Preserving Human Agency in AI-Driven Interactions: The Role of User Experience Design

Hanlie Smuts, Blossom Alloziem, Catherine Dyason, Keiara Govender
University of Pretoria
Department of Informatics
Corner of Lynnwood Road and Roper Street, Hatfield, Pretoria, 0083, South Africa
hanlie.smuts@up.ac.za,{u22604244, u23573806, u23689197}@tuks.co.za

Abstract. *Artificial intelligence (AI) is a foundational component of information systems, influencing decisions across all industries and social platforms. Product and software solution developers devote considerable attention to improving algorithmic performance, with less focus on how customers, communities, and users experience, understand, and influence AI systems. Therefore, this paper seeks to explain how user experience (UX) design can be applied to preserve human agency in AI-driven interactions. The authors analyzed 129 papers using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and derived four UX design considerations: trust, transparency and explainability, inclusion and accessibility, and embedded ethics as key dimensions of responsible AI design preserving human agency. These considerations shift the focus of AI design from technological capability toward human empowerment and agency. Preserving human agency requires more than technical accuracy; it requires systems that support meaningful human participation and oversight.*

Keywords. Artificial Intelligence, User Experience, Explainable AI, Participatory Design, Community Empowerment, AI Ethics

1 Introduction

Artificial intelligence (AI) is no longer an emerging technology. AI is embedded in the infrastructure of everyday life, and AI-driven systems increasingly influence decision-making in education, employment, healthcare, financial services, and other sectors. (Mishra, 2024). This integration of AI technologies shapes opportunities, allocates resources, facilitates interactions between individuals and institutions, and consequently, has become more than just a technological tool (Naikar et al., 2026). Sophisticated AI systems are perceived as more than just a tool for business operations efficacy. It is regarded as a key strategic asset with the potential to produce never-before-experienced outcomes in business, such as hyper-personalization at scale, real-time decision-making, hidden customer needs prediction, and inimitable solution design (Dzreke, 2025). AI systems also provide new opportunities for innovation and value creation beyond realizing operational efficiency (Ali et al., 2024).

In addition, organizations can make more informed decisions and respond more effectively in dynamic environments due to the ability of AI systems to process vast amounts of data (big data), identify complex patterns, and generate actionable insights - in some instances, autonomously (Li & Tian, 2026; Nagaboyina, 2024). AI technologies have the potential to augment human expertise, enhance creativity and support productivity (Smuts & Smith, 2021).

Despite these advances in AI systems, concerns have emerged about fairness, transparency, accountability, and inclusion (Akinrinola et al., 2024; Panarese et al., 2025). Stakeholders (communities, customers, users, and employees) are impacted by opaque algorithmic decisions, creating challenges for engendering trust and establishing democratic oversight (Panarese et al., 2025). Furthermore, stakeholders most impacted by AI remain underrepresented in the design, development and governance of such systems (Kaminski & Malgieri, 2024). Consequently, AI systems risk reinforcing existing inequalities, while simultaneously obscuring the mechanisms through which such inequalities are reproduced (Bircan & Özbilgin, 2025).

The socio-technical nature of user experience (UX) is associated with the usability and usefulness of interface design, mediating relationships between people (stakeholders) and intelligent systems (Cha & Wong, 2025). Through decisions about interaction design, accessibility, explainability, and participation, UX practitioners influence whether AI systems empower users or diminish their agency (Bennett et al., 2023). Therefore, the research question investigated in this paper asked: "*How can user experience design be applied to preserve human agency in AI-driven interactions?*" The findings of this study could assist

AI solution designers in developing AI-driven systems that enhance user autonomy, support informed decision-making, and maintain meaningful human control throughout the interactive process.

Section 2 presents the background of the study, followed by the research methodology in Section 3 and the findings in Section 4. Section 5 presents the study contributions, and Section 6 concludes the paper.

2 Background

2.1 Human agency and explainability in AI systems

The rapid adoption of AI has transformed the relationship between humans and technology (Naikar et al., 2026). Such automated systems increasingly influence decisions and recommendations traditionally made by humans. Scholars report that when stakeholders cannot understand or challenge algorithmic outputs, their capacity for informed decision-making is reduced (Grimmelikhuijsen & Meijer, 2022; Heaton et al., 2023; Masoudi, 2025). Human agency is the individual capacity to make independent choices, take purposeful actions, and shape their own lives and environment. It contrasts with the idea that humans are just passive reactors to natural forces or societal rules; rather, it embodies the power to think, decide, and enact change (Welzel & Inglehart, 2010). Human agency in AI contexts refers to stakeholders' ability to understand, influence, and contest decisions that affect them (Holter & El-Assady, 2024).

