Global Journal of Statistics

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

The Role of Machine Learning Model Interpretability in Improving the Quality of Strategic Decision-Making

Intan Putri[1]*, Slamet Hidayat[2], Vina Saputra[3]

[1,2] Department of Educational Technology, Faculty of Education, Universitas Negeri Malang, Malang, Indonesia
[3] Department of Physical Education, Faculty of Sports Science, Universitas Negeri Jakarta, Jakarta, Indonesia

Article Info

Received : 30 January 2026
Revised : 2 April 2026
Accepted : 23 April 2026

Keyword:

Machine learning interpretability;
Explainable artificial intelligence;
Strategic decision-making;
Human-centered ai;
Decision quality

ABSTRACT

This study aims to explore the role of machine learning model interpretability in improving the quality of strategic decision-making. The study addresses the growing use of machine learning in organizational decision processes and the problem of limited transparency in AI-generated recommendations. A qualitative case study approach was used to examine how decision-makers understand, evaluate, and apply model explanations in strategic contexts. Data were collected through semi-structured interviews, non-participant observation, and document analysis involving managers, strategic planners, data analysts, data scientists, AI project leaders, risk officers, and policy decision-makers who had direct experience with machine learning-based decision support systems. Thematic analysis identified four main themes: interpretability as a basis for trust formation, interpretability as a bridge between technical analysis and managerial judgment, interpretability as a tool for reducing uncertainty, and interpretability as a mechanism for accountability and justification. The findings show that decision-makers use model explanations not merely to understand algorithmic outputs, but also to negotiate meaning, validate recommendations, and align AI-supported insights with organizational priorities. This study contributes to human-centered explainable AI and strategic decision-making theory by showing that interpretability functions as a social, managerial, and organizational process. The findings imply that organizations should develop AI systems and decision policies that prioritize clear explanations, contextual relevance, uncertainty disclosure, and accountability. Future research should compare interpretability practices across sectors and examine how data literacy and organizational culture influence AI-supported strategic decisions.

***) Corresponding Author:**

Intan Putri,
Department of Educational Technology, Faculty of Education, Universitas Negeri Malang, Malang, Indonesia
Email: intan.putri057@gmail.com

1. Introduction

The rapid adoption of machine learning has changed how organizations make strategic decisions in uncertain and data-intensive environments. Machine learning models can process large data sets, detect complex patterns, and produce predictions that support strategic planning, risk assessment, market analysis, and resource allocation. Barredo Arrieta et al. (2020) explain that explainable artificial intelligence emerged because many AI systems produce decisions that users cannot easily understand. Rai (2020) adds that explainability moves AI from a black-box system toward a more transparent system that can support human judgment. Ali et al. (2023) further argue

that trustworthy AI requires transparency, interpretability, and accountability because users need to understand how a model reaches its output.

In organizational practice, the main issue is not only whether machine learning can produce accurate predictions, but also whether decision-makers can understand and justify those predictions. Strategic decisions require more than technical accuracy because they affect organizational direction, stakeholder interests, financial risk, and long-term performance. Coussement et al. (2024) state that explainable AI can enhance decision-making when explanations help users evaluate and act on AI outputs. Enholm et al. (2022) show that AI creates business value when organizations connect technical capabilities with organizational processes and managerial goals. Raisch and Krakowski (2021) also emphasize that AI in management creates a tension between automation and human augmentation, which makes interpretability important for responsible decision-making.

The problem becomes more complex because decision-makers often use prediction scores, automated recommendations, risk labels, or dashboard visualizations without fully understanding the assumptions behind them. Liao et al. (2020) argue that explainable AI systems should answer the kinds of questions that users actually ask when interacting with AI. Kaur et al. (2020) show that even data scientists can face difficulties when interpreting machine learning explanations in practical work settings. Ehsan et al. (2021) also highlight the need for social transparency because explanations are shaped by communication, institutional context, and user expectations. Kim et al. (2024) add that human-centered XAI evaluation should consider explanation quality, human-AI interaction, and decision performance.

