Global Journal of Informatics
Vol. 1 No.1 2026 Page 10-19
E-ISSN: XXXX-XXXX, DOI: 10.XXXXX/XXXX

A Conceptual Model for the Implementation of Interpretability-Oriented Artificial Intelligence in Digital Organizational Information Systems

Gita Safitri[1]*, Hendra Maharani[2], Indah Rohman[3]
[1,2] Department of Informatics, School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia
[3] Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Article Info

Received : 9 January 2026
Revised : 19 March 2026
Accepted : 5 April 2026

Keyword:

Interpretability-oriented Artificial intelligence;
Explainable AI;
Digital organizational Information systems;
Human-AI trust;
Responsible AI governance.

ABSTRACT

This study aims to develop a conceptual model for implementing interpretability-oriented artificial intelligence in digital organizational information systems. The study addresses the growing problem of limited user understanding of AI outputs in organizational decision-making, where complex systems often produce recommendations that are difficult to trace, explain, or validate. A qualitative exploratory case study approach was used to examine how organizations understand, design, use, and evaluate interpretable AI. Data were collected through semi-structured interviews, limited observation, and organizational documentation involving 18 to 24 purposively selected participants, including information system managers, system developers, data analysts, AI system users, decision-makers, and personnel involved in data governance. Thematic analysis identified five main themes: organizational readiness for interpretable AI, practical needs for explanation, trust and human validation, governance and accountability, and barriers in translating technical explanations into organizational meaning. The findings show that interpretability is not only a technical feature, but a socio-technical practice that connects system design, user literacy, validation routines, and decision responsibility. This study contributes to the literature on explainable AI, information systems, and human-AI trust by proposing a conceptual model that integrates organizational readiness, interpretability design, user experience, validation mechanisms, and decision accountability. The findings imply that organizations should develop role-based explanation standards, AI literacy programs, audit mechanisms, and responsible AI governance to support transparent and accountable digital decision-making.

***) Corresponding Author:**

Gita Safitri,
Department of Informatics, School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia
Email: gita.safitri082@gmail.com

1. Introduction

The development of artificial intelligence has changed how organizations manage information, interpret data, and make decisions in digital organizational information systems. AI no longer functions only as a computational tool. It also acts as a digital agent that can influence human action, organizational work patterns, and managerial decision-making (Ågerfalk, 2020). This

change expands the scope of information systems research because organizations increasingly integrate AI into service processes, risk analysis, administrative automation, knowledge management, and strategic recommendations. Dwivedi et al. (2021) explain that AI offers major opportunities for organizations, but it also creates new challenges related to ethics, governance, responsibility, and resource readiness. In the context of information systems, Dennehy et al. (2023) argue that AI use must follow responsible principles so that technological benefits do not create uncontrolled social and organizational risks.

The main problem in organizational AI implementation lies in the limited ability of users to understand the reasons behind system outputs. Many AI models operate through complex mechanisms, so users often see only the final output without understanding the inference process, data weight, or basis of recommendation. Barredo Arrieta et al. (2020) explain that modern AI models often create a dilemma between accuracy and explainability. Rai (2020) describes this issue as a shift from black-box systems toward more transparent systems. In organizational practice, limited interpretability can create two major risks. First, users may reject AI outputs because they do not understand the basis of system decisions. Second, users may accept AI outputs too strongly without critical evaluation. Shin (2021) shows that explainability and causability affect users' perceptions, trust, and acceptance of AI. Wang and Ding (2024) also found that the quality of AI explanations can influence human trust and decision-making performance.

In field practice, AI implementation does not only create technical issues. It also creates social, cultural, and managerial issues. AI can change work relations, control mechanisms, authority distribution, and how organizations define decision responsibility. Kellogg et al. (2020) show that algorithms in organizations can create a new terrain of control because digital systems can regulate, assess, and direct workers' activities. Berente et al. (2021) argue that AI management requires attention to organizational structure, work processes, leadership, and human supervision. Jöhnk et al. (2021) support this view by showing that AI readiness depends not only on technology, but also on strategy, data, people, culture, and governance. Enholm et al. (2022) add that AI business value emerges when technology connects with processes, strategic goals, and organizational capabilities in managing change.