Scholars recognize trust as a prerequisite for successful human-AI interaction (Glikson & Woolley, 2020; Thinnakkakath, 2026). Accordingly, research in explainable AI (XAI) has focused on developing methods that make algorithmic reasoning more interpretable for stakeholders to understand how AI systems function, why decisions are made, and what limitations exist. Effective explanations should be understandable, actionable, and contextually relevant to stakeholders' knowledge and circumstances (Hannibal & Lindner, 2023; Subramanian et al., 2024; Wang et al., 2025). Hannibal and Lindner (2023) elaborate that AI systems should not proffer superficial explanations that create the appearance of openness yet do not enable true understanding. In this context, UX design is vital to ensuring that transparency mechanisms indeed empower users rather than merely legitimize opaque systems (Gray et al., 2018).

Data represent another key resource related to XAI (Durango et al., 2025; Meier, 2025). Carroll et al. (2020) define the people- and purpose-oriented CARE principles (**C**ollective benefit, **A**uthority to control, **R**esponsibility, and **E**thics), which reflect the important role of data in advancing governance and human agency. These principles position stakeholders as data stewards rather than just data sources to attain personal data sovereignty and governance (Baraku et al., 2025). Data sovereignty should be designed to grant stakeholders practical control over their distributed personal data, especially regarding its usage in AI training (Baraku et al., 2025).

2.2 Inclusion and participatory design

The transition from algorithms to human agency is necessary to ensure that AI systems promote equality, accountability, and human dignity (Papagiannidis et al., 2025). Therefore, AI systems should not reproduce social inequalities when they are developed within unequal social contexts (Lainjo, 2023). Such development could cause marginalized groups to experience barriers related to digital access, language support, accessibility, and representation in training datasets. Older adults, rural communities, low-income populations, minority language-speakers, and individuals with disabilities are vulnerable to exclusion (Cunningham et al., 2025; Panarese et al., 2025).

Kang et al. (2026) propose that UX design principles provide a useful framework for addressing these challenges, as inclusion should not be treated as a compliance requirement but rather as a foundational design principle that benefits all stakeholders (Ntoa et al., 2024). Therefore, within the AI context, accessibility extends beyond interface usability to include accessible explanations, adaptive interaction models, and equitable access to algorithmic services (Panarese et al., 2025). Rather than treating stakeholders as passive recipients of technology, UX design positions them as active contributors and co-creators (Olsson, 2004). Hence, participatory design ensures that stakeholders are directly involved in the design process by participating in defining problems, selecting datasets, and evaluating outcomes (Bødker et al., 2022).

2.3 Human agency and oversight models

The tension between human agency and automation in organizations suggests that organizations should balance their productivity gains with professional (human) accountability (Taş et al., 2026). Therefore, oversight routines vary across different task types as transactional work offers limited intervention points, whereas collaborative and expert (knowledge) work provide more opportunities to assert human agency (Krakowski, 2025). In this context, where humans maintain meaningful control in human-AI interaction, the socio-technical nature of AI-enabled work is emphasized (Taş et al., 2026).

To operationalize human agency, human oversight models have been proposed. Based on research over decades, scholars applied these models to human-AI interaction. The Human-in-the-Loop (HITL) model suggest direct human involvement in the AI decision-making process. Human review, approve or modify AI outputs before actions are implemented (EU, 2018;

Kabata & Thaldar, 2024; Singh & Szajnfarber, 2026). HITL is typically applied in high-risk environments such as healthcare, criminal justice and finance (Kabata & Thaldar, 2024).

The Human-on-the-Loop (HOTL) model offers lower level of intervention as in this instance, AI systems operate autonomously while a human supervisor continuously monitors its performance. The human operator can intervene or override decisions when necessary (EU, 2018; Maadi et al., 2021; Singh & Szajnfarber, 2026). HOTL is typically applied in semi-autonomous environments where human supervision is essential for safety and accountability, e.g., autopilot in commercial aviation or in autonomous vehicles.

In the Human-in-Control (HIC) model, humans retain ultimate authority over the deployment, objectives, operation, and governance of AI systems rather than being based on individual decisions (EU, 2018; Singh & Szajnfarber, 2026). Applying the HIC model, humans maintain strategic control (as opposed to operational monitoring only) as they determine when AI should be used, define the operational boundaries, establish governance mechanisms, and remain accountable for system outcomes (Taş et al., 2026).