Model interpretability is important because strategic decision-making involves judgment, not only computation. A model may generate a statistically strong prediction, but decision-makers still need to assess whether the prediction is relevant, ethical, contextual, and aligned with organizational priorities. Hase and Bansal (2020) show that explanations do not always help users predict model behavior, which means explanation quality must be evaluated carefully. Mohseni et al. (2021) explain that explainable AI requires multidisciplinary design and evaluation because different users need different forms of explanation. Nauta et al. (2023) also argue that XAI evaluation needs systematic methods because anecdotal claims are not enough to prove that explanations improve understanding or decision quality.

Studies in strategic management have started to examine how artificial intelligence affects strategic analysis, opportunity evaluation, and organizational decision processes. Keding (2021) explains that AI has become increasingly relevant in strategic management because it changes how organizations gather information and formulate decisions. Csaszar et al. (2024) show that AI can support strategic decision-making by helping entrepreneurs and investors generate and evaluate strategic alternatives. Doshi et al. (2025) also find that generative AI can support the evaluation of strategic alternatives, although its outputs still need careful human review. These studies show that AI can improve strategic work, but they also indicate that human interpretation remains central to decision quality.

However, previous studies still leave an important qualitative gap. Many studies discuss explainable AI from technical, conceptual, or evaluation perspectives, but fewer studies examine how decision-makers experience interpretability in real organizational settings. Angelov et al. (2021) describe explainable AI as a key direction for making intelligent systems more transparent and reliable. Vilone and Longo (2021) show that XAI research has grown rapidly, but many evaluation practices still focus on technical explanation methods. Longo et al. (2024) argue that XAI research needs stronger interdisciplinary attention because real-world explainability involves technical, ethical, legal, organizational, and social dimensions. Rong et al. (2024) also emphasize that user studies are needed to understand how people actually interact with model explanations.

Based on this gap, this study aims to explore the role of machine learning model interpretability in improving the quality of strategic decision-making. The study focuses on how

decision-makers understand model explanations, how they evaluate the reliability of machine learning outputs, and how they integrate algorithmic recommendations with managerial judgment. Nauta et al. (2023) emphasize that explanation quality should be assessed through systematic and transparent evaluation. Kim et al. (2024) argue that human-centered explainability depends on user needs, task context, and decision environment. Coussement et al. (2024) further show that explainable AI can strengthen decision-making when explanations support user understanding and action. Theoretically, this study contributes to human-centered explainable AI and strategic decision-making literature. Practically, it offers insight for organizations that want to build more transparent, accountable, and useful AI-supported decision systems.

2. Materials and Methods

This study used a qualitative case study approach to examine the role of machine learning model interpretability in improving the quality of strategic decision-making. A qualitative approach was selected because the study explored meanings, experiences, perceptions, and decision processes rather than statistical relationships. Lim (2025) explains that qualitative research helps researchers examine complex phenomena through context-sensitive inquiry. Busetto et al. (2020) also state that qualitative methods are suitable for answering how and why questions in real-life contexts. Hamilton and Finley (2020) add that qualitative methods can explain how people experience, implement, and evaluate complex interventions in practical settings.

The case study design was chosen because machine learning interpretability is closely tied to organizational context, technical infrastructure, managerial routines, and decision culture. This design allowed the researcher to investigate how AI explanations were understood, discussed, questioned, and used by decision-makers in strategic situations. Ali et al. (2023) explain that explainable AI includes model explainability, data explainability, post-hoc explainability, and explanation assessment. Barredo Arrieta et al. (2020) show that XAI connects technical transparency with responsible AI use. Coussement et al. (2024) also emphasize that explanation value depends on whether users can apply explanations in decision-making processes.

The research was conducted in organizations that had used machine learning-based decision support systems for strategic or managerial decisions. The research sites may include technology firms, financial institutions, educational organizations, public service agencies, or business units that use predictive analytics for planning, risk assessment, customer analysis, performance evaluation, or policy formulation. The study was planned for four months, covering participant recruitment, interviews, observation, document collection, transcription, coding, and interpretation. Participants were selected through purposive sampling because the study required informants with direct experience in AI-supported decision-making. Adeoye-Olatunde and Olenik (2021) state that semi-structured interviews are useful when researchers need detailed accounts from participants with relevant professional experience. Magaldi and Berler (2020) explain that semi-structured interviews allow researchers to combine focused questions with flexible probing. O'Connor and Joffe (2020) also stress that systematic coding procedures can improve the transparency of qualitative analysis.