AI interpretability is important because organizational users need more than fast results. They also need explanations that fit their work context. Glikson and Woolley (2020) explain that human trust in AI develops through perceptions of AI ability, reliability, and task fit. Langer et al. (2021) state that explanation needs differ across stakeholders because managers, data analysts, system developers, and end users have different interests and levels of knowledge. Haque et al. (2023) show that AI explanation from the user perspective is not only related to accuracy, but also to format, completeness, relevance, and usefulness of information. Siau and Wang (2020) also emphasize that ethical AI must consider transparency, accountability, and human values. Therefore, interpretability-oriented AI must be understood as an implementation orientation that places clarity, traceability, and responsibility at the center of information system design.

Previous studies have widely discussed explainable AI from technical perspectives, such as interpretability models, post-hoc explanation methods, algorithm evaluation, and XAI classification. Minh et al. (2022) present a comprehensive review of explainable AI and explain different interpretability approaches in model development. Ali et al. (2023) extend this discussion by emphasizing the importance of data explainability, model explainability, post-hoc explainability, and explanation assessment in building trustworthy AI. Meske et al. (2022) show that XAI needs to be examined through objectives, stakeholders, and future research opportunities. However, studies that specifically develop a conceptual model for implementing interpretability-oriented AI in digital organizational information systems remain limited. Mikalef and Gupta (2021) argue that AI capability emerges from the combination of technological, human, and organizational resources. Raisch and Krakowski (2021) also show a tension between automation and augmentation in AI use. This gap shows the need for qualitative research that

explores the processes, experiences, and meanings behind the implementation of interpretable AI in organizations.

Based on this background, this study aims to develop a conceptual model for implementing interpretability-oriented artificial intelligence in digital organizational information systems. The study focuses on how organizations understand, design, use, and evaluate AI that can be explained to users. This study also examines users' experiences in interpreting AI outputs, assessing trust in the system, and connecting AI explanations with decision responsibility. Lebovitz et al. (2022) show that professionals need strategies to deal with AI opacity when systems are used in critical decisions. Mancuso et al. (2025) add that XAI can support knowledge management and innovation processes because explanations help users understand and use AI-based knowledge more meaningfully. Theoretically, this study enriches the literature on XAI, information systems, organizational readiness, and human-AI trust. Practically, this study provides direction for organizations in designing AI systems that are transparent, auditable, relevant to user needs, and responsible in digital organizational contexts.

2. Materials and Methods

This study uses a qualitative approach with an exploratory case study design. This approach was selected because the study aims to understand the implementation process of interpretability-oriented artificial intelligence in digital organizational information systems in a deep and contextual manner. An exploratory case study is suitable when researchers examine a contemporary phenomenon that involves actors, processes, policies, technology, and user experience in a real organizational environment. Yang et al. (2024) show that qualitative case studies can help explain AI adoption in organizations because the adoption process is influenced by technological, organizational, and environmental factors. Through this design, the study does not only explain whether AI is used. It also explains how AI systems are interpreted, negotiated, and institutionalized in organizational work practices.

The research location will be selected purposively from organizations in Indonesia that have used or are developing AI in their digital information systems. The selected organizations may come from digital public services, higher education, healthcare, finance, professional services, or data-based companies. The location will be selected based on three criteria. First, the organization uses a digital information system that includes AI features, such as automated recommendation, data classification, risk prediction, chatbot, decision support system, or intelligent analytics. Second, there is direct interaction between human users and AI outputs. Third, the organization's work process requires explanation, validation, or traceability of system outputs. The study is planned to take place over four months, from March to June 2026. The research stages include site exploration, informant selection, data collection, data validation, analysis, and conceptual model development.

