions:

- RQ1: What are common access control models for spreadsheets, both on technical level of modern cloud-powered spreadsheets and on organizational level within multiuser enterprises?
- RQ2: What is a suitable meta model to describe those access control models for spreadsheets?

The remainder of this paper is organized as follows. Section 2 provides summary of related work in the field of access controls models and their suitability for modern cloud-powered spreadsheets. In section 3, research method is presented for deriving ABAC metamodel for spreadsheet programs based on existing literature review, and architecture of modern cloud-powered spreadsheets. Afterwards, in section 4, the derived ABAC metamodel for spreadsheet programs is

presented with descriptions of design challenges and model components. In section 5, initial results of the evaluation of the proposed metamodel are presented. Finally, in section 6, conclusions and critical reflections on research and model evaluation are provided with proposals for future research opportunities.

# 2 Related Work

This paper followed critical review of the literature presented in other studies on spreadsheet errors (Powell et al., 2008). Rather than give a chronological account of the literature, discussion in this paper is focused around formulated research questions.

Authorization and its enforcement (access control) are key components of enterprise information technology systems (Korman, 2016). Focus of researchers have been around modelling different access control systems, evaluation and comparison of access control models deployed to various technical and operational environments, and formal verification of access control models in context of specific algorithms and protocols. Following is the summary of common access control models (Kashmar et al., 2021) that have been analysed for their suitability in development of access control metamodel for spreadsheets;

- Discretionary Access Control (DAC): is access control model built with three major components – objects, subjects, and permissions. DAC allows owners (subjects) to control permissions to their objects and is commonly implemented with Access Control List (ACL). DAC is implemented as integral part of many information technology systems, such as operating systems and databases. DAC is part of commercial spreadsheet implementation for decades and allows spreadsheet owner to control user's access to specific cell or worksheet. Issues with DAC have been studied by researchers (Downs et al., 1985) and this type of access control does not allow granularity for different roles, centralised administration with access policy and monitoring data flow becomes almost impossible as the spreadsheet program grows in complexity.
- Mandatory Access Control (MAC): is access control model managed in centralized manner and is built with four key components – a set of objects, a set of subjects, permissions, and security level. Even though MAC allows centralized policy management, it is very complex to implement this type of access control on all spreadsheet resources due to mandatory security level assigned to both subjects and objects. In addition, security levels specified in traditional MAC has been considered as antiqued (Dholakia, 2017) and do not seem to be in accordance with technical and organizational level of spreadsheet use in modern enterprises.
- Role-Based Access Control (RBAC): is access control model based on following key components – subjects, roles, permissions, actions, operations, and objects. In context of RBAC, role means a group of permissions to use object(s) and perform certain action(s). Only designated administrator has the right to control system security and manage roles assigned to users. RBAC implementation has been studied in hospital management where roles allow modeling complex relationships between doctors, nurses, etc. (Boadu, 2014). However, modeling and maintaining roles for different spreadsheet user roles and groups is very complex, especially in dynamic organizations where business processes and corresponding user roles are rapidly changing.
- Attribute-Based Access Control (ABAC): is access control model based on following three types of dynamic attributes – subjects, objects, and environments. User requests are resolved and determined based on subject attributes, objects attributes, environmental attributes as well as set of conditions specified by access policy. ABAC model is dynamic as it uses state of attributes at the time of access mode resolution. Even though we find limited literature on application of ABAC model to spreadsheets, we decided to further research applicability of ABAC to development of our meta model due to following promising characteristics of ABAC model:
  a. ABAC model is based on dynamic attributes, where object attributes fit to our proposed model of spreadsheet resources and corresponding attributes.
  b. Hierarchy of spreadsheet resources and objects can be modelled with set of ABAC conditions and access rules determinations. This property prevents conflicts in access resolutions and simplifies prototype implementation.
  c. Deployment opportunities for ABAC with spreadsheets are flexible and allows early prototype implementation as detective access control system. This minimizes impact on users and familiarized spreadsheet user interface.
  d. Complexity of ABAC model for spreadsheets depends on number of spreadsheet resource attributes.
  e. Dynamic nature of modern cloud-powered spreadsheets and extensions to spreadsheet formula language fits nicely to ABAC dynamic attribute concept. Potential new functionalities and modules added in cloud-powered spreadsheet can be integrated within existing ABAC concepts.

Unified metamodel of enterprise authorizations (Korman, 2016) is summarizing existing models of access controls. In addition, authors provided mapping between presented unified metamodel and ArchiMate tool that is frequently used in modern enterprises as architecture modeling language. List of generic metamodels for expressing different configurations of access models is valuable starting point for our research and design of ABAC metamodel for spreadsheets.

ABAC modelling and implementation has been recognized by U.S. government as important access control modelling concept for large enterprises and federal information technology systems. National Institute of Standards and Technology (NIST) published in 2014 Special Publication 800-162 " Guide to Attribute Based Access Control (ABAC) Definition and Considerations" (Hu et al., 2014). This publication provides definitions and considerations for using ABAC to improve information sharing, design of systems, while maintain control of that information. Concepts and terminology for ABAC presented in this document has been instrumental for design of our ABAC metamodel for spreadsheets.

To further explore our second research question, in continuation of this literature review we focused our discussion on spreadsheet modelling. Popularity and success of spreadsheets in various domains triggered significant interest of research community (Jannach et al., 2014,), including visualization techniques, static code analysis and reporting, testing approaches, automated fault localization and repair, model-driven development, and design support.

In one of the first publications structured around spreadsheet modelling (Isakowitz, 1995), authors presented model of spreadsheet with separation between logical and physical view. Presented spreadsheet model consists of four key components- "schema" which captures spreadsheet program's logic, "data property" which holds input values of the input cells, "binding property" which maps logical view of schema and data to two-dimensional spreadsheet grid, and finally "editorial property" which provides visual description for headings, labels, tables and documentation on two-dimensional spreadsheet grid. With the help of tools, logic can be extracted from simpler spreadsheet programs and presented in "schema" component with relational logic model. Ideas presented in this paper inspired our work and research related to spreadsheet modelling.

Formal set-oriented spreadsheet modelling approach has been presented by authors as part of their effort to research type inference for spreadsheet programs (Abraham, 2006). A spreadsheet is defined as collection of formulas and values embedded into a spatial structure. Authors provided generalization of spatial structure where two-dimensional grid is just one possible representation. In this paper, spreadsheet formula language is formalized with grammar based on set-oriented and logical representation. Concept of spreadsheet types introduced is still relevant and applicable to modern cloud-based spreadsheets. Ideas presented by authors has been influential in design of presented ABAC metamodel for spreadsheets, specifically in design of access rule resolution and prototype implementation.

The most influential conceptual model of spreadsheet for research presented in this paper have been published by authors researching complexity metrics for spreadsheets (Reschenhofer at al., 2017). Presented spreadsheet metamodel captured and integrated all relevant aspects of spreadsheet, including spreadsheet formula language. Main idea presented by authors has been further explored in this work with inclusion of modern cloud-based spreadsheets.

# 3 Research Method

Work in this paper follows Design Science Research (DSR) approach (Hevener et al., 2004). DSR is structured around three interconnected components or research phases that should jointly deliver design science research artefact. Specifically, application of DSR to this research resulted with following method and research phases:

- Environment – defines set of requirements to the design artifact. In this research proposed ABAC metamodel for spreadsheet represents theoretical foundation for addressing business needs for spreadsheets in multiuser environments and reduction of unauthorized change and spreadsheet errors. Proposed metamodel is agnostic to actual spreadsheet commercial products. In addition, proposed metamodel should address the needs of different business domains and organizations.
- Knowledge Base – we followed focused approach to related work and literature review and actual knowledge base for this work is presented in Section 2.
- Design Science Research – main deliverable of this research is ABAC metamodel for spreadsheets which was designed with iterative approach and continuous refinements and verifications. In addition to design of research artifact, our work contributes with new additions to spreadsheet and access control knowledge base. Our work has been influenced by model of spreadsheet with separation between logical and physical view (Isakowitz, 1995), formal set-oriented spreadsheet modeling approach (Abraham, 2006), conceptual model for measuring the complexity of spreadsheets (Reschenhofer at al., 2017) and unified metamodel for enterprise authorizations (Korman, 2016). The evaluation as part of design science research was done by applying the

designed ABAC metamodel for spreadsheets on use case in multiuser environment.

# 4 ABAC metamodel

According to research method described in Section 3, we designed ABAC metamodel for spreadsheets which is capable to control unauthorized activities with spreadsheets in multiuser environments. Due to limited space in this publication, basic model is depicted in Fig. 1, spreadsheet formula model is further explained in Fig. 2 and lastly unified view to key components of ABAC metamodel for spreadsheet is depicted in Fig 3.

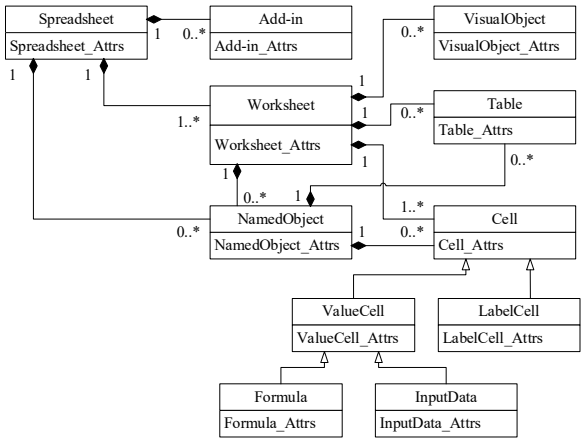



**Figure 1.** Metamodel of spreadsheet resources and associated attributes

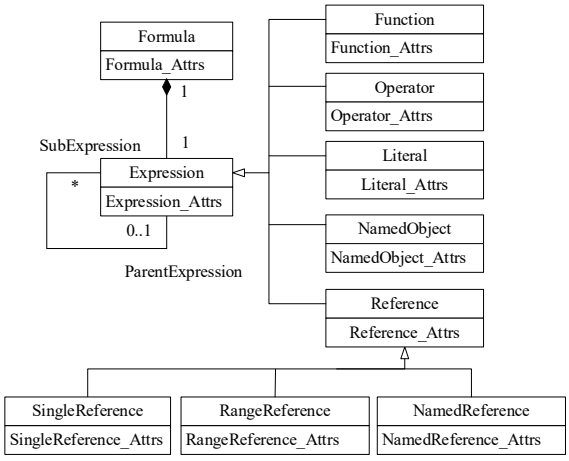



**Figure 2.** Spreadsheet formula metamodel (Continuation from Figure 1)

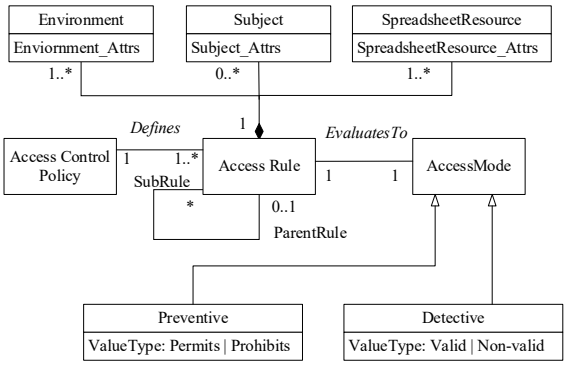



**Figure 3.** Proposed ABAC metamodel for spreadsheets

Design considerations in proposed metamodel are structured around resources that constitute modern cloud-powered spreadsheet. Spreadsheet resources are building blocks for spreadsheet programs and are manipulated by spreadsheet users or change their state during lifecycle of spreadsheet as result of spreadsheet program execution. Spreadsheet resources and their attributes are bounded with ABAC rules and permissible actions performed by spreadsheet users. Key spreadsheet resources unique to proposed ABAC metamodel for spreadsheets are following:

- The class *NamedObject* represents resource in modern spreadsheets that is controlled by user through Named Manager functionality. This naming convention is used in Micorsoft Excel 365 (Excel) product, but other cloud-based spreadsheets offer similar functionality. With recent introduction of *LAMBDA* functions (Gross, 2024) and collection of supporting functions (*MAP, REDUCE, SCAN, MAKEARRAY, BYROW, BYCOL, ISOMITTED)*, Excel is additionally empowered for computational tasks previously reserved for plugins or scripting with embedded macro language. Great demonstration of new Excel capabilities and powerful development strategies using only Excel formulas has been provided by Bartholomew at EuSpRIG conference (Bartholomew, 2023). Lambda functions are essential for creation of reusable software components, empowering spreadsheet developers to develop models quicker with fewer errors (Hatmaker, 2023). Other exciting new feature added to Excel is native support for Python programming language (Microsoft Excel, 2023). Powerful Python computational engine is embedded in Excel and users can integrate Python language code with existing formula language at the level of cell. Python in Excel is compatible with existing tools and libraries for charting and numerical analysis. Among many exciting new features that Python in Excel offer to users is ability to create and dynamically control complex tabular and visual objects directly from python code. This functionality was previously available only through Excel predefined toolbar. The class

*NamedObject* represents all these powerful resources added to Excel and through ABAC rules allows controlled access by users.

- The class *Table* represents Excel named table resources that could be added to worksheet through application interface. However, after manual creation of named table, this spreadsheet resource can be controlled directly from spreadsheet formula language. Structured references are powerful extension to formula language and allows spreadsheet users to reference in their code entire table, columns, rows, and table ranges. Through designated attributes, named table manipulations by users can be controlled with ABAC rules.
- The class *Expression* represents powerful spreadsheet language ability for nesting formulas. Original design introduced by Retschenhofer et al. (Reschenhofer at al., 2017) have been extended to include *NamedObject* as parameters and *NamedReference* as one possible realization of cell references.
- Spreadsheet resources represented in proposed metamodel contains associated set of attributes. We used simple naming convention and set of attributes are represented with suffix *_Attrs* concatenated to Class name that represents spreadsheet resource. For example, set of attributes associated with class Worksheet is represented in model as *Worksheet_Attrs*. Depending on nature and characteristics of modeled spreadsheet resource, corresponding attributes are represented with enumerated lists or key-value HashMap. For example, attribute *type* for *InputData* is represented with enumerated list *[Boolean, Integer, Number, String, Date, Array]* (Reschenhofer at al., 2017). Attribute name for Worksheet is represented with key-value HashMap *{"name":"Worksheet_Name"}*.
- The class *AccessMode* is generalization of *Preventive* and *Detective* classes. In optimal deployment case, proposed model should be implemented as preventive ABAC control system that prevents users for performing unauthorized actions. However, proposed ABAC metamodel can be deployed as detective control mechanism to determine validity of performed user actions with spreadsheet under control.

## 4.1 Access Rules

In proposed ABAC metamodel, access rules are modelled as quadruple with following structure:

$$<S, A, SR, E> \quad (1)$$

*S* represents set of spreadsheet users or roles of the spreadsheet user. Theoretically, *S* represents any *Agent* that might interact with spreadsheet and controlled access is required. For example, other IT system might access spreadsheet via predefined interface or background job might update spreadsheet input data during predetermined period of times.
*A* is set of actions that subject might perform on spreadsheet resource represented with following enumerated list:

$$A \in [CREATE, READ, UPDATE, DELETE] \quad (2)$$

Above enumerated list is well known in computer science as CRUD acronym. Important to note is that proposed ABAC metamodel does not have restrictions to number of actions and if needed in specific deployment scenarios, number of actions could be reduced or extended.
*SR* represents set of spreadsheet resources and corresponding resource attributes on which subject *S* can perform action *A*.
*E* are dynamic conditions, independent of subject and spreadsheet resources that may be used as attributes at decision time to influence an access decision. Examples of environment conditions include time, location, threat level, and temperature (Hu et al., 2014).

## 4.2 Access Mode Determination and Conflict Resolution

In order to formalize access mode determination in proposed metamodel and prevent conflict in access rules resolution, we introduce following definitions:

**Atomic Access Rule (AAR):**
AAR is quadruple of the from (1);

$$<S_i, A_j, SR_k, E_l> \quad (3)$$

Where $S_i \in S$, $A_j \in A$, $SR_k \in SR$, $E_l \in E$.

□

**Composed Access Rule (CAR):**
CAR is conjunction statement composed with AARs:

$$AAR_1 \wedge AAR_2 \wedge AAR_3 \wedge \ldots \wedge AAR_n \quad (4)$$

□

**Access Policy (AP):**
AP is set of all CARs applicable to single organization, and all its users and spreadsheet programs.