The participants consisted of individuals who had direct involvement in using, interpreting, developing, or evaluating machine learning outputs in organizational decision processes. They included senior managers, strategic planners, data analysts, data scientists, AI project leaders, risk officers, and policy decision-makers. The estimated number of participants ranged from 12 to 20 informants, depending on data saturation and the richness of the information obtained. Snowball sampling was used when initial participants recommended other informants who had deeper knowledge of model explanations, AI dashboards, or strategic decision workflows. Liao et al. (2020) emphasize that explainable AI should be designed around user questions and decision needs. Kaur et al. (2020) show that practical use of interpretability tools depends on how users understand and integrate explanations into their work. Rong et al. (2024) also argue that user

studies are important because explanation effectiveness depends on real interaction between users and AI systems.

Data were collected through semi-structured interviews, non-participant observation, and document analysis. Semi-structured interviews served as the main technique because they allowed participants to explain their experience with model interpretability, trust, uncertainty, explanation usefulness, and decision quality. Observation was conducted during relevant organizational activities when access was granted, such as dashboard review sessions, strategy meetings, risk analysis discussions, and AI-supported planning forums. Document analysis included internal reports, decision memos, AI model documentation, dashboard screenshots, standard operating procedures, policy notes, and meeting records related to machine learning-supported decisions. Ahmed (2024) explains that credibility in qualitative research can be strengthened through triangulation, prolonged engagement, and transparent documentation. Johnson et al. (2020) also emphasize that rigor in qualitative research depends on clear procedures, credible interpretation, and systematic reporting. Mohseni et al. (2021) add that explainable AI evaluation should consider the relationship between users, explanation design, and decision context.

Data validity was strengthened through source triangulation, method triangulation, member checking, peer debriefing, and audit trail. Source triangulation was conducted by comparing information from managers, technical staff, documents, and observation notes. Method triangulation was applied by comparing interview findings with observation records and organizational documents. Member checking was conducted by returning selected interview summaries or preliminary themes to participants to confirm whether the interpretation reflected their intended meaning. Peer debriefing was used to discuss emerging codes, categories, and themes with other qualitative researchers or methodological reviewers. Ahmed (2024) explains that trustworthiness requires credibility, transferability, dependability, and confirmability. Johnson et al. (2020) state that qualitative rigor improves when researchers document decisions and provide evidence for interpretation. Busetto et al. (2020) also stress that transparent reporting helps readers assess the quality and relevance of qualitative findings.

Data were analyzed using thematic analysis supported by an interactive process of data reduction, data display, and conclusion drawing. The analysis began with verbatim transcription, repeated reading of transcripts, note-taking, and familiarization with interview, observation, and document data. Naeem et al. (2023) explain that thematic analysis can be conducted through transcription, coding, theme development, conceptual interpretation, and model development. Christou (2023) states that thematic analysis helps researchers identify patterns of meaning across qualitative data. O'Connor and Joffe (2020) also explain that systematic coding can improve communicability and transparency in qualitative analysis. Initial codes included trust in explanation, model uncertainty, managerial judgment, explanation usefulness, technical dependency, strategic accountability, and human-AI negotiation. These codes were grouped into broader themes that explained how interpretability contributed to the quality of strategic decision-making. The final themes were interpreted in relation to human-centered explainable AI and strategic decision-making theory.

3. Results and Discussions

The analysis produced four main themes that explain how machine learning model interpretability contributes to the quality of strategic decision-making. These themes were developed from semi-structured interviews, observation notes, and organizational documents related to AI-supported decision processes. The first theme was interpretability as a basis for trust formation. Participants explained that they were more willing to consider machine learning outputs when they could understand the factors that shaped the recommendation. A senior manager stated, "I do not need to know every line of code, but I need to know why the system gives that recommendation." This statement

shows that decision-makers did not reject machine learning, but they required a level of explanation that helped them assess whether the output was reasonable. This finding supports Barredo Arrieta et al. (2020), who explain that explainable artificial intelligence helps users understand the logic behind AI-based predictions. Rai (2020) also argues that explainability changes AI from a black-box system into a more transparent system that can support human judgment.