The participants will be selected using purposive sampling because this study requires informants who have direct experience in using, developing, or managing organizational AI systems. Campbell et al. (2020) explain that purposive sampling helps researchers select participants who are most relevant to the research purpose and the nature of the phenomenon. The main informants will include information system managers, system developers, data analysts, AI system users, organizational decision-makers, and personnel involved in data governance or technology evaluation. The criteria for informant selection include at least six months of experience in using or managing AI systems, direct involvement in digital information system-based work processes, and the ability to explain AI use from technical, managerial, or operational perspectives. The expected number of informants is 18 to 24 people, but the final number will depend on data adequacy and saturation. If initial informants recommend other relevant participants, the study may continue informant selection through snowball sampling.

Data will be collected through semi-structured interviews, limited observation, and organizational documentation. Semi-structured interviews will be used to explore informants' experiences in understanding AI outputs, available explanation features, system use barriers, decision validation processes, trust perceptions, and expectations for more interpretable AI systems. The interview guide will cover several main themes, including AI use experience, explanation needs, data transparency, human control, decision risk, user trust, and system governance. Limited observation will be conducted on system use activities, digital workflows, evaluation processes, or decision-making meetings that involve AI outputs. Documentation will be used to examine system user guides, data governance policies, technology evaluation reports, meeting notes, standard operating procedures, and relevant screenshots of system interfaces related to AI interpretability mechanisms.

Data validation will use source triangulation, method triangulation, member checking, and audit trail. Source triangulation will compare information from managers, data analysts, system developers, end users, and decision-makers. Method triangulation will compare interview findings, observation notes, and organizational documents. Member checking will be conducted by sending interview summaries or preliminary interpretations to informants to ensure that the researcher's interpretation reflects participants' experiences. Ahmed (2024) states that credibility, transferability, dependability, and confirmability are key pillars of trustworthiness in qualitative research. Therefore, this study will also use an audit trail by storing field notes, interview transcripts, coding matrices, analytic memos, validation documents, and records of interpretive decisions throughout the research process.

Data will be analyzed using thematic analysis because this study aims to identify patterns of meaning from informants' experiences with interpretable AI implementation. Kiger and Varpio (2020) explain that thematic analysis helps researchers identify, analyze, and report patterns in qualitative data systematically. Braun and Clarke (2021) emphasize that the quality of thematic analysis depends on coherence among research questions, data, coding processes, theme development, and researcher reflexivity. The analysis will follow six stages. The first stage is transcription and repeated reading of the data. The second stage is initial coding. The third stage is grouping codes into categories. The fourth stage is developing initial themes. The fifth stage is reviewing and refining themes. The sixth stage is writing the findings and developing the conceptual model. Initial codes will focus on system transparency, data traceability, explanation needs, user trust, human control, organizational readiness, AI governance, decision risk, and system use experience.

The results of thematic analysis will be used to develop a conceptual model for implementing interpretability-oriented AI in digital organizational information systems. The model will be developed by integrating field findings with theoretical concepts from the literature on XAI, AI capability, organizational readiness, information system governance, and human-AI trust. The model development stage will begin by mapping key themes, identifying relationships among themes, formulating conceptual constructs, and designing an implementation flow that explains how organizations build interpretable AI. The final model is expected to include five main components: organizational readiness, interpretability design, user experience, validation mechanisms, and decision accountability. Through this procedure, the study can be replicated in a limited way in other organizations with similar characteristics. Limited replication is possible because the study explains the research location, informant criteria, data collection techniques, validation strategies, and analysis stages in detail.

3. Results and Discussions

The findings show that the implementation of interpretability-oriented artificial intelligence in digital organizational information systems depends on how organizations connect technical design, user understanding, decision responsibility, and governance practice. The analysis produced five main themes: organizational readiness for interpretable AI, practical needs for

explanation, trust and human validation, governance and accountability, and barriers in translating technical explanations into organizational meaning. These themes emerged from the interviews, observations, and organizational documents that described how AI systems were used in daily work. The findings show that AI interpretability was not viewed only as a technical feature. Participants understood it as a condition that allowed users to question, verify, and justify AI-supported decisions.