Taş et al. (2026) present the concept of *episodic oversight sequencing* which suggests that human oversight should not be confined to a single approval point, but rather distributed across multiple workflow stages.

3 Research methodology

This paper aimed to investigate how UX design could be applied to preserve human agency in AI-driven interactions. The study achieved this aim by applying the analysis process suggested by Donthu et al. (2021), consisting of four steps: defining the aims and scope of the study; choosing the techniques for analysis; collecting the data for analysis and executing the analysis; and reporting the findings. The researchers applied the enrichment technique and created a network analysis (Donthu et al., 2021).

The study executed a keyword search based on the research objectives and to collect the data for the analysis in *Scopus*, a large literature repository of peer-reviewed abstract and citation data, applying the following keywords: (TITLE-ABS-KEY (human agency) AND TITLE-ABS-KEY (artificial intelligence) AND TITLE-ABS-KEY (human centered design)) AND (LIMIT-TO (LANGUAGE , "English")). We specifically chose these keywords as we followed a focused strategy centered explicitly on human agency and UX as opposed to broader frameworks such as autonomy, human control, user agency and human-centered AI. The keyword search yielded a result set of 160 papers. The authors screened the papers and removed one duplicate, 18 papers unrelated to the research topic (e.g., assisted building design, media innovation, biomedical visualization, children's fables, agriculture study, or animals), and 11 papers that indeed related to the subject area but did not explicitly deal with human agency. Fig. 1 depicts the PRISMA reporting that resulted in 129 papers from the corpus for analysis.

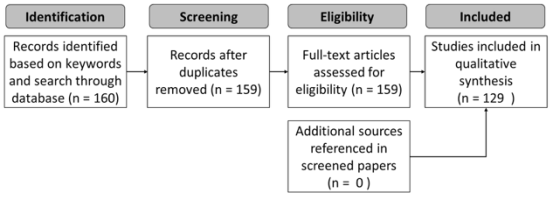


Figure 1. PRISMA reporting

Fig. 2 demonstrates a gradual increase in the number of included studies over time, from a single study in 2014 and 2017–2018 to 13 studies in 2024. The highest number of studies was published in 2025 (n = 54), followed by 2026 (n = 34). Overall, the findings indicate a substantial growth in research activity in recent years, with the majority of included studies published between 2024 and 2026. It must be noted that, as a cross-sectional study that was completed at a snapshot in time, the citations for 2026 will still increase.

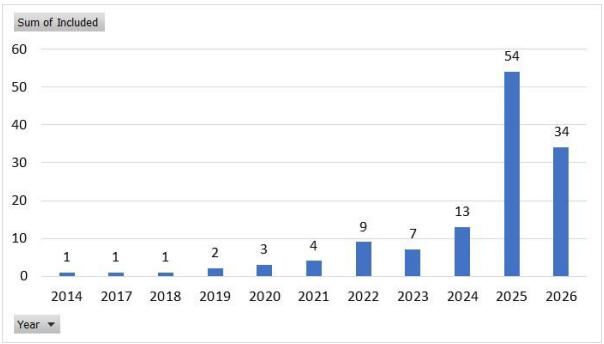


Figure 2. Year of publication

Table 1 presents the distribution of the 129 publications according to publication type. Journal articles constituted the largest proportion of the corpus, accounting for 59 publications (46%). Conference papers represented the second-largest category, with 51 publications (40%). Book chapters accounted for 12 publications (9%), and books contributed five publications (4%). These publication types were less prevalent, yet they provided valuable theoretical perspectives and comprehensive discussions that complemented the journal and conference literature. Editorials constituted the smallest category, with only two publications (2%).

Table 1. Paper type

Publication type	n	%

Article	59	46%
Book	5	4%
Book chapter	12	9%
Conference paper	51	40%
Editorial	2	2%
Grand Total	129	100%

4 Data analysis

The bibliometric information in the 129 papers was downloaded as a .RIS file and imported into *VOSviewer* version 1.6.20 (https://app.vosviewer.com/). *VOSviewer* constructed and visualized the co-occurrence data of the corpus, depicted in Fig. 3. The co-occurrence network consisted of four clusters extracted from 5695 terms, with 120 terms meeting the threshold of a minimum of ten occurrences per term.