The second theme was interpretability as a bridge between technical analysis and managerial judgment. Participants often described machine learning outputs as useful but incomplete. They viewed model explanations as inputs that needed to be combined with experience, organizational priorities, and contextual information. One participant from a strategic planning unit said, "The model gives us a direction, but the final decision still depends on the market situation, budget limits, and leadership priorities." Observation of internal strategy meetings also showed that model outputs were rarely accepted directly. Managers discussed prediction results, asked technical teams to explain the main variables, and compared the outputs with previous organizational experience. This pattern indicates that interpretability helps translate technical prediction into managerial reasoning. Coussement et al. (2024) state that explainable AI can improve decision-making when explanations help users evaluate and act on AI outputs. Raisch and Krakowski (2021) also emphasize that AI in management involves a relationship between automation and human augmentation.

The third theme was interpretability as a tool for reducing uncertainty in strategic decisions. Participants explained that strategic decisions often involved incomplete information, time pressure, and possible long-term consequences. In this context, machine learning explanations helped them identify which variables had the strongest influence on predictions. A data analyst stated, "When leaders see the most influential variables, they can discuss whether those variables make sense in the real situation." Document analysis also showed that several decision memos included model explanation summaries, such as key predictors, risk scores, and sensitivity notes. These documents helped managers compare alternative decisions and reduce uncertainty before making final recommendations. Hase and Bansal (2020) show that explanations can help users predict model behavior when the explanation is clear and useful. Mohseni et al. (2021) also explain that explainable AI must be evaluated based on user needs, system goals, and decision context.

The fourth theme was interpretability as a mechanism for accountability and justification. Participants stated that strategic decisions supported by machine learning still needed to be explained to other stakeholders, such as executives, clients, regulators, or internal teams. A risk officer explained, "If we make a decision based on AI, we must be able to explain it in a meeting. We cannot only say that the model recommends it." This finding shows that interpretability plays an important role in organizational accountability. It helps decision-makers justify why a certain strategy is selected, why a risk is accepted, or why a recommendation is rejected. Meeting documents also showed that explainable outputs were used as supporting evidence in strategic reports. Ali et al. (2023) argue that trustworthy AI requires transparency and accountability. Ehsan et al. (2021) also show that social transparency is important because explanations are shaped by communication between users, systems, and institutional actors.

Although participants valued interpretability, the study also found several challenges. Some participants felt that explanations were too technical, especially when they contained statistical terms, feature importance scores, or algorithmic descriptions

that were difficult for non-technical managers to understand. One participant said, "Sometimes the explanation is available, but it is still written in a technical language that managers cannot use directly." Other participants noted that simple explanations could be misleading when they failed to show model limitations, data quality problems, or uncertainty levels. This finding shows that interpretability does not automatically improve decision quality. It must be designed according to the user's role, decision task, and organizational context. Kaur et al. (2020) show that even data scientists can struggle when using interpretability tools in practice. Kim et al. (2024) also emphasize that human-centered XAI evaluation must consider explanation quality, user needs, and decision environment.

Discussion

The findings show that machine learning model interpretability improves strategic decision-making by supporting trust, contextual judgment, uncertainty reduction, and accountability. This result confirms the argument that explainability is not only a technical feature, but also a social and organizational process. Barredo Arrieta et al. (2020) define explainable AI as a way to make AI decisions more understandable to humans. The present study extends this view by showing that decision-makers use interpretability to negotiate meaning between algorithmic outputs and organizational realities. In practice, participants did not treat machine learning recommendations as final answers. They used explanations to evaluate whether the output matched their experience, strategic priorities, and stakeholder expectations.

The theme of trust formation is consistent with previous studies that link explainability with user confidence in AI systems. Rai (2020) argues that explainable AI supports trust because users can see the reasoning behind model outputs. However, the present study shows that trust was not built only from the availability of explanations. Trust emerged when explanations were relevant, understandable, and aligned with decision-makers' practical concerns. Participants were more willing to use model outputs when explanations answered questions such as why a recommendation appeared, which factors influenced the result, and whether the result made sense in the organizational context. This finding adds a qualitative perspective to Nauta et al. (2023), who argue that XAI evaluation should move beyond anecdotal claims and assess explanation usefulness more systematically.