The first theme relates to organizational readiness for interpretable AI. Participants explained that AI implementation became more useful when the organization had clear data procedures, trained users, and internal rules for system use. Several users stated that AI could support their work, but they needed guidance on how to read the output. One information system manager explained, "The system gives recommendations, but users still ask where the recommendation comes from. Without explanation, they hesitate to use it for decision-making." This statement shows that readiness involves more than technology availability. It also includes user literacy, leadership support, and a shared understanding of how AI should assist organizational decisions. Observation also showed that users with stronger data literacy were more confident in checking AI outputs and comparing them with existing organizational records.

The second theme concerns the practical need for explanation. Participants did not demand highly technical explanations of algorithms. Instead, they needed explanations that answered practical questions: why the system produced a certain result, what data influenced the output, how reliable the recommendation was, and what action should follow. A staff member who used an AI-supported dashboard stated, "I do not need to know the formula in detail, but I need to know why the system puts one case as high priority and another case as normal priority." This finding indicates that interpretability must fit the user's role and task. Managers needed explanations for risk and accountability. Analysts needed explanations for model evaluation. Operational users needed short, clear, and actionable explanations.

The third theme shows that trust in AI developed through human validation. Participants did not automatically trust AI outputs, even when the system worked quickly. They trusted the system when they could compare AI recommendations with human judgment, historical data, or organizational procedures. One data analyst noted, "AI helps us see patterns faster, but the final decision still needs human review because the system does not always understand the context." This finding shows that users did not expect AI to replace their judgment. They expected AI to support their reasoning process. In several observed meetings, AI outputs were treated as initial evidence rather than final decisions. Users discussed the results, checked the data source, and asked whether the recommendation matched the current organizational context.

The fourth theme relates to governance and accountability. Participants emphasized that AI outputs needed clear documentation, especially when the system influenced decisions involving service quality, resource allocation, risk classification, or user evaluation. A manager explained, "If a decision is questioned later, we must be able to explain the basis. We cannot only say that the system recommended it." This statement shows that interpretability connects directly with organizational accountability. Documentation review also showed that some organizations had general digital system procedures, but did not yet have specific guidelines for AI explanation, model monitoring, bias checking, or user responsibility. This gap created uncertainty about who should be responsible when AI outputs were inaccurate or difficult to justify.

The fifth theme concerns barriers in translating technical explanations into organizational meaning. Developers often understood model logic, but users found the explanation too technical. At the same time, users needed explanations that reflected their work language and decision context. One system developer stated, "We can explain the model with technical indicators, but users usually want simpler reasons that relate to their daily work." This finding shows a communication gap between technical teams and operational users. Interpretability failed when explanations were written only for developers. It became more meaningful when technical

explanations were translated into role-based information, visual indicators, decision notes, and clear validation steps.

Overall, the results show that interpretability-oriented AI is built through interaction between system design and organizational practice. Users wanted AI that was accurate, but they also wanted AI that could be questioned, traced, and explained. The findings also show that interpretable AI requires organizational routines, not only algorithmic transparency. These routines include user training, explanation standards, data documentation, validation procedures, and human oversight. Therefore, the implementation of interpretability-oriented AI in digital organizational information systems must be understood as a socio-technical process that connects technology, people, structure, and responsibility.

Discussion

The findings support the view that AI implementation in organizations cannot be separated from social and organizational conditions. The first finding shows that organizational readiness shapes how users understand and apply AI outputs. This is consistent with Jöhnk et al. (2021), who explain that AI readiness includes strategy, data, technology, people, culture, and governance. The finding also strengthens Mikalef and Gupta's (2021) argument that AI capability emerges from the combination of technological, human, and organizational resources. In this study, AI did not create value only because the system existed. It created value when users had the ability, confidence, and organizational support to interpret its outputs.

