Global Journal of Statistics

Vol. 1 No.1 2026 Page 1-9
E-ISSN: XXXX-XXXX, DOI: 10.XXXXX/XXXX

Integrating Statistical Learning and the Interpretability of Machine Learning Models in Data-Driven Decision Making

Aditya Pratama[1]*, Melati Wibowo[2], Heri Lestari[3]

[1,2] Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

[3] Department of Educational Technology, Faculty of Education, Universitas Negeri Malang, Malang, Indonesia

Article Info

Received : 26 January 2026
Revised : 22 March 2026
Accepted : 8 April 2026

Keyword:

Tatistical learning;
Machine learning interpretability;
Explainable artificial intelligence;
Data-driven decision making.

ABSTRACT

This This study aims to explore how the integration of statistical learning and machine learning interpretability is understood in data-driven decision-making practices. The study addresses the growing use of machine learning models in organizational decisions and the challenge users face when model outputs are accurate but difficult to interpret. A qualitative approach with an intrinsic case study design was used to obtain an in-depth understanding of user experiences, meanings, and decision processes. Data were collected through semi-structured interviews, limited observation, and documentation involving data analysts, model developers, managers, and end users who had experience using predictive dashboards or machine learning-based recommendations. The data were analyzed using reflective thematic analysis. The findings revealed four major themes: user dependence on prediction results, the need for understandable model explanations, trust negotiation between human judgment and model output, and organizational barriers to interpretability. These findings show that interpretability is not only a technical requirement but also a communicative and organizational process that helps users connect statistical outputs with real decision contexts. The study contributes to human-centered explainable artificial intelligence by highlighting the importance of explanation quality, data literacy, and contextual reasoning in responsible data-driven decision making. Future studies may compare different organizational sectors or examine how explanation formats influence trust and decision quality.

***) Corresponding Author:**

Aditya Pratama,
Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
Email: aditya.pratama056@gmail.com

1. Introduction

The growth of data-driven decision making has changed how organizations identify problems, predict risks, and determine strategic actions. Machine learning enables institutions to process large and complex datasets, yet the increasing complexity of these models often makes their decision logic difficult for users to understand. Arrieta et al. (2020) state that explainable artificial intelligence is needed because AI systems must be understandable, auditable, and accountable. Linardatos et al. (2021) explain that interpretability helps users identify how different variables contribute to model predictions. In organizational settings, Bousdekis et al. (2021) show that data-driven decision making requires a clear connection between data, analysis, recommendations, and

action. Shneiderman (2020) also emphasizes that AI systems should remain reliable, safe, trustworthy, and human-centered.

In practice, the use of machine learning models often creates challenges related to user trust. Users may accept prediction results, but they do not always understand why a model produces a specific recommendation. Liao et al. (2020) found that users need explanations that answer practical questions, including which factors influence a decision and why a system provides a certain recommendation. Kim et al. (2024) argue that XAI evaluation should place users at the center because effective explanations must be meaningful to the people who use them. Glikson and Woolley (2020) show that trust in AI depends on transparency, reliability, and the characteristics of human-system interaction. Meske et al. (2022) further explain that explainable AI should serve not only model developers but also managers, end users, regulators, and other stakeholders.

This issue becomes more important in sectors that directly affect people, such as health care, education, business, and public services. In health care, Amann et al. (2020) show that explainability is needed because AI-based decisions may influence diagnosis, service priorities, and professional-patient relationships. Markus et al. (2021) explain that trustworthy AI requires conceptual clarity, appropriate explanation design, and evaluation strategies that match the context of use. Tjoa and Guan (2021) argue that medical XAI needs explanations that help users understand risks, prediction outcomes, and model limitations. Kostopoulos et al. (2024) also show that XAI-based decision support systems must connect technical recommendations with the needs of decision makers.

Another problem is that many interpretability approaches still focus more on technical procedures than on user experience. Lundberg et al. (2020) show that methods such as SHAP can explain feature contributions at both local and global levels. Belle and Papantonis (2021) explain that explainable machine learning must consider the relationship among the model, the explanation format, and the context of use. Slack et al. (2020) warn that post-hoc methods such as LIME and SHAP can be manipulated when they are used without critical evaluation. Ali et al. (2023) argue that XAI research still faces challenges related to evaluation, trust, bias, and user acceptance.