□

**Deny-by-default Policy (DP):**
If not explicitly defined in AP, all actions for all users and all spreadsheet resources are prohibited.

□

**Priority of Actions (PA):**
Actions should be evaluated in following order:

$$DELETE > CREATE > UPDATE > READ \quad (5)$$

□

Delete action has highest priority and operator > in context of above relation should be interpreted as "Allow to Perform". For example, if access rule permits user to delete specific spreadsheet resource, according to above definition, user is also allowed to create, update, and read corresponding resource.

**Rule Evaluation Hierarchy (REH):**
All child spreadsheet resources inherit rules applicable to their parents.

□

Proposed metamodel is structured around natural hierarchy of spreadsheet resources. For example, *Spreadsheet* class contains ("HAS A" relationship), *Add-in*, *Worksheet* and *NamedObject* classes, and all applicable access rules for *Spreadsheet* are inherited by its children, *Add-in*, *Worksheet* and *NamedObject*. Determining hierarchy for dynamic spreadsheet resources, such as composed formula expressions and large data tables is challenging, however strict conformance with REH rule during all stages of spreadsheet lifecycle is important for consistent rule's evaluation and user's activity control.

# 5 Model Evaluation

Proposed ABAC metamodel for spreadsheets has been evaluated on use case within analytical laboratory. In order to minimize impact on established organizational processes and spreadsheet use patterns, we deployed ABAC detective controls to determine validity of performed user actions with spreadsheet under control.

In agreement with laboratory management, we selected spreadsheet program for Negative Temperature Coefficient (NTC) probe calibrations. NTC spreadsheet supports important laboratory processes for calibration and management of 65 temperature probes utilized for temperature measurement of air, surface or liquids. Template spreadsheet "NTC" used for calibration of NTC probed in analytical laboratory is depicted in Figure 4.

**NTC Calibration**

**Constants**

| | | |
|---|---|---|
| R0 | 1000 | Ohm |
| Beta | 3930 | |

**Input**

| | | |
|---|---|---|
| t0 | 293,15 | K |
| t1 | 373,15 | K |

**Results**

| | | |
|---|---|---|
| R1 | =R0*EXP(Beta*(1/t_1-1/t_0)) | Ohm |

NTC

**Figure 4.** Worksheet "NTC" used for ABAC model evaluation.

NTC probe calibration spreadsheet is used on daily basis by different team members. Every Monday, calibration expert performs checks and potentially calibrations on all 65 probes utilized in laboratory. One spreadsheet instance is managed for each calibration probe. Laboratory Manager reviews all calibration spreadsheets and if results of NTC probe calibration complies with laboratory guidelines, Manager changes colour of result cell and corresponding worksheet to green as evidence of review and compliant status of performed calibration. Lastly, on Tuesdays, laboratory administrator edits header label and adds information about probe serial number, date of calibration, prints calibrations results on preformatted stickers and attaches them on probe housing. All calibrated probes should be ready for

Based on above process description following sentences in natural language describes user's role:

- Calibration Expert is responsible for calibration of NTC probes. Calibration should be performed every Monday at the beginning of working week. Results of checks and calibrations should be recorded in dedicated spreadsheet for each calibrated NTC probe, in designated cells for calibration results. Other elements of spreadsheet should remain unchanged.
- Laboratory Manager is responsible for reviewing and approving performed NTC probe calibrations. As evidence of approval and satisfactory calibration results, Manager should change color of result cell and of NTC worksheet to green. Yellow color should be used if probe requires recalibration and red color if probe should be inspected by technician or discarded for future use.
- Laboratory administrator is responsible to edit header of all green (approved) spreadsheets every Tuesday, add probe serial number and date to worksheet header label, print final worksheet on sticker and apply sticker to probe housing. This activity should be completed by end of Tuesday.

Based on above role descriptions in natural language we translated them to ABAC access policy (AP) for three different user roles within analytical laboratory.

*<expert, update, R0, Monday>∧*
*<expert, update, Beta, Monday>∧*
*<expert, update, t0, Monday>∧*
*<expert, update, t1, Monday>*

*<manager, update, R1.color="Green", Monday>∧*
*<manager, update, R1.color="Yellow", Monday>∧*
*<manager, update, R1.color="Red", Monday>∧*
*<manager, update, NTC.color="Green", Monday>∧*
*<manager, update, NTC.color="Yellow", Monday>∧*
*<manager, update, NTC.color="Red", Monday>*

*<admin, update, header1.type="Date", Tuesday>∧*
*<admin, update, header2.type="String", Tuesday>*

To align with provided definitions for access mode determination and conflict resolution, above policy is structured with conjunction statements for each user's role that participate in evaluation.

In continuation of evaluation process, we requested analytical laboratory team to continue with their standard weekly procedures and provide us access to all spreadsheet instances generated during the weekly calibration procedures. From the perspective of laboratory users, our evaluation had no impact to their well-established procedures and routines.

To determine users' activity and generate dynamically hierarchy of spreadsheet resources under evaluation we utilized Abstract State Machine (ASM) model for spreadsheets (Zdilar, 2023). ASM model was instrumental to identify all manipulations to spreadsheet resources including formula expressions. ASM has been implemented in Python 3 programming language (Van Rossum & Drake, 2009), with OpenPyXL library utilized for parsing spreadsheet formulae (Zumstein, 2021). ASM model generates directed graph as result of changes performed by users in spreadsheet under evaluation. Edges in graph represents state transitions and corresponds to changes and modifications to spreadsheets resources performed by users. We further utilized NetworkX python library (Hagberg et al. 2008) to generate graph queries and compare identified user changes with allowed actions in defined access policy.

We successfully parsed and analysed all 65 spreadsheets generated by laboratory users during one week of calibration activities. Due to experience of laboratory stuff, only two minor unauthorized changes have been detected by proposed model – cell colour in one spreadsheet was not properly changed and laboratory administrator completed printout of spreadsheet stickers following working day due to other priorities. To confirm correctness of determined changes by our proposed ABAC model, we manually review results for all 65 laboratory spreadsheets.

# 6 Conclusion and Future Research

In this paper we have presented initial results of our research on conceptual modelling of access control for spreadsheets. We structured our research around Design Science paradigm and two formulated research questions. We believe that proposed Attribute-Based Access Control metamodel for spreadsheets and results of its evaluation provides initial answers to formulated research questions. Our initial evaluation in analytical laboratory provided valuable feedback for proposed ABAC metamodel. However, our evaluation is limited to one use case and in our future research we will conduct model evaluation in more comprehensive case studies and large multi-user environments. We will also explore opportunities to apply formal verification techniques for model requirements and correctness of access mode determination and rule conflict resolution. Based on valuable feedback from evaluation participants, we noticed that generation of machine-readable access rules and translation from users' roles documented in natural language is challenging process. We will explore opportunities to automate this process with minimal manual intervention and represent access rules with standardized formats. We focused our initial research on key components of ABAC metamodel, and deployment aspects documented in NIST Special Publication 100-162 (Hu et al., 2014) were out of scope in this research.

# 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. (2006, July). Type inference for spreadsheets. *In Proceedings of the 8th ACM SIGPLAN international conference on Principles and practice of declarative programming* (pp. 73-84).

Bartholomew, P. (2023). Excel as a Turing-complete Functional Programming Environment. *arXiv preprint arXiv:2309.00115*.

Boadu, E. O., & Armah, G. K. (2014). Role-based access control (RBAC) based in hospital management. *Int. J. Softw. Eng. Knowl. Eng, 3*, 53-67.

Dholakia, Y. (2017). Mandatory Access Control – Problems in it and propose a model which overcomes them. *International Research Journal of Engineering and Technology (IRJET), (4)4*, pp.2031-2035

Downs, D. D., Rub, J. R., Kung, K. C., & Jordan, C. S. (1985, April). Issues in discretionary access control. In *1985 IEEE symposium on security and privacy* (pp. 208-208). IEEE.

European Spreadsheet Risk Interest Group. (2023). EuSpRIG Horror Stories. https://eusprig.org/research-info/horror-stories/ (June 10, 2024)

Gross, C. (2021). Announcing LAMBDA Helper Functions: Lambdas as arguments and more. https://techcommunity.microsoft.com/t5/excel-blog/announcing-lambda-helper-functions-

lambdas-as-arguments-and-more/ba-p/2576648 (June 10, 2024)

Hatmaker, C. (2023). Reducing Errors in Excel Models with Component-Based Software Engineering. *arXiv preprint arXiv:2309.00650*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science-Hevner. Design Science in Information Systems Research. *MIS Quarterly, 28 (1)*, 75–105.

Hu, V. C., Ferraiolo, D., Kuhn, R., Schnitzer, A., Sandlin, K., & Scarfone, K. (2014). Sp 800-162. guide to attribute based access control (abac) definitions and considerations. *Nat'l Inst. Standards and Technology*, 800-162.

Isakowitz, T., Schocken, S., & Lucas Jr, H. C. (1995). Toward a logical/physical theory of spreadsheet modeling. *ACM Transactions on Information Systems (TOIS), 13*(1), 1-37.

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.

Kashmar, N., Adda, M., Atieh, M., & Ibrahim, H. (2021). A review of access control metamodels. *Procedia Computer Science, 184*, 445-452.

Korman, M., Lagerström, R., & Ekstedt, M. (2016). Modeling enterprise authorization: a unified metamodel and initial validation. *Complex Systems Informatics and Modeling Quarterly*, (7), 1-24.

Microsoft Excel. (2023). Announcing Python in Excel: Combining the power of Python and the flexibility of Excel. https://techcommunity.microsoft.com/t5/excel-blog/announcing-python-in-excel-combining-the-power-of-python-and-the/ba-p/3893439 (June 10, 2024)

Panko, R. R. (2008). Spreadsheet errors: What we know. what we think we can do. *arXiv preprint arXiv:0802.3457*.

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.

Reschenhofer, T., Waltl, B., Shumaiev, K., & Matthes, F. (2017). A conceptual model for measuring the complexity of spreadsheets. *arXiv preprint arXiv:1704.01147*.

S Scaffidi, C., Shaw, M., & Myers, B. (2005, September). Estimating the numbers of end users and end user programmers. In *2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'05)* (pp. 207-214). IEEE.

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. *CreateSpace*.

Zdilar, M. (2023). Towards Automated Detection of Qualitative Spreadsheet Errors in Multi-user Environments. In *Central European Conference on Information and Intelligent Systems* (pp. 419-424). Faculty of Organization and Informatics Varaždin.

Zumstein, F. (2021). *Python for Excel: A Modern Environment for Automation and Data Analysis* (pp. 155-179). O'Reilly Media.

# Using AI on Open Data: A Systematic Literature Review

**Sara Kijanović, Darko Stefanović, Dušanka Dakić, Teodora Vučković**

University of Novi Sad, Faculty of Technical Sciences, Serbia

Department of Industrial Engineering and Engineering Management

Trg Dositeja Obradovica 6, Novi Sad, 21000

{sara.kijanovic, darko.stefanovic, dakic.dusanka, teodora.vuckovic}@uns.ac.rs

***Abstract****. Open Data, publicly available in machine-readable formats, is a crucial resource driving transformative changes globally. This research uses a systematic literature review to explore the synergies between Open Data and the growth of artificial intelligence, including machine learning and deep learning. The goal was to investigate how integrating Open Data with AI/ML/DL techniques creates intelligent industry solutions. Using the PRISMA framework, the paper analyzes key findings, highlighting Open Data's impact on technological innovation. It also identifies primary data sources for these studies and provides background insights, methodology, results, and conclusions based on the review's insights.*

***Keywords.*** *open data, open government data, artificial intelligence, machine learning, deep learning, literature review*

## 1 Introduction

According to the Law on Electronic Administration of the Republic of Serbia (*Zakon o elektronskoj upravi | Paragraf Lex propisi*, 2018), and also the Open Data Handbook (*What Is Open Data?*, 2024), Open Data (OD) are data that are publicly available for reuse, along with metadata, in a machine-readable and open format. Accessible worldwide, open data empowers governments, businesses, and entrepreneurs to ignite healthcare, education, and agriculture revolutions. These freely available data, accessible online, play a pivotal role in steering social and economic transformations. Innovations propelled by data contribute to heightened government transparency and efficiency, nurture business prospects, generate substantial economic benefits, and address urgent social challenges (*Open Data and AI: A Symbiotic Relationship for Progress | Data.Europa.Eu*, 2023).

We are witnessing the exponential development of artificial intelligence and various segments of AI, such as machine learning (ML), deep learning (DL), natural language processing (NLP) etc. Today, these technologies enhance and boost all aspects of human life. In the context of large amounts of data, advanced methods and techniques can be used for the analysis of such data, obtaining meaningful outputs, as well as intelligent solutions that can find practical applications in everyday life. For instance, advanced technologies can significantly support business processes and process mining, given that this field requires extracting knowledge from vast amounts of data (Dakic et al., 2019).

As mentioned, open data holds tremendous potential as a source of information for creating solutions that can improve all aspects of life. In scientific literature, many authors encounter the problem of data scarcity, which slows down their research. In many scientific papers within the context of humanities and medical sciences, authors such as (Adler et al., 2022; Norori et al., 2021), emphasize the importance of sharing data and materials for the easier implementation of technologies and the detection of anomalies, leading to faster treatments. The authors (Rosés et al., 2021) emphasize the importance of opening various datasets for further research to make progress in the field of crime detection more efficient. This is done to enable the development of spatial crime simulations that contribute to understanding the mechanisms that drive crime (Rosés et al., 2021).

This paper represents a systematic literature review (SLR) providing insights into the current state of the field in the context of open data usage, exploring where researchers gather data and which contemporary techniques from the AI, ML, and DL domains they employ to create intelligent solutions. To achieve the defined research objectives, SLR was conducted using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The main contribution of this paper is to gain insights into the previous application of AI/ML/DL methods and techniques on OD, as well as to identify the primary data sources for these studies over 5 years (2018-2023). Consequently, this may be significant for researchers in the future when it comes to easier identification of suitable datasets. Key research questions were formulated,

along with inclusion and exclusion criteria for the works. The results are presented in the paper.

The rest of the paper is organized into the following sections: Section 2 introduces the background, providing an overview of the fundamental concept of OD and advanced methods and techniques in the context of AI/ML/DL. Section 3 presents related work, Section 4 describes the methodology, Section 5 presents the research results, then Section 6 presents a discussion. Finally, in Section 7, conclusions are drawn based on the results obtained and the review conducted.

# 2 Background

As mentioned previously, Open Data (OD) represents data that is publicly available, accompanied by metadata, in formats that enable its use and reuse. The OD initiative fosters the development of various sectors and promotes data accessibility for everyone's use. Its goal is to encourage transparency and enable equal participation in problem-solving and innovation for all. OD can be used in many fields, and most importantly in the public sector, like public finance researched in (Komosar & Kijanovic, 2023), and education (Kijanovic & Komosar, 2023). Many papers, like those mentioned, provide background on the importance of using OD and highlight the significance of utilizing OD for various purposes.

Governments worldwide often release data, known as Open Government Data (OGD). Many researchers utilize open data from various platforms, and the authors in the paper mention that frequently used platforms include Kaggle, GitHub, Socrata, CKAN, DKAN, etc. (Ali et al., 2022).

The latest reports from the European Union emphasize the immense potential of artificial intelligence, especially when combined with open data. This synergy unlocks new possibilities, not only for extracting fresh insights from open data but also for leveraging AI in novel applications (*Open Data and AI: A Symbiotic Relationship for Progress | Data.Europa.Eu*, 2023).

Open data enables researchers to utilize extensive datasets for testing and scaling their studies and algorithms, overcoming the challenge of accessing real databases. The lack of access or insufficient data can significantly hinder the development of data-driven technologies while harnessing this power can enhance various industrial processes.

Data and technology enhance the industry. Authors (Lopes et al., 2024) stress the importance of precise production data for digital twin models. They tackled data scarcity by simulating production lines and generating artificial data to support deep neural network algorithms. Open data not only promote transparency and collaboration in the research community but also serves as a valuable foundation for advancing industry technologies.

In the subsequent work, insights into the current state of utilizing AI, ML, or DL with OD, using publicly available and open datasets, will be aimed to be gained.

# 3 Related Work

The paper (Hurbean et al., 2021) conducts an SLR on the integration of OD and ML in smart cities (SC). It examines ML applications across six smart city domains using OD, identifying trends, challenges, and potential solutions. It concludes that OD-based ML applications show promise in addressing urban challenges but also highlight data quality and consistency challenges.