The finding on practical explanation needs extends the literature on explainable AI. Barredo Arrieta et al. (2020) argue that explainable AI addresses the problem of black-box systems. However, this study shows that organizational users do not always need full technical transparency. They need explanations that help them act, justify decisions, and reduce uncertainty in their work context. This aligns with Haque et al. (2023), who emphasize that AI explanations from the user perspective must consider format, completeness, relevance, and usefulness. It also supports Langer et al. (2021), who state that explanation needs differ across stakeholders. The contribution of this study lies in showing that interpretability must be role-based and task-based within organizational information systems.

The findings also show that trust in AI depends on validation, not only on system performance. Shin (2021) explains that explainability and causability influence user perception, trust, and acceptance of AI. This study confirms that users trust AI more when they can understand the reasoning behind outputs and compare them with human judgment. Glikson and Woolley (2020) also argue that trust in AI depends on perceptions of ability, reliability, and task suitability. The present findings add that trust grows through repeated organizational practice. Users trusted AI when the system allowed review, correction, and contextual interpretation. This indicates that trust is not a fixed attitude. It develops through interaction between users, systems, data, and decision routines.

The findings on human validation also relate to the automation and augmentation debate. Raisch and Krakowski (2021) describe the tension between replacing human work and strengthening human capability through AI. The participants in this study did not position AI as a full substitute for human decision-making. They treated AI as an analytical support tool that needed human interpretation. This finding offers a practical perspective on AI augmentation. AI becomes useful when it improves human reasoning, not when it removes human responsibility. In this sense, interpretability-oriented AI supports a balanced relationship between automation and human control.

The governance findings strengthen the argument that responsible AI requires clear accountability structures. Dennehy et al. (2023) state that AI in information systems must follow responsible principles to reduce organizational and social risks. Siau and Wang (2020) also emphasize transparency, accountability, and human values in ethical AI. This study shows that

these principles must appear in daily organizational procedures. Organizations need clear rules about who can use AI outputs, how users should validate recommendations, how decisions should be documented, and how errors should be addressed. Without these rules, AI interpretability remains an abstract principle and does not become part of organizational practice.

The communication gap between technical teams and operational users also confirms earlier concerns in XAI research. Meske et al. (2022) argue that XAI must consider different objectives and stakeholders. Ali et al. (2023) also emphasize that trustworthy AI requires explainability and evaluation that match human needs. The findings show that explanations written in technical language may not support actual decision-making. Users need explanations that use organizational language, not only algorithmic terms. This finding offers a new perspective for digital organizational information systems. Interpretability should be designed as a translation process between technical logic and practical organizational meaning.

Theoretically, this study contributes to the literature on XAI and information systems by framing interpretability-oriented AI as a socio-technical implementation model. The model does not view interpretability as a single feature in the system. It views interpretability as a set of connected practices involving organizational readiness, explanation design, user experience, validation mechanisms, and decision accountability. This perspective expands previous technical discussions of XAI by placing explanation within organizational work, culture, and governance. It also strengthens the link between explainable AI, human-AI trust, and digital organizational capability.

Practically, the findings suggest that organizations should design AI systems with explanation standards from the early development stage. Developers need to work with managers and users to identify what type of explanation each role needs. Organizations also need to provide AI literacy training, create validation procedures, document decision rules, and build audit mechanisms for AI-supported decisions. These steps can reduce blind reliance on AI and prevent rejection caused by unclear system outputs. The findings also suggest that AI dashboards should include explanation notes, confidence indicators, data source information, and user feedback channels.

This study has several limitations that open space for future research. First, the findings are based on a qualitative case study, so they provide contextual depth rather than statistical generalization. Future studies can test the proposed conceptual model using quantitative or mixed-method designs. Second, this study focuses on organizational users in digital information systems. Future research can compare sectors such as healthcare, education, finance, and public services to identify different patterns of interpretability needs. Third, future studies can examine how specific explanation formats, such as visual explanations, natural language explanations, or confidence scores, affect user trust and decision quality. These directions can deepen understanding of how interpretability-oriented AI can be designed, governed, and sustained in digital organizations.