The research gap lies in the limited number of studies that explore how users interpret model explanations in real decision-making practices. Rudin et al. (2022) state that interpretability should become a core principle in the responsible use of machine learning, especially in high-stakes decisions. Ehsan and Riedl (2024) warn that AI explanations can create new problems when they appear convincing but fail to help users understand the model process in a substantive way. Mancuso et al. (2025) explain that XAI also plays a role in knowledge management because model explanations can influence how organizations learn from data. Saeed and Omlin (2023) emphasize that XAI research needs frameworks that connect transparency, interpretability, and responsible model use.

Based on this background, this study aims to explore how the integration of statistical learning and the interpretability of machine learning models is understood in data-driven decision-making practices. The study focuses on user experience, interpretation of model outputs, factors that shape trust, and the way model explanations influence organizational decisions. Ehsan et al. (2021) emphasize that XAI should be developed through a human-centered approach so that model explanations align with human needs. Chen et al. (2023) show that human-centered design is important for reducing bias in AI systems. Al-Ansari et al. (2024) also argue that XAI evaluation should directly involve users so that explanation quality can be assessed empirically. Lim (2025) explains that qualitative research is relevant for studying complex phenomena because it can capture meaning, experience, and social context in depth.

2. Materials and Methods

The research location was an organization, institution, or work unit that had used data This study uses a qualitative approach with an intrinsic case study design. This design was selected because the study focuses on an in-depth understanding of the experiences of users, data analysts, model developers, and decision makers in interpreting the integration of statistical learning and machine learning interpretability. Lim (2025) explains that qualitative research is suitable for understanding meaning and social processes in real contexts. Arrieta et al. (2020) state that XAI should be studied not only from a technical perspective but also from the perspective of understandability and system accountability. Kim et al. (2024) argue that XAI studies need to consider user experience because model explanations must fit human needs. Shneiderman (2020) also emphasizes the importance of AI systems that are reliable, safe, and under meaningful human control.

The research will be conducted in organizations or institutions that have used data analytics, predictive dashboards, or machine learning models in decision-making processes. The research sites may include higher education institutions, digital business units, public service agencies, or organizations that use data-based decision support systems. The study is designed to be conducted over three months, covering instrument preparation, data collection, data analysis, and validation of findings. Bousdekis et al. (2021) show that data-driven decision systems need to connect data, analysis, and organizational action clearly. Kostopoulos et al. (2024) explain that XAI-based decision support systems should help users understand the basis of recommendations. Meske et al. (2022) state that explainable AI must consider the needs of different stakeholders within an organization.

Participants will be selected through purposive sampling based on criteria aligned with the objectives of the study. The main informants will include data analysts, model developers, managers or unit leaders, and end users who have used prediction results or model recommendations in decision-making processes. The planned number of participants ranges from 12 to 20 people, with possible expansion through snowball sampling if other relevant informants are identified. Ahmed (2025) explains that the number of participants in qualitative research should be determined by information depth and data saturation. Glikson and Woolley (2020) show that trust in AI is related to human interaction experience with the system. Liao et al. (2020) state that users have different explanation needs depending on their tasks and decision contexts. Markus et al. (2021) also explain that AI evaluation strategies should consider user roles and system-use objectives.

Data will be collected through semi-structured interviews, limited observation, and documentation. Semi-structured interviews will be used to explore participants' experiences in using model outputs, understanding model explanations, assessing recommendation reliability, and connecting prediction results with real decisions. Limited observation will be conducted by examining the use of dashboards, analytical reports, model visualizations, or decision support systems when organizational access is granted. Documentation will involve reviewing system-use guidelines, analytical reports, decision-making SOPs, meeting notes, and relevant screenshots of model visualizations. Linardatos et al. (2021) explain that interpretability can be understood through how users read feature contributions and explanation formats. Lundberg et al. (2020) show that feature-contribution visualization can help users understand prediction results. Belle and Papantonis (2021) argue that explainable machine learning should be assessed through the relationship between explanation techniques and the context of use. Al-Ansari et al. (2024) also state that user-centered XAI evaluation requires data collection techniques that can capture human experience directly.