The authors in (Wirtz et al., 2022) explore the link between open government data (OGD) and the digital economy, emphasizing OGD's potential for innovation and economic growth. Conducting an SLR synthesizes existing knowledge into a theoretical framework and suggests a research agenda for information systems and digital business research. Its conclusions stress OGD's importance for driving innovation and economic value, highlighting a research gap in integrating OGD concepts with digital business research.

The paper (Brinkhaus et al., 2023) concludes that the future of AI applications in chemistry relies heavily on open data initiatives. It emphasizes the necessity of open data infrastructures. Additionally, it underscores the importance of open-source software and molecular string representations in advancing AI research in chemistry. The study suggests that sharing curated data and trained models with the public will accelerate growth in the field. It calls for maintaining high data standards and leveraging public cloud infrastructures to expedite advancements in AI-powered molecular informatics.

The presented papers provide insights and answers to research questions that are concretized through specific examples. Given that no literature review has been conducted so far that could cover general questions applicable to each area, one has been carried out below.

# 4 Methodology

A systematic literature review (SLR) serves the purpose of systematically analyzing, synthesizing, and evaluating available research and papers in a specific field. This approach allows researchers to create a comprehensive overview of relevant studies and literature on a particular topic. An SLR aids in identifying key findings and shortcomings in research and provides a foundation for further investigation.

For this SLR, we used methodology based on Barbara Kitchenham's guidelines for performing Systematic Literature Reviews (Kitchenham &

Charters, 2007), and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). According to (Moher et al., 2009), it consists of four phases: Identification, Screening, Eligibility, and Included phase.

### 4.1 Search Strategy

We used the Scopus index database for this literature review with a query string:

```
("artificial intelligence" OR "AI" OR "Machine learning" OR "ML" OR "Deep Learning" OR "DL")
AND
("Open Data" OR "Open Government Data")
AND
PUBYEAR > 2017
```

Access was granted to open-access papers, meaning they are under a specific Creative Commons license.

### 4.2 Research questions and inclusion and exclusion criteria

To gain a clear understanding and achieve the objective of this review, we have formulated four research questions:

1. Which methods and techniques of AI/ML/DL are predominantly used with OD? (RQ1)
2. Which databases are most commonly used for downloading OD? (RQ2)
3. What are the strengths and limitations of different AI techniques in OD? (RQ3)
4. What is the main purpose of research and the use of AI/ML/DL methods and techniques on the utilized open data? (RQ4)

To enhance the SLR's effectiveness, we defined inclusion and exclusion criteria. The inclusion criteria are:

1. The paper must primarily focus on applying various AI, ML, and DL methods and techniques to OD. (IC1)
2. The paper must contain the source of the Open Dataset, e.g. reference, name of official source/platform, or similar. (IC2)
3. The paper must describe a method or technique applied to OD, provide insight into the algorithm, or include a link to the code in open form. (IC3)
4. The paper that qualitatively describes the measured performance of the applied method or technique. (IC4)
5. The paper must be published after 2017. (IC5)
6. The paper must be written in English. (IC6)
7. The paper must be an article or conference paper. (IC7)
8. Paper must be open accessed. (IC8)

The exclusion criteria are:

1. Papers that do not contain the materials needed to answer the research questions. (EC1)

**2.** If the paper is centered around ML/AI/DL or another technique but does not present a solution related to OD, it will be excluded. (EC2)

### 4.3 SLR Workflow based on PRISMA Workflow

Based on the query string, we conducted the PRISMA workflow shown in Fig. 1. The query result on the Scopus database listed 705 papers, and we excluded 77 based on paper type and 16 based on language. In the screening phase, we had 612 papers, where we excluded 272 papers based on abstract and title. In the next step of this phase, we screened full-text papers, 340 ones, and excluded 44 papers based on open access even though we marked open access on Scopus. And last, we excluded 260 papers based on exclusion and inclusion criteria. In the end, a total of 36 papers were included in the review.

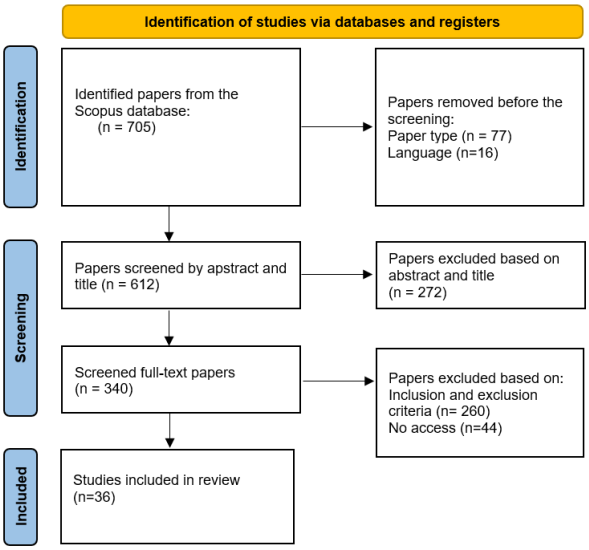



**Figure 1.** PRISMA Workflow diagram

## 5 Results

Within this section, the results obtained through the literature review will be presented. A significant increase in the number of papers is observed, especially in the year 2023 (33,3%), where the majority of selected papers for review were published in that year. Additionally, most of the papers are Journal Articles. This can be correlated with the development of AI/ML/DL in recent years and the increasing availability of OD.

Since visualization is crucial and aids in simplifying the conduct of SLR (Stefanovic et al., 2022), the RQ1, RQ2, and RQ4 have been answered in a tabular format.

To obtain an answer to RQ1, Table 1 was created, which roughly describes the methods and techniques mentioned in the primary papers on open data. In certain papers, multiple methods and techniques were employed, allowing for an experimental insight into which ones exhibit greater efficiency.

The majority of selected papers in their research utilize CNN, precisely 35,3% of them. The papers provide detailed explanations of the architectures used, with the highest number of works employing variations of U-Net (Garcia-Pedrero et al., 2019; Mainali et al., 2023; Rutherford et al., 2022; Stewart et al., 2020). Subsequently, researchers predominantly employ Random Forests for regression and classification. As can be seen in the results, many researchers use different machine learning methods in the context of regression and classification to test which ones perform better on the data and yield better results. 14,71% were used for clustering.

.

**Table 1.** Utilized methods and techniques in primary studies

| Subfield of AI | Methods/techniques | | Primary studies | N | % |
|---|---|---|---|---|---|
| ML | Regression models | Random Forest (RF) | (Boeke et al., 2019; Meneghetti et al., 2023; Nai et al., 2023; Pareeth et al., 2019; Pargent et al., 2023; Shulajkovska et al., 2023) | 15 | 22,06% |
| | | Support Vector Regressor (SVR) | (Boeke et al., 2019; Shulajkovska et al., 2023) | | |
| | | Logistic Regression (LR) | (Ma & Faye, 2022; Shulajkovska et al., 2023) | | |
| | | Linear Regression (LNR) | (Cocca et al., 2020; Shulajkovska et al., 2023) | | |
| | | Elastic Net | (Shulajkovska et al., 2023) | | |
| | | Gradient Boosting Regressor | (Shulajkovska et al., 2023) | | |
| | | Bayesian Ridge (BR) | (Shulajkovska et al., 2023) | | |
| | Classification models | Random Forest | (Boeke et al., 2019; Ma & Faye, 2022) | 11 | 16,18% |
| | | Support Vector Machine (SVM) | (Boeke et al., 2019; Ma & Faye, 2022; Mizuno et al., 2022) | | |
| | | Support Vector Regression (SVR) | (Shulajkovska et al., 2023) | | |
| | | K-nearest neighbor (k-NN) | (Boeke et al., 2019; Shulajkovska et al., 2023) | | |
| | | Adaboost | (Ma & Faye, 2022) | | |
| | | Extreme tree classifier (ETC) | (Verhegghen et al., 2022) | | |
| | | Decision Tree (DT) | (Shulajkovska et al., 2023) | | |
| | Clustering models | K-means | (Mizuno et al., 2022; Trento Oliveira et al., 2023) | 10 | 14,71% |
| | | DBSCAN | (Carter et al., 2021; Mazurek & Hachaj, 2021) | | |
| | | OPTICS | (Mazurek & Hachaj, 2021) | | |
| | | Gradient Boosting Classifier (GBM) | (Ruiz-Rizzo et al., 2022; Veigel et al., 2023) | | |
| | | LightGBM | (H. Liu & Ma, 2023; Ruiz-Rizzo et al., 2022) | | |
| | | XGBoost | (Verhegghen et al., 2022) | | |
| | Visual pattern mining (VPM) | | (Balaniuk et al., 2020) | 1 | 1,47% |
| | Extremely Randomized Tree (ERT) | | (Jin et al., 2022) | 1 | 1,47% |
| | Stochastic Gradient Descent (SGD) | | (Shulajkovska et al., 2023) | 1 | 1,47% |
| DL | Deep learning-based side-channel analysis | | (Imafuku et al., 2023) | 1 | 1,47% |
| | GNN | | (Holmberg et al., 2023) | 1 | 1,47% |
| | ANN | | (Sulavko, 2022) | 1 | 1,47% |
| | RNN | | (Rajan et al., 2020) | 2 | 2,94% |
| | | LSTM | (Ma & Faye, 2022) | | |
| | CNN | DeepLavV3+ | (Edpuganti et al., 2023; Touzani & Granderson, 2021) | 24 | 35,3% |
| | | Faster R-CNN | (Chen et al., 2021) | | |
| | | YOLO | (Chen et al., 2021; Mazurek & Hachaj, 2021) | | |
| | | DarkCovidNet | (Harkness et al., 2022) | | |
| | | COVID-Net | (Harkness et al., 2022) | | |
| | | CoroNet | (Harkness et al., 2022) | | |
| | | U-Net | (Garcia-Pedrero et al., 2019; Mainali et al., 2023; Rutherford et al., 2022; Stewart et al., 2020) | | |
| | | Mask R-CNN | (Carter et al., 2021) | | |
| | | RS-NET | (Jeppesen et al., 2019) | | |
| | | Twin CNN | (P. Liu et al., 2019) | | |
| | | Multilayer CNN | (Sulavko, 2022) | | |
| | | LUN-BiSeNetV2 | (Zhang & Zhang, 2023) | | |
| | | MobileNet | (Mazurek & Hachaj, 2021) | | |
| | | RetinaNet | (Mazurek & Hachaj, 2021) | | |
| | | MAML-CNN-LSTM-Attention-based model | (Xu et al., 2023) | | |
| | | Fully CNN (FCN) | (Balaniuk et al., 2020) | | |
| | | No concrete type specified | (Berlanga et al., 2022; Fibæk et al., 2021; Rajan et al., 2020) | | |

In summary, multiple ML models and techniques are used, highlighting that many researchers use a variety of them in their work for comparisons and similar purposes. As mentioned earlier, DL methods, are predominantly used, often combined with some other ML methods alongside convolutional neural networks. RQ2 is intended to provide an answer to the most common data source for researchers within the selected papers. Table 2 provides an insight into the answer to this question, and the papers were selected based on the origin of the data.

If the datasets are from official portals of cities, countries, or any institutions related to the state and are state-owned, they are presented as such. All individual datasets are presented separately. Each referenced paper indicates where the data was collected from. The answer to RQ2 can be seen in Table 2, indicating that researchers most commonly find relevant datasets on official open data portals belonging to public and state institutions. These data can then be partly classified as open government data. Indeed, in practice, it is observed that governments worldwide are creating many initiatives for data openness. In addition to these portals, Kaggle has significantly assisted researchers in finding data in the selected papers, and other sources are also listed. This insight into sources is significant

in aiding researchers in easily locating data that is in an open format and available for sharing, distribution, and reuse.

It is important to emphasize that, based on the inclusion and exclusion criteria, all reviewed articles are publicly available, and each one provides the exact data source, which can be useful for researchers.

To address RQ3, focusing on the strengths and limitations of AI applied to open data, researchers have consistently highlighted several critical aspects. The majority have mentioned challenges related to data quality (Cocca et al., 2020; Meneghetti et al., 2023; Mizuno et al., 2022; Veigel et al., 2023), followed by data inhomogeneity (Moher et al., 2009) and limitations in the available data (Rutherford et al., 2022; Stewart et al., 2020). Additionally, the authors emphasize the necessity for statistical analysis to assess the accuracy and reliability of models (Shulajkovska et al., 2023). Concerns regarding model obsolescence (Cashman et al., 2016; Hurbean et al., 2021) and issues related to generalization (Mizuno et al., 2022).

**Table 2**. Source of datasets

| Source of dataset | Primary studies | % |
|---|---|---|
| Government or public institutions | (Carter et al., 2021; Chen et al., 2021; Cocca et al., 2020; Ma & Faye, 2022; Meneghetti et al., 2023; Mizuno et al., 2022; Ruiz-Rizzo et al., 2022; Veigel et al., 2023; Zhang & Zhang, 2023) | 25,2% |
| Airconstructor | (Sulavko, 2022) | 2,8% |
| AmesHousing; Titanic | (Pargent et al., 2023) | 2,8% |
| ASCAD | (Imafuku et al., 2023) | 2,8% |
| ANAC | (Nai et al., 2023) | 2,8% |
| CERN OD Portal | (Holmberg et al., 2023) | 2,8% |
| Copernicus | (Balaniuk et al., 2020; Verhegghen et al., 2022) | 5,6% |
| GDSC | (P. Liu et al., 2019) | 2,8% |
| GIS | (Touzani & Granderson, 2021) | 2,8% |
| Kaggle | (Edpuganti et al., 2023; Rajan et al., 2020; Stewart et al., 2020; Xu et al., 2023) | 11,2% |
| Landstat 8 | (Jeppesen et al., 2019; Pareeth et al., 2019) | 5,6% |
| Kitti | (Mazurek & Hachaj, 2021) | 2,8% |
| LPIS | (Garcia-Pedrero et al., 2019) | 2,8% |
| ODR | (Berlanga et al., 2022) | 2,8% |
| OPENAQ | (Jin et al., 2022) | 2,8% |
| OpenNeuro | (Rutherford et al., 2022) | 2,8% |
| OSF | (H. Liu & Ma, 2023; Mainali et al., 2023) | 5,6% |
| RICORD | (Harkness et al., 2022) | 2,8% |
| Sentinel 1/2 | (Fibæk et al., 2021) | 2,8% |
| Urbanite | (Shulajkovska et al., 2023) | 2,8% |
| WorldPop | (Trento Oliveira et al., 2023) | 2,8% |
| ICEAS | (Boeke et al., 2019) | 2,8% |

The identified key strengths of implementing AI on OD include the enhancement of public administration systems through machine learning (Nai et al., 2023), machine learning models providing information about factors predicting outcomes while offering specific predictions for each case (Ruiz-Rizzo et al., 2022), and the use of open data that additionally supports research in this field (Jin et al., 2022).

To answer RQ4, papers and their purposes were observed, i.e., the purposes for which they used different methods and techniques on open data were examined. Table 3 provides an insight into the focus of the selected papers. All papers are grouped into several categories, depending on whether they propose a new solution (method, model, approach), are of a research nature, or involve analysis, comparison, or enhancement of existing systems. Some papers deal with the development of a new approach or method.

The majority of papers focus on proposing a new approach, model, or method, which is certainly since AI/ML/DL methods and techniques are quite promising and applicable in practice. Considering the huge potential of open data, the application of modern methods and techniques with open data can significantly influence the development and proposal of new methods, models, and approaches to problem-solving. Moreover, many research papers use open data, as well as analyses and comparisons of existing methods and techniques.

Additionally, it is important to emphasize that most of the papers are published in journals with a good impact factor (IF). Some of the papers are published in journals with excellent impact factors, such as (Mizuno et al., 2022) with 12,4 IF, and (Fibæk et al., 2021) with 7,6 IF.