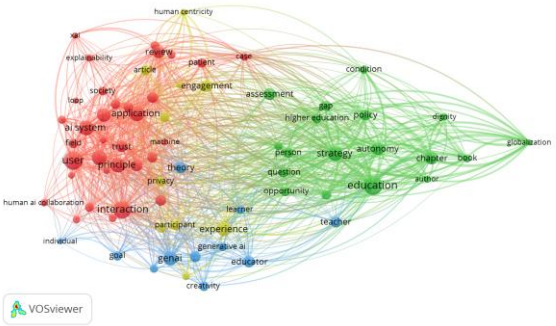


Figure 3. UX, Human agency, and AI cluster analysis

Table 2. Cluster analysis description

	Items (n)	Derived cluster label	Cluster description
1	33	Trust	Trust must be built through consistent accountability mechanisms and genuine stakeholder responsiveness.
2	18	Transparency and Explainability	Stakeholders must be able to understand, contest, and shape the AI systems that affect their lives, and not merely receive their outputs.
3	11	Inclusion and Accessibility	Inclusive design should consider who the design excludes and what is required to reach them.
4	10	Embedded Ethics	Ethics must be a structural feature of the design process, governance, and iteration cycle.

Table 2 presents the four clusters in more detail, highlighting the cluster number, the number of themes in a cluster, and a detailed list of the themes. A cluster label and a short description were derived for each cluster from the cluster themes.

5 Results

This paper investigated the role of UX design in preserving human agency in AI-driven interactions. AI systems consist of algorithms, datasets, predictive models, and automated decision-making processes. Stakeholders bring lived experiences, cultural knowledge, values, needs, and expectations. UX may function as the socio-technical interface through which these domains interact. Based on the bibliographic coupling of the corpus, the clusters identified in Table 2 are described in more detail in the following sections and visualized in Fig. 4.

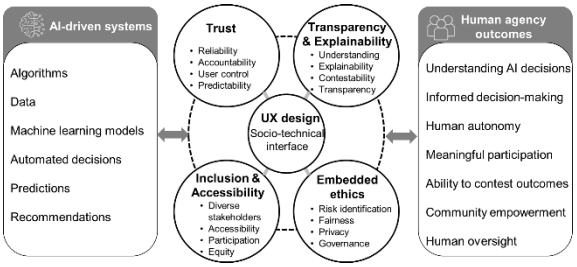


Figure 4. UX principles for preserving human agency in AI-driven interactions

Fig. 4 illustrates the relationship between AI-driven systems, user experience (UX) design, and human agency outcomes. On the left, AI-driven systems are depicted as algorithms, data, machine learning models, automated decisions, and predictive recommendations that influence user interactions. At the centre of the visualization is User Experience (UX) Design, which serves as a socio-technical interface connecting AI systems with human outcomes. Four key UX design principles surround the central component. Together, these design principles shape positive human agency outcomes, including informed decision-making, human autonomy, meaningful participation, community empowerment, and human oversight of AI-enabled decisions.

5.1 Trust

Trust emerged as a key requirement for preserving human agency in AI-driven interactions. Stakeholders are more willing to engage with AI systems when they perceive them as reliable, predictable, and aligned with their interests (Bach et al., 2024; Glikson & Woolley, 2020). UX engenders trust through interaction

mechanisms that enable stakeholders to understand, monitor, and influence AI system behaviour, e.g., when farmers use an AI crop advisory system but can compare AI recommendations with historical local farming practices. In the context of trust, UX design emphasizes the importance of communicating system capabilities, limitations, confidence levels, and uncertainty to support appropriate reliance rather than merely accepting algorithmic outputs (Stimson & Raper, 2025). This status is achieved by creating interfaces that provide feedback, support user control, and enable system predictability across contexts (Nam & Ki Chan, 2025; Renata et al., 2023). Thus, UX design encourages informed reliance on AI systems, while providing opportunities for human judgment and intervention (Doh et al., 2025).

5.2 Transparency and explainability

Stakeholders can understand and contest AI-driven systems through a transparency and explainability lens. UX design holds that explanations should be meaningful, contextual, and tailored to the needs of different stakeholders, e.g., an AI welfare allocation system provides citizens with clear explanations of eligibility decisions and appeal processes (Bozkurt & Sharma, 2024). Transparency contributes to trust when stakeholders understand why a recommendation or decision has been generated and what factors influence the outcome (Zerilli et al., 2022). However, transparency alone is not sufficient if explanations are overly technical or merely create the illusion of openness without supporting genuine understanding, e.g., when a bank loan is denied by an AI system without providing the customer with reasons, nor the process for improving or appealing the decision (Rudin, 2019; Xie et al., 2020). UX design ensures multi-layered, visual representations and interactive feedback mechanisms (Bozkurt & Sharma, 2024).