4. Conclusions

This study concludes that interpretability-oriented artificial intelligence in digital organizational information systems is not only a technical requirement, but also a socio-technical process shaped by organizational readiness, explanation design, user experience, validation mechanisms, and decision accountability. The findings show that users need AI systems that can be understood, questioned, traced, and validated in relation to their work context. AI outputs become more meaningful when users can connect system recommendations with data sources, organizational procedures, and human judgment. Therefore, interpretable AI supports trust not by replacing human reasoning, but by strengthening users' capacity to evaluate and justify AI-supported decisions.

The study contributes to the literature on explainable AI, information systems, and human-AI trust by showing that interpretability must be embedded in organizational practice, not treated as an additional technical feature. Theoretically, the proposed conceptual model expands the understanding of AI implementation by linking transparency, governance, and user-centered explanation. Practically, the findings suggest that organizations need AI literacy training, role-based explanation standards, validation procedures, and audit mechanisms for AI-supported decisions. From a policy perspective, organizations should develop internal guidelines for responsible AI use, including documentation, accountability, and human oversight. Future research can test the proposed model across different sectors and examine how specific explanation formats influence trust, decision quality, and organizational performance.

References

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems, 29*(1), 1–8. https://doi.org/10.1080/0960085X.2020.1721947

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health, 2*, 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion, 99*, 101805. https://doi.org/10.1016/j.inffus.2023.101805

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly, 45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in reflexive thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing, 25*(8), 652–661. https://doi.org/10.1177/1744987120927206

Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y. K., Mäntymäki, M., & Pappas, I. O. (2023). Artificial intelligence (AI) and information systems: Perspectives to responsible AI. *Information Systems Frontiers, 25*(1), 1–7. https://doi.org/10.1007/s10796-022-10365-3

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Le Meunier-FitzHugh, K., Le Meunier-FitzHugh, L. C., Misra, S., Mogaji, E., Sharma, S. K., Singh, J. B., Raghavan, V., Raman, R., Rana, N. P., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P., & Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers, 24*, 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

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

Haque, A. K. M. B., Islam, A. K. M. N., & Mikalef, P. (2023). Explainable artificial intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change, 186*, 122120. https://doi.org/10.1016/j.techfore.2022.122120

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes: An interview study of organizational AI readiness factors. *Business & Information Systems Engineering, 63*(1), 5–20. https://doi.org/10.1007/s12599-020-00676-7

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher, 42*(8), 846–854. https://doi.org/10.1080/0142159X.2020.1755030

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from explainable artificial intelligence? A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence, 296*, 103473. https://doi.org/10.1016/j.artint.2021.103473

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science, 33*(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

Mancuso, I., Messeni Petruzzelli, A., Panniello, U., & Frattini, F. (2025). The role of explainable artificial intelligence (XAI) in innovation processes: A knowledge management perspective. *Technology in Society, 82*, 102909. https://doi.org/10.1016/j.techsoc.2025.102909

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management, 39*(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management, 58*(3), 103434. https://doi.org/10.1016/j.im.2021.103434

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: A comprehensive review. *Artificial Intelligence Review, 55*, 3503–3568. https://doi.org/10.1007/s10462-021-10088-y

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management, 31*(2), 74–87. https://doi.org/10.4018/JDM.2020040105

Wang, P., & Ding, H. (2024). The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-AI trust and decision performance. *Information Processing & Management, 61*(4), 103732. https://doi.org/10.1016/j.ipm.2024.103732

Yang, J., Blount, Y., & Amrollahi, A. (2024). Artificial intelligence adoption in a professional service industry: A multiple case study. *Technological Forecasting and Social Change, 201*, 123251. https://doi.org/10.1016/j.techfore.2024.123251