**Table 3.** Research goal/purpose

| Purpose | | Primary studies | % |
|---|---|---|---|
| Proposing | Method | (Chen et al., 2021; Jin et al., 2022; Mizuno et al., 2022; Stewart et al., 2020; Zhang & Zhang, 2023) | 14% |
| | Model | (Sulavko, 2022; Xu et al., 2023) | 5,6% |
| | Approach | (Fibæk et al., 2021; Ma & Faye, 2022; Mazurek & Hachaj, 2021; Touzani & Granderson, 2021) | 11,2% |
| Exploratory | | (Balaniuk et al., 2020; Carter et al., 2021; Edpuganti et al., 2023; Garcia-Pedrero et al., 2019; Holmberg et al., 2023; Nai et al., 2023; Ruiz-Rizzo et al., 2022; Veigel et al., 2023) | 22,4% |
| Developing approach/method | | (Berlanga et al., 2022; Pareeth et al., 2019; Shulajkovska et al., 2023) | 8,4% |
| Analysis | | (Harkness et al., 2022; Mainali et al., 2023; Meneghetti et al., 2023; Rajan et al., 2020; Trento Oliveira et al., 2023; Verhegghen et al., 2022) | 16,8% |
| Comparison | | (Boeke et al., 2019; Cocca et al., 2020) | 5,6% |
| Improvement of already existing solutions | | (Imafuku et al., 2023; H. Liu & Ma, 2023; P. Liu et al., 2019; Rutherford et al., 2022) | 11,2% |
| Other | | (Jeppesen et al., 2019; Pargent et al., 2023) | 5,6% |

# 6 Discussion

The results obtained after the SLR, which were presented in the previous chapter, will be briefly discussed in this section.

It can be concluded that there has been a trend in the development and application of AI/ML/DL methods and techniques on open data in recent years. This is inevitably influenced by the development of information technologies and the increasing availability of OD. Primarily, institutions open portals for open data, but the research indicates that there are also other relevant sources assisting researchers in using, distributing, and downloading data, as shown in Table 2. Additionally, some platforms have collected data, combined it from other open portals, and enabled free usage, such as Kaggle. This way, the answer to RQ2 was obtained.

Regarding RQ1, DL methods like CNN and other neural networks, along with ML models for regression, classification, and clustering (especially RF), are most commonly used.

Addressing the problems that researchers encounter, as shown through RQ3, and improving open data sets and their sharing can lead to the development and use of AI/ML/DL on OD, and such solutions provide additional value. As shown through RQ4, there is an increasing number of papers proposing new solutions addressing the problems highlighted by researchers, and this segment can be further improved.

This paper enhances our understanding of the synergy between AI and OD and shows how this synergy can be applied in industry and society to create value and innovation.

# 7 Conclusion

In recent years, there has been significant development and awareness of the importance of Open Data and its potential for implementing smart solutions with advanced technologies. Researchers often face the challenge of finding high-quality, relevant data. Open data not only enhances transparency but also drives progress in research and industry.

To date, a detailed literature review on the main data sources, methods, and techniques used with advanced technologies has been lacking. This paper addresses this gap by analyzing trends and aiding in the discovery of various datasets. The review highlights the increasing use of AI/ML/DL in processing and extracting knowledge from these datasets. It provides code and algorithm descriptions, offering insights for future researchers.

The findings from this literature review shed light on the current state of the field and support the further development and application of advanced technologies on open data, aiming to improve various industry processes.

# Acknowledgment

This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-65/2024-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through the project “Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad” (No. 01-3394/1).

# References

Adler, D. A., Wang, F., Mohr, D. C., Estrin, D., Livesey, C., & Choudhury, T. (2022). A call for open data to develop mental health digital biomarkers. *BJPsych Open*, *8*(2), e58. https://doi.org/10.1192/bjo.2022.28

Ali, M., Alexopoulos, C., & Charalabidis, Y. (2022). A comprehensive review of open data platforms, prevalent technologies, and functionalities. *15th International Conference on Theory and Practice of Electronic Governance*, 203–214. https://doi.org/10.1145/3560107.3560142

Balaniuk, R., Isupova, O., & Reece, S. (2020). Mining and Tailings Dam Detection in Satellite Imagery Using Deep Learning. *Sensors*, *20*(23), 6936. https://doi.org/10.3390/s20236936

Berlanga, G., Williams, Q., & Temiquel, N. (2022). Convolutional Neural Networks as a Tool for Raman Spectral Mineral Classification Under Low Signal, Dusty Mars Conditions. *Earth and Space Science*, *9*(10), e2021EA002125. https://doi.org/10.1029/2021EA002125

Boeke, S., Van Den Homberg, M. J. C., Teklesadik, A., Fabila, J. L. D., Riquet, D., & Alimardani, M. (2019). TOWARDS PREDICTING RICE LOSS DUE TO TYPHOONS IN THE PHILIPPINES. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-4/W19*, 63–70. https://doi.org/10.5194/isprs-archives-XLII-4-W19-63-2019

Brinkhaus, H. O., Rajan, K., Schaub, J., Zielesny, A., & Steinbeck, C. (2023). Open data and algorithms for open science in AI-driven molecular informatics. *Current Opinion in Structural Biology*, *79*, 102542. https://doi.org/10.1016/j.sbi.2023.102542

Carter, B. P., Blackadar, J. H., & Conner, W. L. A. (2021). When Computers Dream of Charcoal: Using Deep Learning, Open Tools, and Open Data to Identify Relict Charcoal Hearths in and around State Game Lands in Pennsylvania. *Advances in Archaeological Practice*, *9*(4), 257–271. https://doi.org/10.1017/aap.2021.17

Cashman, S. A., Meyer, D. E., Edelen, A. N., Ingwersen, W. W., Abraham, J. P., Barrett, W. M., Gonzalez, M. A., Randall, P. M., Ruiz-Mercado, G., & Smith, R. L. (2016). Mining Available Data from the United States Environmental Protection Agency to Support Rapid Life Cycle Inventory Modeling of Chemical Manufacturing. *Environmental Science & Technology*, *50*(17), 9013–9025. https://doi.org/10.1021/acs.est.6b02160

Chen, L., Grimstead, I., Bell, D., Karanka, J., Dimond, L., James, P., Smith, L., & Edwardes, A. (2021). Estimating Vehicle and Pedestrian Activity from Town and City Traffic Cameras. *Sensors*, *21*(13), 4564. https://doi.org/10.3390/s21134564

Cocca, M., Teixeira, D., Vassio, L., Mellia, M., Almeida, J. M., & Couto Da Silva, A. P. (2020). On Car-Sharing Usage Prediction with Open Socio-Demographic Data. *Electronics*, *9*(1), 72. https://doi.org/10.3390/electronics9010072

Dakic, D., Sladojevic, S., Lolic, T., & Stefanovic, D. (2019). Process Mining Possibilities and Challenges: A Case Study. *2019 IEEE 17th International Symposium on Intelligent Systems and Informatics (SISY)*, 000161–000166. https://doi.org/10.1109/SISY47553.2019.9111591

Edpuganti, A., Akshaya, P., Gouthami, J., Sajith Variyar, V. V., Sowmya, V., & Sivanpillai, R. (2023). EFFECT OF DATA QUALITY ON WATER BODY SEGMENTATION WITH DEEPLABV3+ ALGORITHM. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-M-3–2023*, 81–85. https://doi.org/10.5194/isprs-archives-XLVIII-M-3-2023-81-2023

Fibæk, C. S., Keßler, C., & Arsanjani, J. J. (2021). A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning. *International Journal of Applied Earth Observation and Geoinformation*, *105*, 102628. https://doi.org/10.1016/j.jag.2021.102628

Garcia-Pedrero, A., Lillo-Saavedra, M., Rodriguez-Esparragon, D., & Gonzalo-Martin, C. (2019). Deep Learning for Automatic Outlining Agricultural Parcels: Exploiting the Land Parcel Identification System. *IEEE Access*, *7*, 158223–158236. https://doi.org/10.1109/ACCESS.2019.2950371

Harkness, R., Hall, G., Frangi, A. F., Ravikumar, N., & Zucker, K. (2022). The Pitfalls of Using Open Data to Develop Deep Learning Solutions for COVID-19 Detection in Chest X-Rays. In P. Otero, P. Scott, S. Z. Martin, & E. Huesing (Eds.), *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI220164

Holmberg, D., Golubovic, D., & Kirschenmann, H. (2023). Jet Energy Calibration with Deep Learning as a Kubeflow Pipeline. *Computing and Software for Big Science*, *7*(1), 9. https://doi.org/10.1007/s41781-023-00103-y

Hurbean, L., Danaiata, D., Militaru, F., Dodea, A.-M., & Negovan, A.-M. (2021). Open Data Based Machine Learning Applications in Smart Cities: A Systematic Literature Review. *Electronics*, *10*(23), 2997. https://doi.org/10.3390/electronics10232997

Imafuku, K., Kawamura, S., Nozaki, H., Sakamoto, J., & Osuka, S. (2023). Non-Profiled Deep Learning-Based Side-Channel Analysis With Only One Network Training. *IEEE Access*, *11*, 83221–83231. https://doi.org/10.1109/ACCESS.2023.3301178

Jeppesen, J. H., Jacobsen, R. H., Inceoglu, F., & Toftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, *229*, 247–259. https://doi.org/10.1016/j.rse.2019.03.039

Jin, C., Wang, Y., Li, T., & Yuan, Q. (2022). Global validation and hybrid calibration of CAMS and MERRA-2 PM2.5 reanalysis products based on OpenAQ platform. *Atmospheric Environment*, *274*, 118972. https://doi.org/10.1016/j.atmosenv.2022.118972

Kijanovic, S., & Komosar, A. (2023). Research on Open Data Practice Within Education. *2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, 1–6. https://doi.org/10.1109/INFOTEH57020.2023.10094111

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. *2*.

Komosar, A., & Kijanovic, S. (2023). Research on the Application of Open Data in the Public Finance Sector. *2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, 1–6. https://doi.org/10.1109/INFOTEH57020.2023.10094103

Liu, H., & Ma, E. (2023). An Explainable Evaluation Model for Building Thermal Comfort in China. *Buildings*, *13*(12), 3107. https://doi.org/10.3390/buildings13123107

Liu, P., Li, H., Li, S., & Leung, K.-S. (2019). Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinformatics*, *20*(1), 408. https://doi.org/10.1186/s12859-019-2910-6

Lopes, P. V., Silveira, L., Guimaraes Aquino, R. D., Ribeiro, C. H., Skoogh, A., & Verri, F. A. N. (2024). Synthetic data generation for digital twins: Enabling production systems analysis in the absence of data. *International Journal of Computer Integrated Manufacturing*, 1–18. https://doi.org/10.1080/0951192X.2024.2322981

Ma, T.-Y., & Faye, S. (2022). Multistep electric vehicle charging station occupancy prediction using hybrid LSTM neural networks. *Energy*, *244*, 123217. https://doi.org/10.1016/j.energy.2022.123217

Mainali, K., Evans, M., Saavedra, D., Mills, E., Madsen, B., & Minnemeyer, S. (2023). Convolutional neural network for high-resolution wetland mapping with open data: Variable selection and the challenges of a generalizable model. *Science of The Total Environment*, *861*, 160622. https://doi.org/10.1016/j.scitotenv.2022.160622

Mazurek, P., & Hachaj, T. (2021). SLAM-OR: Simultaneous Localization, Mapping and Object Recognition Using Video Sensors Data in Open Environments from the Sparse Points Cloud. *Sensors*, *21*(14), 4734. https://doi.org/10.3390/s21144734

Meneghetti, N., Pacini, F., Biondi Dal Monte, F., Cracchiolo, M., Rossi, E., Mazzoni, A., & Micera, S. (2023). Predicting party switching through machine learning and open data. *iScience*, *26*(7), 107098. https://doi.org/10.1016/j.isci.2023.107098

Mizuno, S., Ohba, H., & Ito, K. (2022). Machine learning-based turbulence-risk prediction method for the safe

operation of aircrafts. *Journal of Big Data*, *9*(1), 29. https://doi.org/10.1186/s40537-022-00584-5

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

Nai, R., Meo, R., Morina, G., & Pasteris, P. (2023). Public tenders, complaints, machine learning and recommender systems: A case study in public administration. *Computer Law & Security Review*, *51*, 105887. https://doi.org/10.1016/j.clsr.2023.105887

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

*Open data and AI: A symbiotic relationship for progress | data.europa.eu*. (2023, June 9). https://data.europa.eu/en/publications/datastories/open-data-and-ai-symbiotic-relationship-progress

Pareeth, S., Karimi, P., Shafiei, M., & De Fraiture, C. (2019). Mapping Agricultural Landuse Patterns from Time Series of Landsat 8 Using Random Forest Based Hierarchial Approach. *Remote Sensing*, *11*(5), 601. https://doi.org/10.3390/rs11050601

Pargent, F., Schoedel, R., & Stachl, C. (2023). Best Practices in Supervised Machine Learning: A Tutorial for Psychologists. *Advances in Methods and Practices in Psychological Science*, *6*(3), 25152459231162559. https://doi.org/10.1177/25152459231162559

Rajan, K., Zielesny, A., & Steinbeck, C. (2020). DECIMER: Towards deep learning for chemical image recognition. *Journal of Cheminformatics*, *12*(1), 65. https://doi.org/10.1186/s13321-020-00469-w

Rosés, R., Kadar, C., & Malleson, N. (2021). A data-driven agent-based simulation to predict crime patterns in an urban environment. *Computers, Environment and Urban Systems*, *89*, 101660. https://doi.org/10.1016/j.compenvurbsys.2021.101660

Ruiz-Rizzo, A. L., Archila-Meléndez, M. E., & González Veloza, J. J. F. (2022). Predicting the probability of finding missing older adults based on machine learning. *Journal of Computational Social Science*, *5*(2), 1303–1321. https://doi.org/10.1007/s42001-022-00171-x

Rutherford, S., Sturmfels, P., Angstadt, M., Hect, J., Wiens, J., Van Den Heuvel, M. I., Scheinost, D., Sripada, C., & Thomason, M. (2022). Automated Brain Masking of Fetal Functional MRI with Open Data. *Neuroinformatics*, *20*(1), 173–185. https://doi.org/10.1007/s12021-021-09528-5

Shulajkovska, M., Smerkol, M., Dovgan, E., & Gams, M. (2023). A machine-learning approach to a mobility policy proposal. *Heliyon*, *9*(10), e20393. https://doi.org/10.1016/j.heliyon.2023.e20393

Stefanovic, D., Havzi, S., Spasojevic, I., Lolic, T., & Ristic, S. (2022). Software Tools to Support Visualising Systematic Literature Review. In B. Lalic, D. Gracanin, N. Tasic, & N. Simeunović (Eds.), *Proceedings on 18th International Conference on Industrial Systems – IS'20* (pp. 79–86). Springer International Publishing. https://doi.org/10.1007/978-3-030-97947-8_11

Stewart, C., Lazzarini, M., Luna, A., & Albani, S. (2020). Deep Learning with Open Data for Desert Road Mapping. *Remote Sensing*, *12*(14), 2274. https://doi.org/10.3390/rs12142274

Sulavko, A. (2022). Biometric-Based Key Generation and User Authentication Using Acoustic Characteristics of the Outer Ear and a Network of Correlation Neurons. *Sensors*, *22*(23), 9551. https://doi.org/10.3390/s22239551

Touzani, S., & Granderson, J. (2021). Open Data and Deep Semantic Segmentation for Automated Extraction of Building Footprints. *Remote Sensing*, *13*(13), 2578. https://doi.org/10.3390/rs13132578

Trento Oliveira, L., Kuffer, M., Schwarz, N., & Pedrassoli, J. C. (2023). Capturing deprived areas using unsupervised machine learning and open data: A case study in São Paulo, Brazil. *European Journal of Remote Sensing*, *56*(1), 2214690. https://doi.org/10.1080/22797254.2023.2214690

Veigel, N., Kreibich, H., & Cominola, A. (2023). Interpretable Machine Learning Reveals Potential to Overcome Reactive Flood Adaptation in the Continental US. *Earth's Future*, *11*(9), e2023EF003571. https://doi.org/10.1029/2023EF003571

Verhegghen, A., Kuzelova, K., Syrris, V., Eva, H., & Achard, F. (2022). Mapping Canopy Cover in African Dry Forests from the Combined Use of Sentinel-1 and Sentinel-2 Data: Application to Tanzania for the Year 2018. *Remote Sensing*, *14*(6), 1522. https://doi.org/10.3390/rs14061522

*What is Open Data?* (2024, October 8). https://opendatahandbook.org/guide/en/what-is-open-data/

Wirtz, B. W., Weyerer, J. C., Becker, M., & Müller, W. M. (2022). Open government data: A systematic literature review of empirical research. *Electronic Markets*, *32*(4), 2381–2404. https://doi.org/10.1007/s12525-022-00582-8

Xu, M., Ren, H., Chen, P., & Xin, G. (2023). On the V2G capacity of shared electric vehicles and its forecasting through MAML-CNN-LSTM-Attention algorithm. *IET Generation, Transmission & Distribution*, gtd2.12921. https://doi.org/10.1049/gtd2.12921

*Zakon o elektronskoj upravi | Paragraf Lex propisi*. (2018, April 6). http://www.paragraf.rs/propisi/zakon-o-elektronskoj-upravi-republika-srbija.html

Zhang, Y., & Zhang, M. (2023). LUN-BiSeNetV2: A lightweight unstructured network based on BiSeNetV2 for road scene segmentation. *Computer Science and Information Systems*, *20*(4), 1749–1770. https://doi.org/10.2298/CSIS221205029Z

# Reevaluation of intentions of dynamic agent utilizing BDI reasoning in ABAsim architecture

**Michaela Pauriková, Michal Varga**
University of Žilina
Faculty of Management Science and Informatics
Department of informatics
Univerzitná 8215/1, 010 26, Žilina
paurikova2@stud.uniza.sk
michal.varga@fri.uniza.sk

**Norbert Adamko**
Simcon, s.r.o.
J. Kráľa 19, 010 01, Žilina
norbert.adamko@simcon.sk

***Abstract**. Paper proposes a mechanism of reevaluation of intentions of ABAsim architecture's dynamic agent utilizing BDI (Belief-Desire-Intention) reasoning. Universal and easily extensible solution to reevaluate agent's intentions is presented. Capabilities of dynamic agent are expanded to choose user specified plan to fulfil desire, to reevaluate chosen plan dependent on the state of the environment in which dynamic agent is situated, and to interrupt interpreted plan if the dynamic agent changes its beliefs. The mechanism is implemented as part of architecture library, what makes it easy to use in any simulation model using ABAsim architecture. We present its implementation in simulation tool PedSim developed at the Faculty of Management Science and Informatics of the University of Žilina and demonstrate the functionality in simple scenarios created in the simulation tool.*

**Keywords.** agent-based simulation; ABAsim architecture; BDI reasoning; reevaluation of intentions

## 1 Introduction

Decision-making is one of the key abilities of humans in achieving their goals. It may be used when one decides how to pursue a goal or the subsequent reevaluation of those chosen decisions. Implementation of decision-making mechanism into simulation tools for modeling of pedestrian movement and behavior, not only enhances modeling capabilities of such software but, most importantly, it makes it possible to obtain more accurate results.