5.3 Inclusion and accessibility

The inclusion and accessibility cluster highlights the importance of ensuring that AI systems serve diverse stakeholders rather than reinforcing social inequalities through underrepresented training data, e.g., using AI systems trained predominantly on data from one demographic group in healthcare (Murphy & Largacha-Martínez, 2022; Panarese et al., 2025). Underrepresentation might also be included in training data based on minority language speakers, differently abled people, rural populations, and historically marginalized communities. For example, AI services through mobile phones and short message service (SMS) are unavailable to stakeholders without broadband internet access, or with indigenous language interfaces that do not enable participation by stakeholders who are typically excluded from mainstream AI systems (Cunningham et al., 2025). Inclusive UX design extends beyond usability and addresses inequities through accessible interface, adaptive interaction models, and culturally sensitive design practices (Jacucci et al., 2014; Nkwo et al., 2025). Hence, by treating stakeholders as co-creators rather than passive users, UX design supports more equitable outcomes and reduces the likelihood of exclusion, ensuring the incorporation of local contexts, values, and knowledge structures into AI systems (Ahrweiler et al., 2025; Stimson & Raper, 2025).

5.4 Embedded ethics

Embedded ethics imply that ethical considerations should be integrated into the design and governance of AI systems rather than applied retrospectively (after deployment), e.g., stakeholders participate in defining acceptable uses of their data before AI development begins (Bryson, 2018; Sanderson et al., 2023). UX designers operationalize principles, such as fairness, accountability, contestability, privacy, and human-centred values, throughout the system lifecycle by translating them into practical design features that stakeholders can experience and evaluate (Papagiannidis et al., 2025). This outcome is achieved through interface design that makes responsibility visible and actionable, identifies risks, and evaluates impacts throughout development (Lindgren et al., 2025; Young et al., 2019). UX design supports human agency and embedded ethics by ensuring that AI systems remain accountable to the stakeholders they serve by balancing automation with human oversight, fairness, and user control (Bryson, 2018; Mittelstadt et al., 2016).

6 Conclusion

The objective of this paper was to understand how UX design could be applied to maintain human agency in AI-driven interactions. The study analyzed 129 papers by applying the PRISMA method. Subsequently, a conceptual framework was derived consisting of four UX design considerations: trust, transparency and explainability, inclusion and accessibility, and embedded ethics, as key responsible AI design dimensions for preserving human agency. These considerations elevate AI system design beyond technological capability towards human empowerment. It emphasizes systems that enable meaningful participation, informed decision-making, and effective human oversight, i.e., human agency.

Methodological decisions such as restricting the review to Scopus-indexed publications and applying predefined search terms, were addressed in the methodology section. However, these limitations also provide opportunities for future research. Future empirical studies should validate the proposed conceptual framework across discrete AI system domains. In addition, future research could investigate the experiences of diverse and underrepresented user

groups to ensure that AI systems support agency inclusively and equitably.

References

Ahrweiler, P., Späth, E., Siqueiros García, J. M., Capellas, B. L., & Wurster, D. (2025). Inclusive Technology Co-design for Participatory AI. In P. Ahrweiler (Ed.), *Participatory Artificial Intelligence in Public Social Services: From Bias to Fairness in Assessing Beneficiaries* (pp. 35-62). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71678-2_2

Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Advanced Research and Reviews*, *18*(3), 050-058. https://doi.org/https://doi.org/10.30574/gscarr.2024.18.3.0088

Ali, M., Khan, T. I., Khattak, M. N., & Şener, İ. (2024). Synergizing AI and business: Maximizing innovation, creativity, decision precision, and operational efficiency in high-tech enterprises. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(3), 100352. https://doi.org/https://doi.org/10.1016/j.joitmc.2024.100352

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2024). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. *International Journal of Human–Computer Interaction*, *40*(5), 1251-1266. https://doi.org/10.1080/10447318.2022.2138826

Baraku, V., Paraskakis, I., Veloudis, S., & Yadav, P. (2025). *Extending Personal Data Sovereignty by Enabling Governance of AI Training on Personal Data* Working Conference on Virtual Enterprises, https://doi.org/10.1007/978-3-032-05673-3_2

Bennett, D., Metatla, O., Roudaut, A., & Mekler, E. D. (2023). *How does HCI understand human agency and autonomy?* Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, https://doi.org/10.1145/3544548.3580651

Bircan, T., & Özbilgin, M. F. (2025). Unmasking inequalities of the code: Disentangling the nexus of AI and inequality. *Technological Forecasting and Social Change*, *211*, 123925. https://doi.org/https://doi.org/10.1016/j.techfore.2024.123925