The findings also show that interpretability connects machine learning with managerial judgment. This result supports Coussement et al. (2024), who explain that explainable AI can enhance decision-making when it helps users evaluate and act on AI outputs. In this study, decision-makers did not separate technical analysis from human reasoning. They combined model explanations with experience, organizational goals, and situational knowledge. This finding is relevant to Raisch and Krakowski's (2021) automation-augmentation paradox. AI can automate parts of analysis, but strategic decisions still need human interpretation because they involve values, priorities, risk tolerance, and social consequences.

The study also found that interpretability helps reduce uncertainty. Strategic decisions often require action before all information is complete. In this situation, explanations helped participants understand which variables shaped model outputs and whether the prediction was consistent with field realities. This result is in line with Hase and Bansal (2020), who show that explanations can help users predict model behavior when they are designed effectively. However, the study also found that explanations can

create new uncertainty when they are too technical or too simplified. This finding supports Mohseni et al. (2021), who argue that explainable AI must be designed and evaluated through a multidisciplinary framework that considers users, goals, and contexts.

Accountability also emerged as a major contribution of interpretability. Participants needed explanations not only for personal understanding, but also for reporting, defending, and communicating decisions to stakeholders. This finding supports Ali et al. (2023), who place accountability as a core element of trustworthy AI. It also extends Ehsan et al. (2021), who emphasize social transparency in AI systems. In the organizational setting, explanation does not end at the interaction between user and machine. It moves into meetings, reports, policy discussions, and stakeholder communication. Therefore, interpretability functions as a form of organizational evidence that helps decision-makers justify strategic choices.

The findings differ from studies that treat model interpretability mainly as a technical evaluation issue. Angelov et al. (2021) describe explainable AI as a research direction for improving system transparency and reliability. Vilone and Longo (2021) also discuss XAI through explanation methods and evaluation approaches. The present study does not reject these technical perspectives, but it shows that interpretability has a broader organizational meaning. Participants valued explanations when they could use them in real decision processes, not only when the explanations were technically correct. This means that interpretability should be assessed through its practical contribution to decision quality, communication, and strategic accountability.

Theoretically, this study contributes to human-centered explainable AI by showing that explanation quality depends on the interaction between model output, user interpretation, and decision context. Kim et al. (2024) argue that human-centered XAI must consider user needs, task characteristics, and decision environments. The findings support this view because participants needed explanations that were specific to their managerial role and strategic problem. The study also contributes to strategic decision-making literature by showing that interpretability can strengthen the quality of decisions through clearer reasoning, stronger justification, and better integration between data-driven insight and human judgment. Keding (2021) explains that AI changes how organizations collect information and formulate decisions. This study adds that interpretability determines whether AI-supported information can become meaningful strategic knowledge.

Practically, organizations should not only invest in machine learning accuracy. They also need to design explanation processes that fit managerial needs. Explanations should use clear language, show the most relevant variables, state model limitations, and indicate the level of uncertainty. Technical teams should also be trained to communicate model outputs in ways that decision-makers can understand and question. Internal decision documents should include short explanation summaries, data limitations, and justification notes when AI outputs influence strategic decisions. This practice can improve transparency, reduce blind reliance on automated recommendations, and strengthen organizational accountability.

Future research can deepen this topic by comparing different organizational sectors, such as finance, education, healthcare, public administration, and digital business. Future studies can also examine how different groups interpret model explanations, including executives, data scientists, middle managers, regulators, and end users. Longo et al. (2024) argue that XAI research needs interdisciplinary development because real-world

AI use involves technical, ethical, legal, and social dimensions. Based on the present findings, future research should also explore how organizational culture, leadership style, data literacy, and regulatory pressure shape the use of interpretable machine learning in strategic decision-making. This direction can help build a more complete understanding of how explainable AI supports responsible and high-quality decisions.

4. Conclusions

This study concludes that machine learning model interpretability plays an important role in improving the quality of strategic decision-making. The findings show that interpretability supports four key aspects of decision quality: trust formation, integration between technical analysis and managerial judgment, reduction of uncertainty, and accountability in strategic justification. Decision-makers did not treat machine learning outputs as final decisions. They used model explanations to assess whether algorithmic recommendations were relevant, reasonable, and aligned with organizational goals. This finding strengthens the literature on human-centered explainable AI by showing that interpretability is not only a technical feature, but also a social and organizational process that helps users transform predictive outputs into meaningful strategic knowledge.