Pedestrian simulation is relevant method for solving various problems, such as evacuation in emergency situations (Siyam et al., 2020) or studying people's movement in transportation systems (Wang, 2021). Therefore, it is essential to continually enhance the models and provide more precise and detailed interpretations of human behavior. One such enhancement includes the universal mechanism of decision-making and intention reevaluation for dynamic agents. The foundations of this mechanism were laid out in the work by (Varga & Adamko, 2014), and its extension is the aim of this paper.

One approach, by which an agent, representing a pedestrian, can define and manage its intentions, is based upon the BDI (Belief-Desire-Intention) paradigm (Varga & Adamko, 2014). The agent strives to fulfill its desires by achieving its goals, which it does through the execution of defined plans. In our work, we focus on the method of defining various rules for selecting or canceling these plans and reevaluating them during plan interpretation. This mechanism is implemented in the ABAsim architecture (Adamko, 2013). ABAsim architecture is utilized by the PedSim simulation tool, which we will use to demonstrate the effectiveness of proposed mechanism. However, the enhancement of the mechanism is not limited to aforementioned software, but is relevant for any tool based on the ABAsim architecture and can also be adopted by other agent-based architectures as well.

The paper is organized as follows. Section 2 focuses on basic concepts such as agent-based simulation, the ABAsim architecture, and the BDI paradigm. In Section 3, we introduce the proposed extension and compare it with the current modeling approach. Fourth section present the results achieved by applying the proposed mechanism and the last section contains conclusion and introduces future work.

## 2 Basic concepts

Investigation of properties of complex systems is not intuitive since factors such as emergent behavior come in place. Agent-based simulation belongs to relevant techniques for capturing properties of such systems, experimenting with changes within them and evaluating impact of those changes. When using agent-based simulation, the system is decomposed into autonomous agents. As software agent we consider an

encapsulated computer system situated in a certain environment, in which the agent acts flexibly and autonomously to achieve a given goal (Jennings, 2001). The operation of system is achieved by defining behavior and relations between agents. The agents can be classified from several perspectives. According to mobility, we classify agents as either stationary or mobile. Based on their application purpose, one can identify e.g. internet and information agents. The most important classification of an agent for our purposes is based on the level of initiative (Adamko, 2013): reactive and deliberative agents. Deliberative/intelligent agent evaluates the situation and considers an appropriate action using its internal symbolic model of the environment or works according to a predetermined plan.

## 2.1 ABAsim architecture

ABAsim (Agent-Based Architecture of Simulation Models) architecture has been developed at the Faculty of Management Science and Informatics of the University of Žilina. The application of the ABAsim architecture can be found in simulation software Villon (Klima et al., 2001) and PedSim (Varga & Adamko, 2014). Villon is a simulation tool capable to create detailed microscopic models of transportation terminals. PedSim is pedestrian behavior and movement simulation model. Utilization of ABAsim architecture in both tools enables simple integration of PedSim simulation model into Villon.

The main components of ABAsim architecture are agents of two types: *managing* and *dynamic*. The managing agent is responsible for a specific part of the modeling system. This agent is part of the hierarchical structure of agents, which tries to achieve a common goal through their cooperation. Each ABAsim model must contain at least one managing agent. The dynamic agent represents an autonomous intelligent entity. It can control its behavior in order to achieve given goal. Goal may be assigned by the managing agent or by the dynamic agent itself (if it is not in contrary to the goal assigned by the managing agent). The last important components of the ABAsim architecture are *entities*. An entity usually represents the physical entities of modelled systems without autonomous and intelligent behavior. In an example of an airport service system, managing agent can represent the airport dispatcher, the dynamic agents can represent the visitors, and the entities can represent planes. Example of such system topology is depicted in Fig. 1.

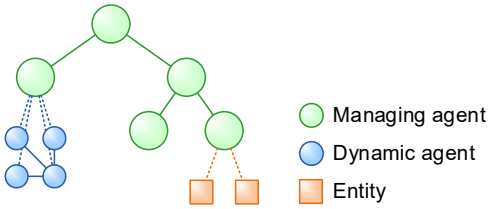



**Figure 1**. Example of model topology in architecture ABAsim containing managing agents, dynamic agents and entities

Each agent can be decomposed into several components (Adamko, 2013). Core component is agent's *manager* which is responsible for agent's reasoning. Reasoning of an agent can be defined in several ways:

1. explicitly in program code,
2. using ABAgraph (modified Petri net), or
3. by defining components, which are used by manager utilizing specialized kind of reasoning (example is BDI reasoning, which we will discuss later in this paper).

Manager cooperates with *assistants,* which can be classified according to their usage. *Sensors* are used to provide information about the state of the environment, in which the agent operates; *solvers* provide suggestions for problem-solving; and *effectors* perform actions in the environment. Lastly, two internal components responsible for synchronization and communication are part of an agent. The internal structure of an agent is depicted in Fig. 2.

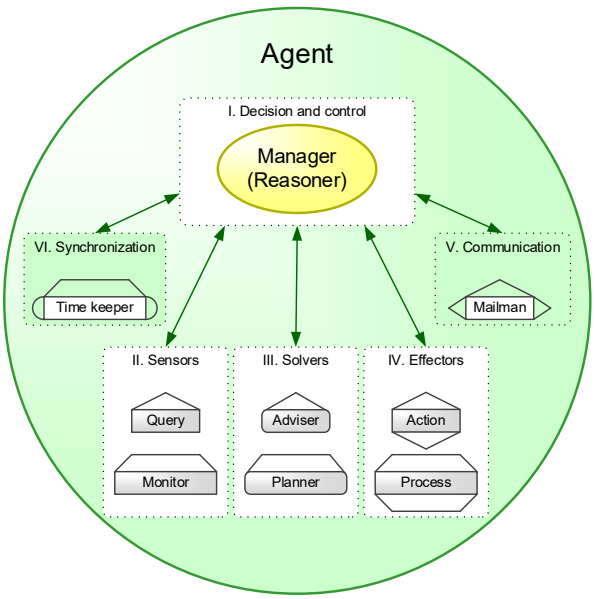



**Figure 2.** Decomposition of an agent in ABAsim architecture

## 2.2 BDI paradigm

The ability to manage goals and make decisions according to acquired information is one of the key requirements when modeling human behavior. The principle of BDI, as one of the paradigms utilized to model human behavior, is based on practical human reasoning (Bratman, 1987) (Kinny & Georgeff, 1991) and distinguishes three main components – beliefs, desires, and intentions. *Beliefs* represent agent information about the surrounding environment (beliefs do not need to be necessarily true, e.g. agent may believe that train departs at certain time, however it may be delayed); *desires* are agent's possible goals that can be fulfilled; and *intentions* represent agent's will to fulfill chosen desire using specific *plan*.

The BDI paradigm offers not only intuitive definition of agent's reasoning, but also the possibility of extending it with various plan selection and reconsideration mechanisms. In the ABAsim architecture, only a dynamic agent is capable to reason using BDI paradigm. This is performed by specifically defined reasoner and assistants – *sensors*, whose task is to gather information from the simulation environment. This information is transformed into beliefs and stored in the belief store, which represents the agent's virtual memory and can process, update, and forget individual beliefs. Recall that dynamic agent's task is to achieve a goal set by the managing agent. This goal is treated as a desire (to satisfy order from managing agent) and stored in the desire store as a new desire awaiting to be fulfilled. Intentions are created from desires that the agent is capable to fulfill. Generated intentions are registered in the intention structure. Fulfillment of an intention is performed through set of plans defined for the specific desires. Individual plans are stored in the plan library.

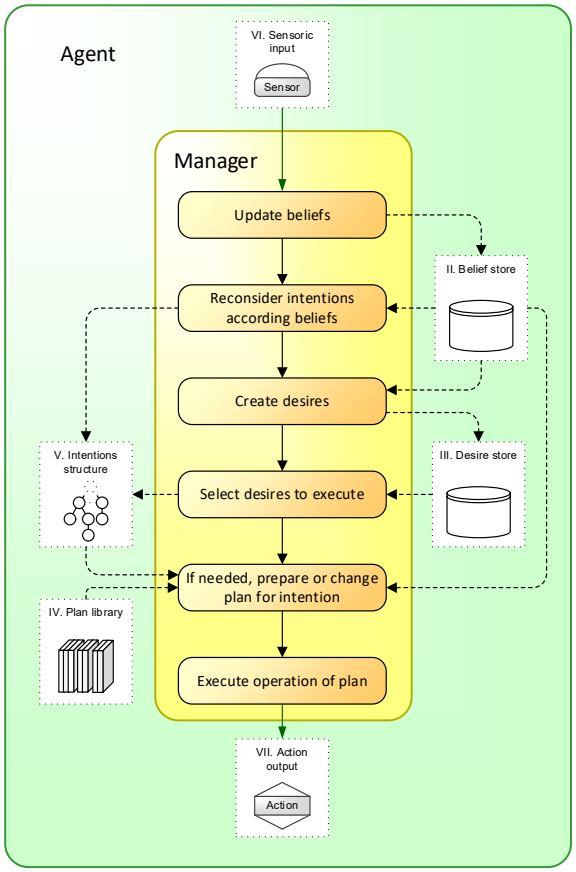



**Figure 3**: Principle of BDI reasoning

The principle of BDI reasoning, from belief acquisition and formation to successful or unsuccessful fulfillment of a chosen desire, is described in Fig. 3.

BDI plan is the key element that defines agent's activity and the fulfillment of its desire. Like missions (Varga & Adamko, 2019) used to specify agent's strategical decisions, BDI plan is defined in a form of a graph. Plan consists of nodes that dynamic agent interprets (Varga & Adamko, 2014) and thus executes its decision making and perform actions. The nodes are:

- *Start* representing the beginning of the plan, i.e. where dynamic agent starts its interpretation;
- *Finish* and *Abort* nodes representing successful or failed intention to fulfill a desire.
- *Activity* is linked to specific action performed by dynamic agent. The interpretation of an activity node usually ends in future (simulation) time, similarly to execution of events in event-driven simulation.
- *Operation* represents internal agent activity without impact on the environment. Typically, extracting or preparing information from beliefs for later usage in activities is performed when this node is interpreted.
- *Goal* interpretation leads to formulation of new desire, the node splits the execution flow of BDI plan into two branches linked to successful or failed intention used to fulfill the formulated desire.

- *Decision* is used to split execution flow into branches according to decision making process. Every branch is linked to possible value(s) of evaluation of decision. When agent interprets decision, current beliefs are considered.

Formally the BDI plan is an oriented graph $P = (N, B)$ composed of set of nodes $N$ and set of branches $B$:

$$N = \{n_1 .. n_k\}; k = |N|$$
$$B = \{b_1 .. b_l\}; l = |B|$$
$$\forall b \in B: b = (n^{from}, n^{to});$$
$$n^{from}, n^{to} \in N; n^{from} \neq n^{to}$$

We define index set $T$ reflecting types of nodes; and disjoint subsets containing nodes with given $T$:

$$T = \{Start, Finish, Abort, Activity, Operation, Goal, Decision\}$$
$$N^t = \bigcup_{n \in N} n: type(n) = t$$
$$N = \bigcup_{t \in T} N^t$$
$$\bigcap_{t \in T} N^t = \emptyset$$

Lastly, using the out neighbourhood of a node

$$N^+(n) = \bigcup_{m \in N} m: \exists b \in B: b = (n, m)$$

and outdegree

$$\deg^+(n) = |N^+(n)|$$

topology of the graph can be formalized as follows:

$$|N^{Start}| = 1;\ \forall n \in N^{Start}: \deg^+(n) = 1$$
$$|N^{Finish}| = 1;\ \forall n \in N^{Finish}: \deg^+(n) = 0$$
$$|N^{Abort}| = 1;\ \forall n \in N^{Abort}: \deg^+(n) = 0$$
$$|N^{Activity}| \geq 0;\ \forall n \in N^{Activity}: \deg^+(n) = 1$$
$$|N^{Operation}| \geq 0;\ \forall n \in N^{Operation}: \deg^+(n) = 1$$
$$|N^{Goal}| \geq 0;\ \forall n \in N^{Goal}: \deg^+(n) = 2$$
$$|N^{Decision}| \geq 0;\ \forall n \in N^{Decision}: \deg^+(n) \geq 1$$

Dynamic agent's plans library stores prepared plans. The plans are organized according to desires that agent may satisfy using them. To satisfy a desire, multiple plans may be defined. The process of selection of proper plan in given situation plays vital role in agent's behavior. The selection of plan is performed according to agent's preference, which may be defined in many different ways.

# 3 Reevaluattion of intentions

Consider a person who must get to work. The person may either walk or take some means of public transportation. The person decides to take a walk, since weather seems fine. However, halfway the person had to take a tram because it started raining. The goal remained the same – to get to work. Only the utilized plan used to fulfill the goal has changed from "take a walk" to "take a tram" with respect to current environment conditions. It is an important feature of a simulation tool to enable such reassessment. With it, it is possible to model situations, in which one is unable to complete the goal in the originally intended way but is forced to choose an alternative method to achieve the goal, or even reconsider the possibility of goal achievement itself. Absence of such mechanism can have negative consequences, for example when multiple dynamic agents have the same desire, due to the identical behavior, they would intend to achieve the goal the same way without considering their knowledge or personal preferences. This is insufficient for modeling realistic situations and human reason, that motivated the usage of BDI paradigm.

There are two possible places to reevaluate agent intentions, which we will discuss in more detail in next subsections:

1. *To select a plan to fulfill the desire* (i.e. formulation of an intention). Recall that plans are stored in a plan library. We propose universal extension to the first-fit strategy originally implemented in the library.
2. *To reconsider current intentions according to updated beliefs*. This way it will be possible for an agent to adapt to the current environment conditions as well as to reassess the feasibility of desires in relation to the current state of the environment. The fundamentals of this concept were included in the original implementation of the BDI reasoner of dynamic agent in ABAsim architecture; however, we present universal way to define reassessment using the formalism of BDI plans.