Bødker, S., Dindler, C., Iversen, O. S., & Smith, R. C. (2022). What is participatory design? In *Participatory design* (pp. 5-13). Springer. https://doi.org/https://doi.org/10.1007/978-3-031-02235-7_2

Bozkurt, A., & Sharma, R. (2024). Trust, Credibility and Transparency in Human-AI Interaction: Why We Need Explainable and Trustworthy AI and Why We Need It Now. *19*, i-ix. https://doi.org/10.5281/zenodo.14599168

Bryson, J. J. (2018). Patiency is not a virtue: the design of intelligent systems and systems of ethics [Article]. *Ethics and Information Technology*, *20*(1), 15-26. https://doi.org/10.1007/s10676-018-9448-6

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., & Rowe, R. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, *19*(43), 1-12. https://doi.org/https://doi.org/10.5334/dsj-2020-043

Cha, I., & Wong, R. Y. (2025). *Understanding Socio-technical Factors Configuring AI Non-Use in UX Work Practices* Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, https://doi.org/10.1145/3706598.3713140

Cunningham, J. L., Adjagbodjou, A., Basoah, J., Jawara, J., Kadoma, K., & Lewis, A. (2025). *Toward Responsible ASR for African American English Speakers: A Scoping Review of Bias and Equity in Speech Technology* [Conference paper]. Proceedings of the 8th AAAI/ACM Conference on AI, Ethics, and Society (AIES 2025), Madrid, Span. https://doi.org/10.1609/aies.v8i1.36580

Doh, W., Goh, Y., & Kim, S. H. (2025). *Beyond Overreliance: The Human-AI-System Concordance (HASC) Matrix and the Cognitive Dynamics of AI-Assisted Decision-Making* [Conference paper]. Proceedings of the Human Factors and Ergonomics Society, https://doi.org/10.1177/10711813251358240

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285-296. https://doi.org/https://doi.org/10.1016/j.jbusres.2021.04.070

Durango, I., Gallud, J. A., & Penichet, V. M. R. (2025). The data dance: choreographing seamless partnerships between humans, data, and GenAI [Article]. *International Journal of Data Science and Analytics*, *20*(4), 3613-3640. https://doi.org/10.1007/s41060-024-00684-w

Dzreke, S. S. (2025). The competitive advantage of AI in business: A strategic imperative. *International Journal for Multidisciplinary Research*, *7*(4), 50400. https://doi.org/https://doi.org/10.36948/ijfmr.2025.v07i04.50400

EU. (2018). European Commission Ethics guidelines for trustworthy AI 2019 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai accessed June 2026.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). *The dark (patterns) side of UX design* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, https://doi.org/10.1145/3173574.3174108

Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. *Perspectives on Public Management and Governance*, *5*(3), 232-242. https://doi.org/https://doi.org/10.1093/ppmgov/gvac008

Hannibal, G., & Lindner, F. (2023). *Towards a Questions-Centered Approach to Explainable Human-Robot Interaction* [Conference paper]. Frontiers in Artificial Intelligence and Applications, https://doi/org/10.3233/FAIA220641

Heaton, D., Clos, J., Nichele, E., & Fischer, J. E. (2023). *The social impact of decision-making algorithms: reviewing*

the influence of agency, responsibility and accountability on trust and blame Proceedings of the First International Symposium on Trustworthy Autonomous Systems, https://doi.org/10.1145/3597512.3599706

Holter, S., & El-Assady, M. (2024). Deconstructing Human-AI Collaboration: Agency, Interaction, and Adaptation [Article]. *Computer Graphics Forum*, *43*(3), Article e15107. https://doi.org/10.1111/cgf.15107

Jacucci, G., Spagnolli, A., Freeman, J., & Gamberini, L. (2014). Symbiotic interaction: A critical definition and comparison to other human-computer paradigms [Article]. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8820*, 3-20. https://doi.org/10.1007/978-3-319-13500-7_1

Kabata, F., & Thaldar, D. (2024). Human in the loop requirement and AI healthcare applications in low-resource settings: A narrative review. *South African Journal of Bioethics and Law*, *17*(2), 70-73. https://doi.org/https://hdl.handle.net/10520/ejc-m_sajbl_v17_n2_a6

Kaminski, M. E., & Malgieri, G. (2024). Impacted stakeholder participation in AI and data governance. *Forthcoming on Yale Journal of Law and Technology (2024-2025), U of Colorado Law Legal Studies Research Paper*(24-23). https://doi.org/https://ssrn.com/abstract=4836460