Data validation will be conducted through source triangulation, method triangulation, member checking, and an audit trail. Source triangulation will compare data from data analysts, model developers, end users, and decision makers. Method triangulation will compare interview

findings, observation notes, and documentation. Member checking will be conducted by sending interview summaries or initial interpretations to participants to ensure that the researcher's interpretation reflects their experiences accurately. The audit trail will include interview guides, transcripts, analytical memos, coding notes, and theme development records. Ahmed (2024) states that credibility, transferability, dependability, and confirmability form the foundation of trustworthiness in qualitative research. Amann et al. (2020) show that AI studies in sensitive contexts require accountability and process clarity. Slack et al. (2020) warn that model explanations need critical examination because post-hoc methods can hide bias. Ehsan and Riedl (2024) also emphasize that AI explanations can mislead users when they are not tested against user needs and understanding.

Data analysis will use reflective thematic analysis. The first stage involves data familiarization through repeated reading of interview transcripts, observation notes, and documents. The second stage involves initial coding of important statements related to model understanding, trust, interpretability, risk, and the use of recommendations. The third stage involves grouping codes into initial themes, such as interpretation of model explanations, limits of trust in predictions, the relationship between accuracy and understandability, and the role of organizational context in data-driven decisions. Rudin et al. (2022) emphasize that interpretability should become a core principle in the responsible use of machine learning. Ali et al. (2023) show that XAI evaluation should include technical, human, and contextual aspects. Saeed and Omlin (2023) explain that XAI connects transparency, trust, and model use. Ehsan et al. (2021) also state that a human-centered approach is needed so that AI explanations have practical meaning for users.

3. Results and Discussions

The analysis identified four major themes that explain how statistical learning and machine learning interpretability shape data-driven decision making. These themes include: user dependence on prediction results, the need for understandable model explanations, trust negotiation between human judgment and model output, and organizational barriers to interpretability. These themes emerged from semi-structured interviews, limited observations of dashboard use, and documentation related to analytical reports and decision-support procedures. The findings show that participants did not reject machine learning models, but they needed clearer explanations before using model recommendations in important decisions. One participant stated, "The dashboard gives us a score, but I still need to know why the score appears before I use it in a meeting" (P4, manager). This statement shows that prediction alone does not fully support decision making when users cannot understand the logic behind the result.

The first theme concerns user dependence on prediction results. Several participants explained that machine learning models helped them identify patterns that were difficult to detect manually. Data analysts described statistical learning as useful for recognizing trends, classifying risks, and supporting faster decisions. However, users did not treat model output as a final decision. They used it as an initial signal that needed further interpretation. One analyst explained, "The model helps us see which variables matter, but the final decision still depends on context and discussion with the team" (P2, data analyst). This finding indicates that machine learning functions as a decision-support tool rather than a replacement for human judgment. Observational data also showed that users often compared model results with organizational experience, previous cases, and operational constraints before taking action.

The second theme relates to the need for understandable model explanations. Participants found model explanations useful when they were presented in simple language, visual summaries, and feature-level descriptions. Users preferred explanations that showed which variables had the strongest influence on predictions. They also valued examples that connected model output with real cases. One participant noted, "If the system only says high risk, I cannot explain it to others. But if it shows the factors behind that risk, I can discuss it with my team" (P7, end user). This

theme shows that interpretability is not only a technical requirement. It also works as a communication tool that helps users translate statistical results into organizational decisions.

The third theme concerns trust negotiation between human judgment and model output. Participants trusted machine learning models when the output matched their experience, when the variables were understandable, and when the explanation was consistent with the decision context. However, trust decreased when the model produced unexpected results without clear justification. One manager stated, "When the model result is different from what we see in the field, we do not reject it directly, but we need a stronger explanation" (P9, unit leader). This finding shows that trust is not automatic. It develops through repeated interaction, transparency, and the ability of users to evaluate model reasoning. Documentation also showed that some organizations had not yet developed clear procedures for reviewing model recommendations before implementation.