## 3.1 Selection of a plan

The choice of plan from the plan library depends on the agent's preferences and beliefs. The preferences represent extra information, thanks to which the agent can evaluate the relative quality of the plan to satisfy a desire (Baier & McIlraith, 2007). Recall that dynamic agent can choose from many plans to satisfy sole goal. For this reason, preferences may not be associated with the goal itself. Therefore, (Hindriks & van Riemsdijk, 2009) name the preference as "soft restriction". On the other hand, the goal is a "hard restriction" because it must always be achieved. There are several ways to define plan preferences. It is possible to allow the user to define preferences for each plan separately (Dasgupta & Ghose, 2011). Another approach can be found in (Visser et al., 2015) whose basic desire formula consists of the type of attribute to which the preference applies and the value indicating its priority over other types.

To expand dynamic agent's decision-making capabilities when selecting a plan to fulfil a desire, we introduce agent component named *strategy*. Strategy combines model designer's will to give dynamic agent set of rules when selecting a plan, together with preferences of the dynamic agent. Strategy can be defined for each goal separately. Cooperation of new

component with existing components of agents during the reasoning process is depicted in Fig. 4.

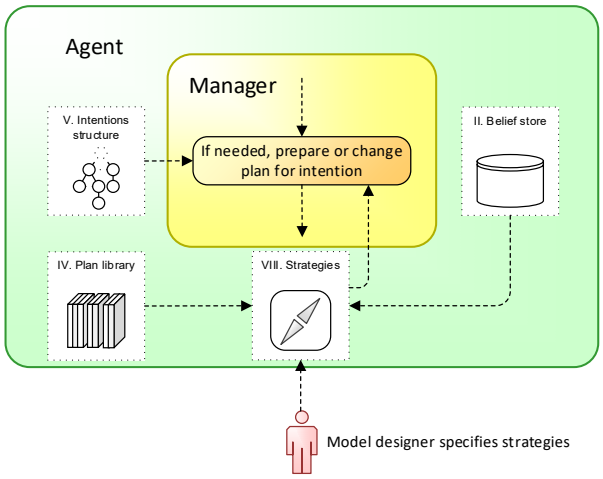



**Figure 4**. Cooperation of strategies module with relevant components of an agent

This approach was chosen for several reasons. It makes possible (1) to model different reactions of agents to the same goal, (2) to model the need to prefer the plan over other plans as well as (3) to dynamically update preferences according to the beliefs of agent, or (4) to reflect the original implementation (and ensure backward compatibility) and also (5) it enables simple future expansion with a new strategy.

We implemented three basic strategies:

a) *First fit* – to achieve the goal dynamic agent chooses the first suitable plan from the plans library.

*b)* *Best fit* – to achieve the goal dynamic agent chooses the best suitable plan from the plans library. This is performed based on the priority defined by the model designer. The priority is represented as an integer in the range from 1 to 100. The value 100 represents the most preferred plan, and the value 1 represents the least preferred plan.

c) *Random fit* – to achieve the goal dynamic agent chooses random suitable plan from the plans library. The model designer can specify the probability distribution.

## 3.2 Evaluation of a plan

Recall example with a person walking to work. The preferred selection of plan "take a walk" over "take a tram" in the first place would be realized via aforementioned strategy mechanism. Note that when the dynamic agent selects a plan via some strategy, only suitable plans can be used, e.g. plan "take a walk" can be considered only when "person has enough time" and "weather is fine". Moreover, the person was capable to drop already taken plan (after it started raining), this leads to another mechanism related to the plans. BDI models such behavior using plan reassessment via trigger and abort conditions of plans.

To expand dynamic agent's decision-making capabilities when performing plan reassessment, we introduce agent component called *evaluator*. Evaluator utilizes beliefs of dynamic agent to evaluate trigger and abort condition. Evaluators can be defined for each plan separately. Cooperation of new component with existing components of agents during the reasoning process is depicted in Fig. 5.

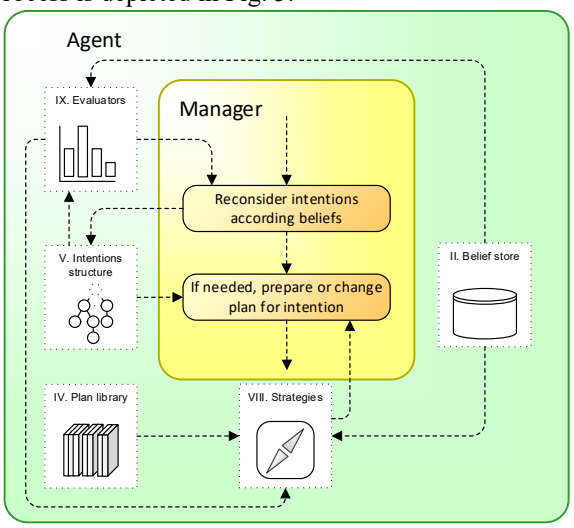



**Figure 5**. Cooperation of evaluators module with relevant components of an agent

The evaluator represents function, which processes actual state of the agent and return Boolean value expressing weather plan can be utilized (trigger) or it must be dropped (abort).

Analogously to a manager, an evaluator can also be defined in multiple ways:

1. Explicitly using program code,
2. Using *reassessment plans*. If the evaluator is used to evaluate trigger condition, we refer to *trigger plan*, if the evaluator is used to evaluate abort condition, we refer to *abort plan*. Using plans has advantages – especially when building complex conditions that cannot be easily explicitly expressed using sequential approach.
3. Thanks to the universal implementation in ABAsim architecture to defining evaluators, it will be possible to use other sophisticated methods, such as scripts or neural networks, in the future.

Reassessment plan is modified BDI plan, formally $P^{Reassessment} = (N^{Reassesment}, B)$. Modifications can be formally summarized as follows:

Since evaluator is a Boolean function, return value may be only True or False. This leads to modification of meaning of indexing type $Finish$ to $True$(or answer $Yes$); similarly, indexing type $Abort$ is considered as $False$ (or answer $No$). Depending on whether one defines a trigger plan (to determine if the main plan is executable) or an abort plan (to handle interruptions during plan execution), interpreting $Yes$ and $No$ node

have different meaning. Indexing set of reassessment plan is defined as:

$$T^{\text{Reassessment}} = \{Start, Yes, No, Operation, Decision\} \subset T$$

Nodes of reassessment plan must be of specified type, therefore:

$$N^{Reassessment} = \bigcup_{t \in T^{\text{Reassessment}}} N^t$$

We point out that all nodes of reassessment plan are interpreted in zero simulation time. Therefore, it has no impact on the execution of *actions*. The benefit of this approach lies

- in the possibility to use the same mechanism to define BDI plans and reassessment plans,
- in using of the same interpreter (since the type of nodes available to define reassessment plans are subset of the type of nodes used to define BDI plan),
- in the possibility to share beliefs between BDI plan and reassessment plan.

# 4 Experiments

ABAsim architecture with proposed modifications is part of the simulation software PedSim. This software is developed at the Faculty of Management Science and Informatics of the University of Žilina. While PedSim is primarily intended for pedestrian simulation, the versatility of ABAsim architecture allows us to utilize proposed mechanism in simulations of other systems as well – if sophisticated decision making of dynamic agents is relevant in those systems. We will use PedSim to demonstrate functionality of the proposed reevaluation of agent's goals.

PedSim provides an easy way of defining BDI plans and reassessment plans using a graph structure. This allows the simple definition of proposed BDI components of dynamic agents. Examples of such definitions are depicted in Fig. 6.

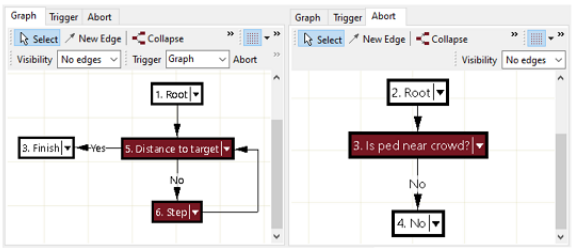



Figure 6. Definition of BDI plan and reassessment plan in PedSim

Proposed approach for reevaluation of intentions has been successfully tested through several experiments. These experiments focused on introducing strategies and creating situations where dynamic agents needed to reevaluate their initial plan choices.

As an example, let us mention a scenario in which we assigned the pedestrian, represented by a dynamic agent utilizing BDI reasoning, a goal to "reach the target area". While walking, pedestrian tried to avoid any crowd that appeared in his path. To monitor the presence of crowds, we utilized specialized sensor (component of ABAsim BDI paradigm) that updated belief "there is a crowd". The sensor was scanning polygonal area for presence of other pedestrians, see Fig. 7 for illustration.

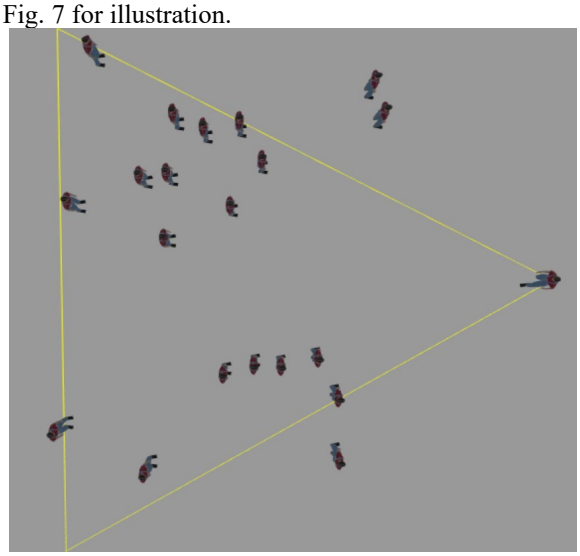


**Figure 7**. Visualization of polygon used by the crowd detection sensor

We defined two BDI plans with two different routings, as depicted in Fig. 8. First plan (in the figure referenced as "Experiment 3 – Plan 2") was more straightforward, however it led through typically crowded areas. The second plan (in the figure referenced as "Experiment 3 – Plan 3") was designed to avoid those areas, but the resulting path would be longer. Initially, the plan selection process involved a strategy based on defined priorities, with higher priority for the first plan.

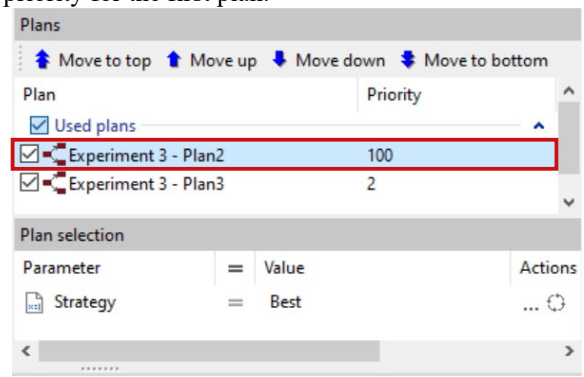



**Figure 8**. Definition of strategy to select from two plans based on priorities

When the sensor detected a crowd, the evaluator reevaluated the selected plan, decreased its priority and eventually aborted the plan. Subsequently, BDI reasoner changed the plan for the intention by using

strategy based on priorities. This led to selection of the other plan (originally with the second highest priority) to fulfill the agent's goal. Validation of the change of plan can be seen in a change of color representing agent's path from red to green, in Fig. 9, where colors are assigned according to used plans. The identification of the crowd on the square (blue area), led to a change in the plans. The pedestrian took longer route, but with higher personal preference according to current conditions caused by the presence of crowd.

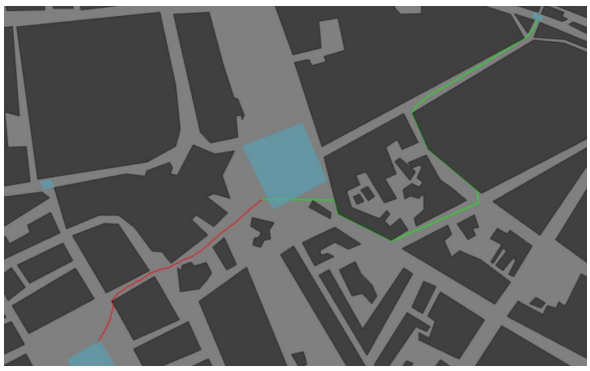


**Figure 9**. Change of routing plan of a pedestrian according to current situation in the simulation model

# 5 Conclusion and future work

With a focus on the dynamic agent's ability to reevaluate intentions, we have extended ABAsim architecture and thus enabled the creation of more reliable models based on human reasoning. Tools utilizing proposed approach will be capable to model the behavior of agents in complex systems more precisely.

Proposed mechanism lays foundation for future advancements. One potential direction we see is the integration of artificial intelligence. An agent would be capable of making the decisions and reevaluations based on its experiences and learning from similar situations, without the need for predefined reaction plans by the model designer. This would further enhance the modeling capabilities of simulation tools utilizing the ABAsim architecture.

# Acknowledgments

This work was supported by the grant VEGA 1/0654/22 Cost-effective design of combined charging infrastructure and efficient operation of electric vehicles in public transport in sustainable cities and regions

# References

Adamko, N. (2013). Agent-oriented simulation of complex service systems. University of Zilina.

Baier, J. A., & McIlraith, S. A. (2007). On Domain-Independent Heuristics for Planning with Qualitative Preferences. *Logical Formalizations of Commonsense Reasoning, Papers from the 2007 AAAI Spring Symposium*, Technical Report SS-07-05, Stanford, California, USA.

Hindriks, K.V., & van Riemsdijk, M.B. (2009). Using Temporal Logic to Integrate Goals and Qualitative Preferences into Agent Programming . *Declarative Agent Languages and Technologies VI*. Lecture Notes in Computer Science, vol 5397.

Bratman, M. (1987). Intentions, Plans and Practical Reason. Cambridge: Harvard University Press.

Dasgupta, A., & Ghose, A.K. (2011). BDI Agents with Objectives and Preferences. *Declarative Agent Languages and Technologies VIII*. Lecture Notes in Computer Science, vol 6619. Springer, Berlin, Heidelberg.

Jennings, N. R. (2001). An agent-based approach for building complex software systems. *Communications of the ACM*.

Kinny, D. N., & Georgeff, M. P. (1991). Commitment and Effectiveness of Situated Agents. *Sydney, Australia: Proceedings of theTwelfth International Joint Conference on Artificial Intelligence*.

Klima, V., Kavička, A., & Adamko, N. (2001). Software tool VirtuOs – simulation of railway station operation. *Proceedings of European simulation multiconference*, SCS, Prague, 2001, pp.84-88. ISBN 1-56555-225-3.

Siyam, N., Alqaryouti, O., & and Abdallah S. (2020). Research Issues in Agent-Based Simulation for Pedestrians Evacuation. *IEEE Access*, vol. 8, pp. 134435-134455. DOI 10.1109/ACCESS.2019.2956880.

Varga, M., & Adamko N. (2019). Definition of agent goals by means of missions. *2019 International Conference on Information and Digital Technologies (IDT)*. ISBN 978-1-7281-1401-9.

Varga, M., & Adamko N. (2014). Integration of BDI paradigm into ABAsim architecture. *Journal of Information Control and Management Systems,* 12.

Visser, S., Thangarajah, J., Harland, J., & Dignum, F. (2015). Preference-based reasoning in BDI agent systems. *Autonomous Agents and Multi-Agent Systems*.

Wang, J. (2021) Research on the Planning and Design of Transportation Hubs Based on Passenger Flow Simulation: Taking Tianjin Railway Station Transportation Hub as an Example. *2021 4th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA)*, Suzhou, China, pp. 1-6, DOI 10.1109/ISTTCA53489.2021.9654566.

# Bridging Academia and Industry: A Collaborative Approach to Software Development Education

**Lea Masnec, Zlatko Stapić**
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin
{lmasnec, zlatko.stapic}@foi.unizg.hr

**Maja Benkus**
SPAN d.d.
Koturaška cesta 47, 10 000 Zagreb
mbenkus@span.eu

***Abstract****. This paper showcases an innovative approach to coursework by integrating practical knowledge with specific challenges posed by companies. The Joint Creative Classroom (JCC) program, involving both universities and companies, has proven effective in teaching software engineering, focusing on software analysis and development. A student survey indicates significant improvement in knowledge across topics related to software development and project management. The paper also highlights curriculum planning and the use of contemporary tools. The primary aim is to demonstrate student-company networking through teaching within the JCC program.*

**Keywords:** education with industry collaboration, interdisciplinary learning, joint creative classroom program, networking, software development education

## 1 Introduction

Today's market demands are on the rise, leading to an increased need for a broader spectrum of knowledge necessary for operating and educating in the information technology sector. To stay up to date with market trends, it is important to invest efforts in educating potential employees, namely students. In addition to college courses providing essential knowledge in industry-related areas, collaboration with companies offers a significant advantage. Through such partnerships, students and faculty can gain first hand insights into market needs and conditions, potentially specializing in specific industry branches and acquiring competitive expertise.