Kang, E., Choi, S., Li, X., & Hwang, H. (2026). *Exploring Factors Influencing Perceived Usefulness and Adoption Intention for Financial AI Services Across User Experience Levels* [Conference paper]. Communications in Computer and Information Science, https://doi.org/10.1007/978-3-032-12767-9_14

Krakowski, S. (2025). Human-AI agency in the age of generative AI. *Information and Organization*, *35*(1), 100560. https://doi.org/https://doi.org/10.1016/j.infoandorg.2025.100560

Lainjo, B. (2023). The Global Social Dynamics and Inequalities of Artificial Intelligence. *International Journal of Innovation Scientific Research and Review*, *5*(8), 4966-4974. https://doi.org/10.13140/RG.2.2.20078.79689

Li, H., & Tian, F. (2026). Advancing Decision-Making through AI-Human Collaboration: A Systematic Review and Conceptual Framework. *Group Decision and Negotiation*, *35*(2), 26. https://doi.org/10.1007/s10726-026-09980-1

Lindgren, H., Lindvall, K., & Richter-Sundberg, L. (2025). Responsible design of an AI system for health behavior change-an ethics perspective on the participatory design process of the STAR-C digital coach. *Front Digit Health*, *7*, 1436347. https://doi.org/10.3389/fdgth.2025.1436347

Maadi, M., Akbarzadeh Khorshidi, H., & Aickelin, U. (2021). A Review on Human–AI Interaction in Machine Learning and Insights for Medical Applications. *International Journal of Environmental Research and Public Health*, *18*(4), 2121. https://doi.org/https://doi.org/10.3390/ijerph18042121

Masoudi, M. (2025). Algorithmic Governance, Data-Driven Decision Making, and the Transformation of Democratic Accountability in Contemporary States. *Advanced Journal of Management, Humanity and Social Science*, *2*(1), 10-22. https://doi.org/10.5281/zenodo.18009536

Meier, R. (2025). Balancing Minds and Data: The Privacy Dilemma of LLMs and Anthropomorphism in LLMs [Article]. *Journal of Social Computing*, *6*(3), 173-183. https://doi.org/10.23919/JSC.2025.0014

Mishra, P. (2024). Artificial Intelligence in Everyday Life. *Journal of Innovations in Computer Science And Trends in IT*, *1*(2), 3048-4707. https://doi.org/https://bciit.ac.in/pdf/eJournal/Issue%202/3Formatted_PreetiAlok.pdf

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1-21. https://doi.org/10.1177/2053951716679679

Murphy, J. W., & Largacha-Martínez, C. (2022). Decolonization of AI: a Crucial Blind Spot [Article]. *Philosophy and Technology*, *35*(4), Article 102. https://doi.org/10.1007/s13347-022-00588-2

Nagaboyina, P. K. (2024). AI-Driven Big Data Analytics: Transforming Decision-Making and Strategic Insights. *INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS*, *11*, 786-791. https://doi.org/https://www.researchgate.net/publication/395464648_AI-Driven_Big_Data_Analytics_Transforming_Decision-Making_and_Strategic_Insights

Naikar, N., Hoffman, R., Roth, E. M., Klein, G., Militello, L. G., & Dominguez, C. (2026). Should we Make AI More Tool-like or Teammate-Like? *Journal of Cognitive Engineering and Decision Making*, *20*(2), 137-157. https://doi.org/https://doi.org/10.1177/15553434251346904

Nam, S., & Ki Chan, C. C. (2025). *Towards Designing User Interfaces for Optimized Human-AI Communication and Supervisory Control in Software Engineering* [Conference paper]. Proceedings - 29th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2025-Summer, https://doi.org/10.1109/SNPD65828.2025.11252932

Nkwo, M., Ikwunne, T., Adejoro, C., & Anuyah, O. (2025). Exploring Usability of AI Systems in the Global South—Toward Responsible Human-Centered AI for Sustainable Cities and Communities. In *Human - Computer Interaction Series* (Vol. Part F9478, pp. 125-165). https://doi.org/10.1007/978-3-031-88326-2_5

Ntoa, S., Margetis, G., Valakou, A., Makri, F., Dimitriou, N., Karakostas, I., Kokkinis, G., Apostolakis, K. C., Tzovaras, D., & Stephanidis, C. (2024). *A Mixed-Methods Approach for the Evaluation of Situational Awareness and User Experience with Augmented Reality Technologies* [Conference paper]. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), https://doi.org/10.1007/978-3-031-61569-6_13