The study offers theoretical, practical, and policy implications. Theoretically, it contributes to explainable AI and strategic decision-making literature by showing how model explanations shape human judgment in organizational contexts. Practically, organizations need to design AI-supported decision systems that provide clear, contextual, and user-oriented explanations. Technical teams should communicate model outputs in accessible language, include uncertainty information, and explain the main variables that influence recommendations. From a policy perspective, organizations need internal guidelines that require transparency, documentation, and accountability when machine learning outputs influence strategic decisions. Future research can compare different sectors, examine larger groups of decision-makers, and explore how organizational culture, data literacy, and regulatory pressure shape the use of interpretable machine learning in strategic decision-making.

References

Adeoye-Olatunde, O. A., & Olenik, N. L. (2021). Research and scholarly methods: Semi-structured interviews. *JACCP: Journal of the American College of Clinical Pharmacy, 4*(10), 1358–1367. https://doi.org/10.1002/jac5.1441

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

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: An analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11*(5), e1424. https://doi.org/10.1002/widm.1424

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, 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

Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and Practice, 2*, 14. https://doi.org/10.1186/s42466-020-00059-z

Christou, P. A. (2023). How to use thematic analysis in qualitative research. *Journal of Qualitative Research in Tourism, 3*(2), 79–95. https://doi.org/10.4337/jqrt.2023.0006

Coussement, K., Abedin, M. Z., Kraus, M., Maldonado, S., & Topuz, K. (2024). Explainable AI for enhanced decision-making. *Decision Support Systems, 184*, 114276. https://doi.org/10.1016/j.dss.2024.114276

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Science, 9*(4), 322–345. https://doi.org/10.1287/stsc.2024.0190

Doshi, A. R., Bell, J. J., Mirzayev, E., & Vanneste, B. S. (2025). Generative artificial intelligence and evaluating strategic decisions. *Strategic Management Journal, 46*(3), 583–610. https://doi.org/10.1002/smj.3677

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in AI systems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3411764.3445188

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

Hamilton, A. B., & Finley, E. P. (2020). Reprint of: Qualitative methods in implementation research: An introduction. *Psychiatry Research, 283*, 112629. https://doi.org/10.1016/j.psychres.2019.112629

Hase, P., & Bansal, M. (2020). Evaluating explainable AI: Which algorithmic explanations help users predict model behavior? *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5540–5552. https://doi.org/10.18653/v1/2020.acl-main.491

Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative research. *American Journal of Pharmaceutical Education, 84*(1), 7120. https://doi.org/10.5688/ajpe7120

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376219

Keding, C. (2021). Understanding the interplay of artificial intelligence and strategic management: Four decades of research in review. *Management Review Quarterly, 71*, 91–134. https://doi.org/10.1007/s11301-020-00181-x

Kim, J., Liao, Q. V., & Wortman Vaughan, J. (2024). Human-centered evaluation of explainable AI applications: A systematic review. *Frontiers in Artificial Intelligence, 7*, 1456486. https://doi.org/10.3389/frai.2024.1456486

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376590

Lim, W. M. (2025). What is qualitative research? An overview and guidelines. *Australasian Marketing Journal, 33*(2), 199–229. https://doi.org/10.1177/14413582241264619

Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion, 106*, 102301. https://doi.org/10.1016/j.inffus.2024.102301

Magaldi, D., & Berler, M. (2020). Semi-structured interviews. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 4825–4830). Springer. https://doi.org/10.1007/978-3-319-24612-3_857

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems, 11*(3–4), Article 24. https://doi.org/10.1145/3387166

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods, 22*, 1–18. https://doi.org/10.1177/16094069231205789

Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., van Keulen, M., & Seifert, C. (2023). From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable AI. *ACM Computing Surveys, 55*(13s), Article 295. https://doi.org/10.1145/3583558

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*, 1–13. https://doi.org/10.1177/1609406919899220

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*, 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

Rong, Y., Leemann, T., Nguyen, T. T., Fiedler, L., Qian, P., Unhelkar, V., Seidel, T., Kasneci, G., & Kasneci, E. (2024). Towards human-centered explainable AI: A survey of user studies for model explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(4), 2104–2122. https://doi.org/10.1109/TPAMI.2023.3331846

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion, 76*, 89–106. https://doi.org/10.1016/j.inffus.2021.05.009