The demand for students with practical experience is continuously increasing. Similarly, software development and the management of digital platforms are currently pivotal concepts for innovation and success in business (Scaturro, 2023). The rationale behind this research lies in recognizing the transformative nature of software development and the imperative to align educational programs with the evolving needs of the industry (Umar & Rana, 2024). We aim to dissect the intricacies of the JCC program, examining its role in skill development and practical knowledge acquisition. Additionally, this paper seeks to present the specific benefits and challenges associated with incorporating one of the leading cloud platforms, namely Microsoft Azure (Peddireddy, 2024), into the educational landscape. This research explores the concept of JCC as an integrative approach to training students in the field of software development. We focus on the skill development, ranging from the fundamental principles of digital platforms to the implementation of applications on the Microsoft Azure cloud to combine the knowledge learned in the program with the knowledge from the business profession.

The research question to be answered in this paper is: *What are the aspects of utility in integrating the JCC program into existing courses for a better comprehension of practical knowledge?*

The structure of the paper consists of five chapters contributing to the answer to stated research question. Firstly, we present the related work, and the current state-of-the-research and practice related to the agility and collaboration between industry and academia in education. Upon aligning the need for our research, we present our research method and performed activities in the third chapter. We have used a qualitative research methods involving both company and students. The backbone of the research is related to the questionnaire survey performed within the group of students who participated in the JCC program. In the fourth chapter the course and learning design are presented along with the obtained results. This chapter includes the rationale on defined learning outcomes as well as insights in final software solutions created by teams of students. However, the most important aspect of the chapter gives the results on received students feedback. In the final chapter of this paper the conclusions on this innovative approach are given.

## 2 Related works

Collaboration between the academic community and industry stands as a cornerstone for shaping innovative educational programs geared towards preparing

literature review that we performed shows that in the last years several authors have been investigating and contributing to this field of agile and collaborative teaching and learning.

Devedžić (2010) offers insights into agile software development education via a case study, aligning with the agile approach of JCC and emphasizing adaptability and continuous improvement in students' knowledge development. This focus on agility forms the crux of this scientific paper.

Matus et al. (2023) delve into factors influencing student satisfaction within JCC, assessing changes in individual student skills post-course completion, and the impact of student engagement and peer collaboration on satisfaction levels.

Gannod et al. (2015) and Liebenberg et al. (2015) explore the agile approach in education, with the latter underscoring the importance of tailoring educational programs to industry requirements, a core objective of JCC. Martínez Fernández et al. (2023) and Bocconi et al. (2012) contribute to the discourse by advocating project-based learning for teaching software analytics and data science best practices, enriching our understanding of collaborative education practices and fostering creativity in classrooms.

Easterling (2022) delves into collaboration dynamics among university faculties, pertinent to the theme of connecting with industry partners. Guo & Yang (2012), Layman (2006), and Nwodo (2023) further this understanding by providing insights into integrating education and industry in software development contexts, with a focus on project-based learning, changes in students' perceptions, and practical tools like Azure DevOps for collaborative software development.

Abegglen et al. (2021) stress the significance of partnerships in the educational ecosystem, echoing the ethos of JCC, while Malik & Zhu (2023) explore the impact of project-based learning on student success, reinforcing the practical dimension of JCC's objectives. Oda et al. (2017) assess engineering students' global competence through project-based learning, embracing the multidisciplinary approach of JCC.

Scarani et al. (2020) and Zantedeschi et al. (2020) offer insights into technological tools supporting collaborative learning and decentralized collaborative learning, respectively, enriching our understanding of collaborative software development aligned with JCC principles.

# 3 Method

This research relies on the three-phase methodological approach to explore the effectiveness of integrating the JCC program into existing courses in software development. This study has several limitations that should be considered when interpreting the results. The research was conducted at a single institution, which may limit the generalizability of the findings to a broader population. The sample consisted of a relatively small group of only 18 students, which could affect the statistical power and reliability of the quantitative analyses. In terms of data protection, special attention was given to the privacy and anonymity of the participants. The names of the students were not collected or mentioned in the study, and all survey responses were anonymous. These steps were taken in accordance with ethical guidelines for research to ensure the protection of personal data and the confidentiality of participant information.

## 3.1 Learning design

In the initial stage of the research, special attention was devoted to planning and developing the learning design for the JCC program. This process involved a comprehensive analysis of industry needs in software development to identify key themes and program objectives. By collaborating with industry representatives, we selected teaching and evaluation methods that best aligned with both student needs and industry expectations. This curriculum design effectively integrated practical skills with theoretical concepts, creating a well-rounded educational program that prepares students for the job market while ensuring compliance with the official accreditation standards required in university education.

## 3.2 Questionnaire survey

Quantitative evaluation of the program's effectiveness was conducted through a survey among students who took the program. The survey questionnaire was designed to quantitatively assess students' progress in understanding concepts of software development before and after participating in the JCC program. The questions in the survey were focused on various aspects of software development, and the responses used a scaled metric for quantitative analysis.

## 3.3 Qualitative evaluation by experts

In addition to quantitative evaluation, the qualitative aspect of the program's effectiveness was obtained through evaluation by experts participating in the JCC program. Company representatives assessed the practical skills and competencies students acquired through the program, providing deeper insights into the actual benefits of the program for students. Qualitative evaluation facilitated the assessment of students' readiness for the job market and identification of areas where the program could be further improved to better meet industry needs.

# 4 Program implementation and obtained results

Throughout the program, students successfully mastered key concepts and principles of digital platforms. They developed a profound understanding of the role of digital platforms in modern business. Agile software development became a daily practice, and students acquired the ability to craft high-quality user stories. Through active participation in projects on the Azure DevOps platform, they applied their acquired knowledge in real-world scenarios.

However, the overall process of planning and designing this program started a months before, incorporating the several learning design activities and phases. Thus, in this chapter we present both program planning and program execution aspects, including the results obtained with questionnaires given to the students before and after the program.

## 4.1 Learning design and program implementation

A previously mentioned comprehensive method of learning design was performed to ensure the best alignment of all teaching and learning activities and students' coursework with desired learning outcomes and overall workload of the students.

The JCC aspect was incorporated in a graduate level program *Analysis and program development* and was called *Agile Development of Digital Platforms*. Program syllabus covered various topics relevant to modern software engineering. Those topics are all covered in a way that only latest trends in software tools and methodologies were followed and considered. However, to build up on this university version of the program we wanted to replace the activities related to the students' project with the innovative version of the curriculum supported by industry experts from partnering company.

The tool used for learning design is Balanced Design Planning (BDP) Tool (Divjak et al., 2023) which is research-based, innovative, constructive-alignment oriented and modular solution based on learning outcomes (LOs) and learner workload as (Divjak et al., 2022). An excerpt from the tool showing the detailed plan of one teaching topic with several teaching units and different teaching and learning activities is presented in the Fig. 1.

The tool enabled us to define and connect all the activities with the learning outcomes and to explore the overall coverage of the learning outcomes with both teacher present and non-teacher present activities.
All topics covered are listed in section 4.2. in Table 1. Each topic was covered by lectures held by thirteen company experts. In classic academic courses all lectures are usually held by one or two course professors, so this was a great opportunity for students to have each topic covered by industry experts specialized for each topic. Also, they had an opportunity to experience different teaching styles.

The JCC program attended 18 students. Since one of the program goals was teamwork skills development, students were divided into four teams each containing four or five team members. Each team could choose one of three different project assignments. Team's progress was assessed in three project submissions. Evaluation criteria were clearly defined and available to students on program's *e-learning* page. Evaluation criteria followed topics covered in lectures which were held by the time assessment is done. Beside scores, in each project assessment students were provided with feedback about what was done well and what needs to be improved for the next phase. Project assessments were done by industry experts, so students had an opportunity to gain valuable feedback from industry standpoint. Since this program was project-oriented, scores gained through project submissions had a major weight in overall student score. Students were also assessed in three knowledge quizzes held during

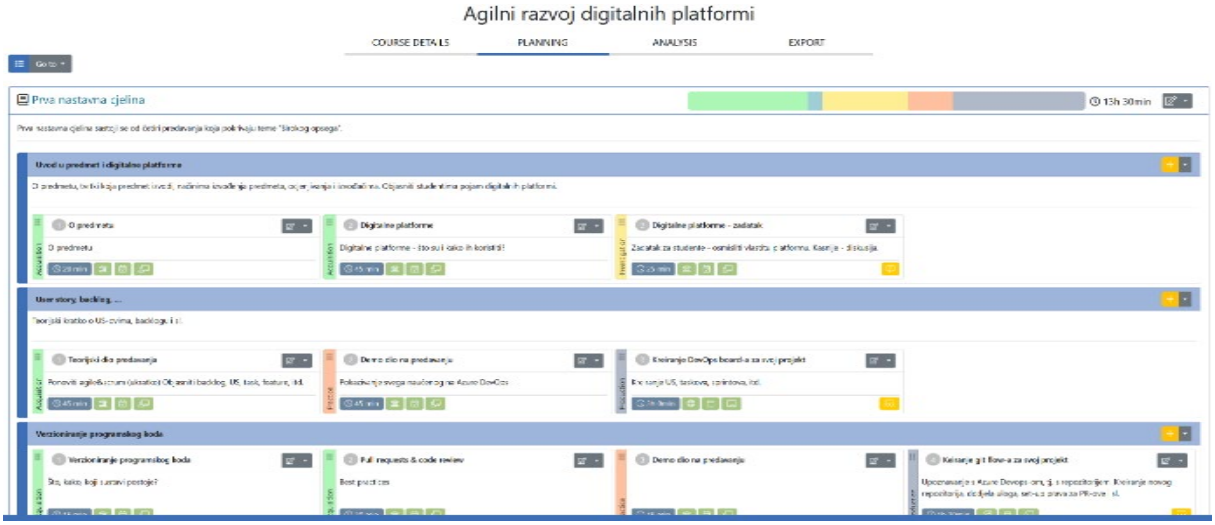



**Figure 1.** Excerpt from the learning design tool

program duration. In those quizzes theoretical knowledge is tested from the topics covered in lectures.

The results from the JCC aspect of the overall program were taken into account and joined with the results in other evaluation elements which were performed in the main program (such as assessment of theoretical knowledge) in order to define the final grade of the students.

## 4.2 Learning outcomes

Table 1 summarizes the achieved training outcomes of the JCC program, highlighting their innovation in applying tools like Azure DevOps and their practical utility in the real software development. The table shows the total of seven topics that were part of the JCC program and contribute to the practival knowledge of the students necessary to implement their project assignments. All other topics were part of the main university course and are not listed in this table.

**Table 1.** Achieved training outcomes

| Topic | Learning Outcome | Innovation on the Program | Practical Utility |
|---|---|---|---|
| Digital platforms and agile development | Understanding key concepts of digital platforms and applying agile principles in practice. | Implementation of Azure DevOps as part of agile development. | Improved team collaboration, faster development cycles. |
| User story & backlog | Ability to define, manage, and prioritize user stories and product backlog. | Use of Azure DevOps for backlog management and user story creation. | Enhanced clarity in project requirements and prioritization. |
| Code versioning and best practices | Ability to manage code versions and maintain code quality through best practices. | Active collaboration within the team through code evaluation. | Maintenance of clean and quality code. |
| Architecture (monolithic vs. microservices) | Understanding the differences between monolithic and microservice architectures and their applications. | Application of best practices in designing both types of architectures. | Development of scalable and sustainable applications. |
| Data layer and entity framework | Designing and implementing data layers and using the Entity Framework for database management. | Use of Appropriate data management techniques and tools. | Efficient data handling and management in applications. |
| Business logic and API layer | Developing robust business logic and effectively creating API layers. | Implementation of strategic API endpoints and business rules. | Creation of essential application functionalities. |
| Authentication, validation, and logging | Mastering security practices including authentication, validation, and logging. | Application of security measures and logging in real projects. | Development of secure, reliable applications with comprehensive logging. |
| Debugging and exception handling | Proficiency in debugging techniques and handling exceptions effectively. | Use of advanced debugging tools and methodologies. | Improved application stability and reliability. |
| Clean code and SOLID (principles of object oriented design) | Applying clean code principles and SOLID design patterns in software development. | Application of clean coding practices and SOLID principles in projects. | Development of maintainable and scalable software. |
| Microsoft Azure | Understanding and using Microsoft Azure for application deployment and development. | Utilization of Azure Cloud services for real-world projects. | Gained experience with cloud-based application development. |
| Deploy and CI/CD on Azure | Configuring deployment pipelines and | Implementation of CI/CD using | Faster and more reliable software |

| | | | |
|---|---|---|---|
| | automating CI/CD processes on Azure. | Azure DevOps for streamlined workflows. | deployment processes. |

## 4.3 Finalized student solutions

After working throughout the entire semester on the comprehensive solution, encompassing frontend, backend, Azure management, and documentation guided by agile development, all four student teams were required to showcase their final solutions to the company through a presentation and project defense. While students had already defended their projects in front of the faculty professors, the additional review within the company provided them with a unique opportunity to present their final product before industry experts. Notably, the company representatives who attended the presentations had also served as mentors to the teams throughout the semester. Consequently, students had the invaluable chance to receive feedback directly from experts in the field on which they had been working.
Fig. 2 illustrates the students' presentation within the company.

**Figure 2.** Presentation of student work in *the SPAN d.d.*

This culmination event served as a platform for the teams to demonstrate their achievements and innovations to professionals within the industry, creating a bridge between academic learning and real-world application. It further reinforced the collaborative nature of the educational program, emphasizing the practical relevance and impact of the students' efforts on the industry landscape.

Following the questionnaire-based assessment, the utilization of qualitative evaluation by companies becomes evident.

The qualitative evaluation of the JCC program was conducted in three phases, where participants projects were systematically assessed based on predefined criteria. Each evaluation phase involved a thorough review of the projects, reassessment of identified issues, and the provision of detailed feedback to the teams, which facilitated continuous improvement and a significant increase in the quality of the projects. The evaluation criteria included project management using Azure DevOps, technical implementation according to best practices (such as layered architecture, clean code, and SOLID principles), functionality coverage and accuracy, successful deployment on Azure Cloud, and the quality of the final project presentation.

The evaluation process involved experts from various fields: Software Developers and Business Analysts assessed business analysis, git flow, and architecture, while two Software Developers evaluated the programming aspect of the projects. The final project defense was assessed by considering the feedback from all evaluators, including Team Leads, Software Architects, and QA specialists. During the evaluation process, each evaluator independently reviewed and tested the projects and assigned scores based on the defined criteria. After individual assessments, the evaluators convened to discuss their scores and reach a consensus on the final ratings, ensuring fairness and objectivity.

The evaluation results were documented through evaluators notes, feedback provided to the teams, and Teams received personalized comments with specific recommendations for improvement, such as code optimization and better project management. Throughout the iterative evaluation process, there was a notable improvement in the quality of the solutions, underscoring the importance of continuous support and feedback in the development of participants skills.

## 4.4 Analysis of student feedback

For the purpose of this scientific study, a survey questionnaire was conducted to evaluate the program's impact on students' knowledge across various topics related to software development and project management. Out of the 18 students enrolled in the program, only 11 chose to complete the survey. This response rate reflects the voluntary nature of survey participation, where not all students opted to provide their feedback. It is important to note that the decision to participate in the survey was left entirely up to the students, which may have influenced the number of responses received. The arithmetic mean of their scores for each topic before and after the program was presented in Fig. 3. Students assessed their knowledge on 11 topics using a scale from 1 to 5, where 1 indicates "no knowledge" and 5 indicates "complete knowledge". The topics included digital platforms, user stories, code versioning and best practices, software architecture, data layers, business logic, security, debugging, clean code, Microsoft Azure, and CI/CD. The top line with the topic title represents the knowledge before, while the bottom line represents the knowledge after.

The survey results show a significant increase in average scores across all topics after completing the

program (Carver et al., 2010). The most substantial improvements were observed in *User Story & Backlog* from 2.73 to 4.55 and *Architecture (Monolithic vs. Microservices)* from 2.36 to 4.09. These topics are crucial for successful project management and software development, making the knowledge gains in these areas particularly noteworthy.