Olsson, E. (2004). What active users and designers contribute in the design process. *Interacting with Computers*, *16*(2), 377-401. https://doi.org/10.1016/j.intcom.2004.01.001

Panarese, P., Grasso, M. M., & Solinas, C. (2025). Algorithmic bias, fairness, and inclusivity: A multilevel framework for justice-oriented AI. *AI & SOCIETY*, *41*, 1-23. https://doi.org/https://doi.org/10.1007/s00146-025-02451-2

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/https://doi.org/10.1016/j.jsis.2024.101885

Renata, V., Lee, J. D., & Kamaraj, A. V. (2023). *Trading Degrees of Freedom: User Control in Pens, GPT, and Automated Vehicles* [Conference paper]. Proceedings of the Human Factors and Ergonomics Society, https://doi.org/10.1177/21695067231192270

Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Sanderson, C., Douglas, D., Lu, Q., Schleiger, E., Whittle, J., Lacey, J., Newnham, G., Hajkowicz, S., Robinson, C., & Hansen, D. (2023). AI Ethics Principles in Practice: Perspectives of Designers and Developers [Article]. *IEEE Transactions on Technology and Society*, *4*(2), 171-187. https://doi.org/10.1109/TTS.2023.3257303

Singh, A., & Szajnfarber, Z. (2026). Architecting Human-AI Systems for Effective Collaboration and Oversight: Making Sense of Human/AI-in/on/Over/Under/Along-the-Loop. *Systems Engineering*, *29*(2), 337-353. https://doi.org/https://doi.org/10.1002/sys.70024

Smuts, H., & Smith, A. (2021). Collaboration of Human and Machine for Knowledge Work: An Organisational Transformation Framework for Data-driven Decision-making. In Z. W. Y. Lee, T. K. H. Chan, & C. M. K. Cheung (Eds.), *Information Technology in Organisations and Societies: Multidisciplinary Perspectives from AI to Technostress* (pp. 0). Emerald Publishing Limited. https://doi.org/10.1108/978-1-83909-812-320211002

Stimson, C. E., & Raper, R. (2025, 2025//). *Participatory AI: A Method for Integrating Inclusive and Ethical Design Considerations into Autonomous System Development* Towards Autonomous Robotic Systems, Cham. https://doi.org/10.1007/978-3-031-72059-8_13

Subramanian, H. V., Canfield, C., & Shank, D. B. (2024). Designing explainable AI to improve human-AI team performance: A medical stakeholder-driven scoping review [Review]. *Artificial Intelligence in Medicine*, *149*, Article 102780. https://doi.org/10.1016/j.artmed.2024.102780

Taş, E., Memmert, L., & Bittner, E. (2026). Episodic oversight in generative AI workflows: A nine-step protocol for preserving human agency (OP-9). *Electronic Markets*, *36*(1), 58. https://doi.org/https://doi.org/10.1007/s12525-026-00915-x

Thinnakkakath, G. (2026). The Future of Digital Trust: Designing Transparent and Ethical Information Ecosystems. In *Media Literacy and the Politics of Digital Misinformation* (pp. 1-27). https://doi.org/10.4018/979-8-3373-6566-4.ch001

Wang, L., Anyi, C. L., Xu, K., Liu, Y., Arriaga, R. I., & Goel, A. K. (2025). *Explainable AI for Daily Scenarios from End-Users' Perspective: Non-Use, Concerns, and Ideal Design* [Conference paper]. DIS 2025 - Proceedings of the 2025 ACM Designing Interactive Systems Conference, https://doi.org/10.1145/3715336.3735796

Welzel, C., & Inglehart, R. (2010). Agency, values, and well-being: A human development model. *Social indicators research*, *97*(1), 43-63. https://doi.org/https://doi.org/10.1007/s11205-009-9557-z

Xie, Y., Chu, M. T., Li, G., & Li, Y. (2020). Explainable artificial intelligence for designers: A human-centered perspective. *International Journal of Human-Computer Interaction*, *36*(16), 1487-1499. https://doi.org/10.1080/10447318.2020.1801171

Young, M., Katell, M., Krafft, P. M., & Dailey, D. (2019). Practical fairness: Dilemmas in the everyday work of machine learning application development. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-23. https://doi.org/10.1145/3359236

Zerilli, J., Bhatt, U., & Weller, A. (2022). How transparency modulates trust in artificial intelligence. *Patterns (N Y)*, *3*(4), 100455. https://doi.org/10.1016/j.patter.2022.100455