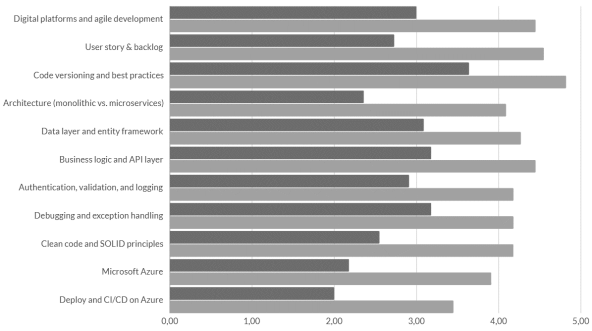



**Figure 3.** Analysing student proficiency in program topics

The average score for *Code versioning and Best Practices* also saw a significant rise, from 3.64 to 4.82, indicating successful adoption of best practices in code management. This is important because proper version control reduces the risk of errors and enhances team collaboration.

Security-related topics, such as *Authentication, Validation, and Logging*, increased from 2.91 to 4.18, while *Debugging and Exception Handling* rose from 3.18 to 4.18. These results demonstrate that students have gained a deeper understanding and the skills necessary to develop secure and reliable applications.

The smallest but still significant increases were observed in topics related to Microsoft Azure. *Deploy and CI/CD on Azure* increased from 2.00 to 3.45, while *Microsoft Azure* rose from 2.18 to 3.91. Although these results show significant progress, the relatively lower initial scores indicate that these areas were less familiar to students before the program. Nonetheless, the increase reflects successful transfer of fundamental knowledge and skills for using cloud platforms.

# 5 Conclusion

In the conclusion of this research, we emphasize the importance of collaboration between the academic community and industry in the development of innovative educational programs. Through the analysis of various approaches to agile software development, research on collaborative initiatives such as the JCC, and the application of project-based learning, key guidelines for successful integration of education and industry have been identified. Case studies of four student teams, who developed comprehensive solutions throughout the semester, underscore the significance of practical application of acquired knowledge. Presentations before industry experts allowed students not only to convey their creative ideas but also to receive constructive feedback directly from industrial professionals who served as mentors during the development process.

Our study confirms that collaboration with the industry provides students with useful insights, encourages a practical approach to learning, and enables the development of skills necessary for successful functioning in the job market. It is noteworthy to highlight that the agile development approach, supported by mentorship from industry experts, has empowered students not only to develop software solutions but also to gain relevant experience and skills essential for future professional endeavours.

The survey results indicate that the JCC program has significantly improved students' knowledge across all analysed topics. The greatest progress was achieved in project management and software architecture, while knowledge in Microsoft Azure and CI/CD processes also showed substantial improvement. These results confirm the program's effectiveness in preparing students for the challenges of modern software development and IT project management.

The research question we addressed in the research was: *What are the aspects of utility in integrating the JCC program into existing programs for a better comprehension of practical knowledge?* With the execution of all activities, starting from the research design throughout the implementation and feedback analysis we conclude that the integration of JCC program into existing programs emerges as a transformative approach that significantly enhances students' comprehension of practical knowledge in the realm of software development. This conclusion is rooted in several key aspects that underscore the utility of JCC integration:

- *Real-World application as a learning cornerstone*: JCC's emphasis on real-world application serves as a cornerstone for practical knowledge acquisition. By engaging students in hands-on projects mirroring industry scenarios, the program bridges the gap between theoretical understanding and pragmatic implementation, ensuring a more profound comprehension of software development concepts.
- *Industry mentorship as a guiding light*: The integration of industry experts as mentors throughout the JCC program brings a unique utility by providing students with direct guidance from professionals actively involved in the field. This mentorship model fosters a deeper understanding of industry best practices, offering students insights that extend beyond theoretical frameworks.
- *Technological relevance for current industry standards*: The utility of JCC lies in its commitment to technological relevance. Alignment with current technologies equips students with skills directly applicable to the ever-evolving landscape of software development.

*Continuous feedback loop for iterative improvement*: Integral to the JCC integration is the establishment of a continuous feedback loop. Regular interactions with industry mentors and presentations to experts create a platform for iterative improvement. This ongoing feedback mechanism refines students' practical skills, encouraging constant learning and adaptation.

The suggestion for further research is to investigate how continuous feedback and iterative improvements affect long-term retention and application of software development skills among students.

In essence, the integration of JCC into existing programs not only enriches theoretical understanding but propels students towards a practical comprehension of developing digital platforms. By addressing the dynamics of real-world application, mentorship, collaboration, technological relevance, continuous improvement, and job market preparedness, JCC emerges as a powerful catalyst in shaping the next generation of skilled and industry-ready software developers.

# References


Abegglen, S., Burns, T., & Sinfield, S. (2021). Collaboration in higher education: Partnering with students, colleagues and external stakeholders. Journal of University Teaching & Learning Practice, 18(7), 1-6. doi:10.53761/1.18.7.01

Bocconi, S., Kampylis, P., & Punie, Y. (2012). Innovating teaching and learning practices: Key elements for developing creative classrooms in Europe. eLearning papers, 30, 1-13.

Carver, J. C., Jaccheri, L., Morasca, S., & Shull, F. (2010). A checklist for integrating student empirical studies with research and teaching goals. Empirical Software Engineering, 15, 35-59. doi:10.1007/s10664-009-9109-9

Devedžić, V. (2010). Teaching agile software development: A case study. IEEE transactions on Education, 54(2), 273-278. doi:10.1109/TE.2010.2052104

Divjak, B., Bađari, J., Grabar, D., Horvat, D., Vondra, P., & Rienties, B. (2023). Enhancing Learning Design through User Experience Research: Insights from a Survey in Four European Countries. In *Central European Conference on Information and Intelligent Systems* (pp. 213-221). Faculty of Organization and Informatics Varaždin.

Divjak, B., Grabar, D., Svetec, B. & Vondra, P. (2022). Balanced Learning Design Planning: Concept and Tool. Journal of information and organizational sciences, 46 (2022), 2; 361-375 doi:10.31341/jios.46.2.6

Easterling, L. A. (2022). Collaborative Faculty Development Activities in Higher Education Institutions: An Activity System Analysis. Indiana University.

Gannod, G. C., Troy, D. A., Luczaj, J. E., & Rover, D. T. (2015). Agile way of educating. In 2015 IEEE Frontiers in Education Conference (FIE) (pp. 1-3). IEEE. doi:10.1109/FIE.2015.7344019

Guo, S., & Yang, Y. (2012). Project-based learning: An effective approach to link teacher professional development and students learning. Journal of Educational Technology Development and Exchange (JETDE), 5(2), 5. doi: 10.18785/jetde.0502.04

Layman, L. (2006). Changing students' perceptions: an analysis of the supplementary benefits of collaborative software development. In 19th Conference on Software Engineering Education & Training (CSEET'06) (pp. 159-166). IEEE. doi:10.1109/CSEET.2006.10

Liebenberg, J., Huisman, M., & Mentz, E. (2015). The relevance of software development education for students. IEEE Transactions on Education, 58(4), 242-248. doi:10.1109/TE.2014.2381599

Malik, K. M., & Zhu, M. (2023). Do project-based learning, hands-on activities, and flipped teaching enhance student's learning of introductory theoretical computing classes?. Education and Information Technologies, 28(3), 3581-3604. doi:10.1007/s10639-022-11350-8

Martínez Fernández, S. J., Gómez Seoane, C., & Lenarduzzi, V. (2023). Applying project-based learning to teach software analytics and best practices in data science. International journal of engineering education, 39(2), 476-487.

Matus, M., Hajdin, G., & Balaban, I. (2023). Joint Creative Classroom as a Response to Industry 4.0 Demands. *In Central European Conference on Information and Intelligent Systems* (pp. 199-204). Faculty of Organization and Informatics Varaždin.

Nwodo, A. (2023). Beginning Azure Devops: Understanding and Using Azure Developer Services to Plan Work, Collaborate on Code Development, and Build and Deploy Applications. Wiley. isbn:978-1-394-16588-9

Oda, S., Inoue, M., & Yamazaki, A. K. (2017, November). Assessment of global competency for engineering students in a multicultural and multidisciplinary project based learning course. In *2017 7th World Engineering Education Forum (WEEF)* (pp. 439-443). IEEE. doi:10.1109/WEEF.2017.8467071

Peddireddy, K. (2024). Cloud Computing Platforms. In Emerging Trends in Cloud Computing Analytics,

Scalability, and Service Models (pp. 186-199). IGI Global.

Scarani, S., Muñoz, A., Serquera, J., Sastre, J., & Dannenberg, R. B. (2020). Software for interactive and collaborative creation in the classroom and beyond: An overview of the soundcool software. Computer Music Journal, 43(4), 12-24.

Scaturro Heil, M. R. (2023). The Power of Digital Platforms: Facilitating the Collaboration of Undergraduate Computer Science Project Teams. International Association for Development of the Information Society. *Proceedings of the 20th International Conference on Cognition and Exploratory Learning in the Digital Age*, Madeira Island, Portugal.

Umar, U. S., Rana, M. E. (2024, January). Cloud Revolution in Manufacturing: Exploring Benefits, Applications, and Challenges in the Era of Digital Transformation. In 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS) (pp. 1890-1897). IEEE. doi:10.1109/ICETSIS61505.2024.10459473

Zantedeschi, V., Bellet, A., & Tommasi, M. (2020, June). Fully decentralized joint learning of personalized models and collaboration graphs. In International Conference on Artificial Intelligence and Statistics (pp. 864-874). PMLR.

# 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 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. Decisions 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 the 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 directors.

### 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 publishers 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 editors 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 (Computer Games, Cyber-Physical and Embedded systems, Data and Knowledge Bases, Digital transformation and corporate culture, Innovation, entrepreneurship and economic growth, Intelligent Energy, Intelligent Systems, Learning analytics, AI and technologies in education, Strategic planning and decision making, Systems for Smart Mobility, Trends in Software Engineering) 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.

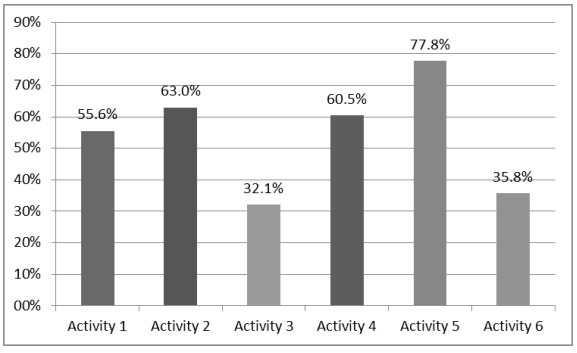



**Figure 1.** Times, 10 points, normal face

## 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, Cyber-Physical and Embedded systems, Data and Knowledge Bases, Digital transformation and corporate culture, Innovation, entrepreneurship and economic growth, Intelligent Energy, Intelligent Systems, Learning analytics, AI and technologies in education, Strategic planning and decision making, Systems for Smart Mobility, Trends in Software Engineering) 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.

# 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, Cyber-fizički i ugrađeni sustavi, Baze podataka i baze znanja, Digitalne transformacije i korporativna kultura, Inovacije, poduzetništvo i ekonomski rast, Pametna energija, Inteligentni sustavi, Analitike učenja, umjetna inteligencija i tehnologije u obrazovanju, Strateško planiranje i donošenje odluka, Sustavi za pametnu mobilnost, Trendovi u softverskom inženjerstvu), 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

# 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).

# 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.

- Autor(i), pripadnost(i), adrese i e-mail adrese
- Sažetak
- 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.

# 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 `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.

# 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 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). 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.

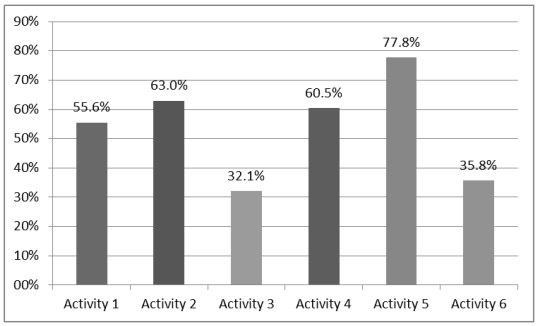



**Figure 1.** Times, 10 points, normal face

## 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

## 9.1 Slike

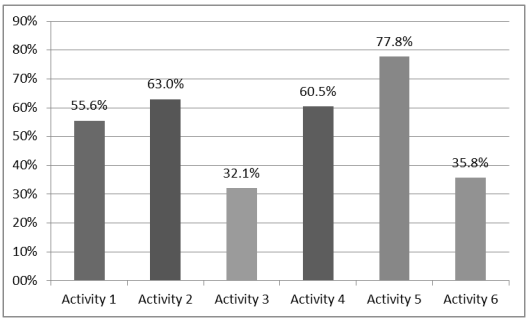
Sve slike trebaju biti uključene u tekst. Nazivi slika trebaju biti ispod slika, u fontu Times veličine 10 (ili sličnom fontu). Samo prvu riječ svakog naziva slike pišite velikim početnim slovom.

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 |

| | 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 (*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

[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).

### 9.3 Jednadžbe

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.

[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).

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/

## 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 2024 Sponsors

CECIIS 2024 GOLD SPONSOR

INFODOM

# CECIIS 2024 GOLD SPONSOR

# CECIIS 2024 SILVER SPONSORS

City of Novi Marof

HBOR

# CECIIS 2024 BRONZE SPONSORS

Varaždin County

VARAŽDIN

City of Varaždin

# CECIIS 2024 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
Emer. Leo Budin, University of Zagreb, Croatia
Tonči Carić, University of Zagreb, Croatia
Carlos Carrascosa, Universitat Politècnica de València, Spain
Aruna Chandra, Indiana State University, USA
Luis de Marcos, 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 Gašpar, 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
Marina Klačmer Čalopa, University of Zagreb, Croatia
Michal Kohani, University of Zilina, Slovakia
Mario Konecki, University of Zagreb, Croatia
Vladimir Kovšca, 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
Peter Marton, University of Žilina, Slovakia
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, CAMPUS 02 - University of Applied Sciences, Austria
Emer. 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
Marko Tadić, University of Zagreb, Croatia
Martina Tomičić Furjan, University of Zagreb, Croatia
Katarina Tomičić Pupek, University of Zagreb, Croatia
Jan Tretmans, Radboud University Nijmegen, 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

# Honorary Members - Former Program Committee Chairs

Tihomir Hunjak, University of Zagreb, Croatia
Mirko Maleković, University of Zagreb, Croatia
Boris Aurer, University of Zagreb, Croatia

# CECIIS 2024 Organizing Committee

Dina Korent, Organizing Committee Chair, University of Zagreb, Croatia

Ena Barčić, University of Zagreb, Croatia
Marijana Bubanić, University of Zagreb, Croatia
Antonela Devčić, University of Zagreb, Croatia
Martina Đuras Sekovanić, University of Zagreb, Croatia
Ivana Fojs, University of Zagreb, Croatia
Lea Masnec, University of Zagreb, Croatia
Marko Matus, University of Zagreb, Croatia
Jelena Gusić Munđar, University of Zagreb, Croatia
Karlo Nadoveza, University of Zagreb, Croatia
Karlo Pepelnik, University of Zagreb, Croatia
Laura Strupar, University of Zagreb, Croatia

# CECIIS 2024 Research Tracks' Chairs

**Computer Games:** Markus Schatten, University of Zagreb, Croatia

**Cyber-Physical and Embedded Systems:** Ivan Magdalenić, University of Zagreb, Croatia

**Digital Transformation and Corporate Culture:** Katarina Tomičić Pupek, University of Zagreb, Croatia

**Innovation, Entrepreneurship and Economic Growth:** Vladimir Kovšca, University of Zagreb, Croatia

**Intelligent Energy:** Danijel Topić, University of Osijek, Croatia

**Intelligent Systems:** Jasminka Dobša, University of Zagreb, Croatia and Dunja Mladenić, Jožef Stefan Institute, Slovenia

**Learning Analytics, AI and Technologies in Education:** Blaženka Divjak, University of Zagreb, Croatia and Igor Balaban, University of Zagreb, Croatia

**Strategic Planning and Decision Making:** Nina Begičević Ređep, University of Zagreb, Croatia

**Trends in Software Engineering:** Marko Mijač, University of Zagreb, Croatia

# CECIIS 2024 Poster Section Chairs

Dina Korent, University of Zagreb, Croatia

Marijana Bubanić, University of Zagreb, Croatia

# CECIIS 2024 Student Poster and Presentation Section Chairs

Violeta Vidaček Hainš, University of Zagreb, Croatia

Victoria Appatova, University of Cincinnati, USA

Antonela Devčić, University of Zagreb, Croatia

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