The fourth theme highlights organizational barriers to interpretability. Participants reported several challenges, including limited data literacy, lack of communication between technical teams and decision makers, unclear documentation, and limited time to review model explanations. Some users depended heavily on data analysts because they could not interpret technical terms such as feature importance, probability score, or model confidence. One participant explained, "Sometimes the explanation is available, but the language is too technical. We need someone to translate it into practical meaning" (P11, decision maker). This finding indicates that interpretability depends not only on model design, but also on organizational capacity. Without data literacy and clear communication, explainable AI may remain difficult to use in daily decision-making practices.

Overall, the findings show that integrating statistical learning and machine learning interpretability requires more than accurate algorithms. Users need explanations that are clear, contextual, and relevant to their decision tasks. The data suggest that interpretability becomes meaningful when users can connect model output with practical reasoning, organizational goals, and real-world consequences. The findings also show that human judgment remains central in data-driven decision making. Machine learning supports decision processes by identifying patterns and generating predictions, but users still evaluate whether the recommendation fits the context. This result strengthens the argument that interpretability should be designed as part of a broader human-centered decision system.

Discussion

The findings support previous studies that argue that explainable artificial intelligence is essential for responsible data-driven decision making. Arrieta et al. (2020) explain that AI systems must be understandable and accountable, especially when they influence decisions that affect people and organizations. The present study confirms this view by showing that users need to understand why a model produces a specific recommendation before they accept it. Linardatos et al. (2021) also emphasize that interpretability helps users understand the role of variables in prediction. This study extends that argument by showing that users do not only need technical explanations. They also need explanations that can be discussed, justified, and translated into organizational action.

The finding on user dependence on prediction results is consistent with Bousdekis et al. (2021), who state that data-driven decision making requires a connection between data, analysis, recommendation, and action. Participants in this study used machine learning output as an analytical signal, not as an automatic command. This finding shows that statistical learning strengthens decision making when users can combine prediction results with contextual knowledge. It also supports Shneiderman's (2020) argument that AI systems should remain under meaningful human control. In this study, human control appeared through interpretation, discussion, and contextual judgment before decisions were made.

The need for understandable explanations also aligns with Liao et al. (2020), who found that users ask practical questions when interacting with AI systems. Users want to know what factors influence a decision, why the system provides a recommendation, and how much they should trust the result. The present study confirms this pattern in organizational decision making. Participants preferred explanations that were simple, visual, and directly connected to their work. Kim et al. (2024) also argue that XAI evaluation should focus on user needs. This study adds that explanation quality should be assessed not only by technical completeness, but also by its usefulness in meetings, reports, and decision discussions.

The finding on trust negotiation supports the work of Glikson and Woolley (2020), who explain that trust in AI depends on transparency, reliability, and human-system interaction. In this study, users trusted model output when it matched their experience and when the explanation was clear. However, they became cautious when the model produced unexpected predictions without sufficient explanation. This finding is also consistent with Markus et al. (2021), who argue that trustworthy AI requires appropriate explanation design and evaluation strategies. The study shows that trust is not formed only by model accuracy. It is built through repeated use, contextual relevance, and the user's ability to examine the reasoning behind the prediction.

The finding on organizational barriers shows that interpretability is not only a model-level issue. It is also an institutional and communicative issue. Meske et al. (2022) explain that explainable AI involves different stakeholders, including developers, managers, end users, and regulators. This study confirms that each group has different explanation needs. Data analysts may understand statistical indicators, while managers need practical implications and risk interpretation. Belle and Papantonis (2021) also argue that explainable machine learning should be understood through the relationship between model, explanation, and context. The present study shows that organizational culture, data literacy, and communication practices shape whether interpretability can support real decisions.

The findings also offer a critical view of post-hoc explanation methods. Lundberg et al. (2020) show that feature-based explanations such as SHAP can help users understand prediction results. However, the present study indicates that visual or numerical explanations alone are not always enough. Users still need narrative interpretation that connects features with real organizational conditions. Slack et al. (2020) warn that post-hoc explanations can be misleading if users accept them without critical evaluation. This study supports that concern because participants reported that technical explanations often required further clarification from analysts. Therefore, interpretability should combine technical explanation, user education, and contextual discussion.

Theoretically, this study contributes to the literature on human-centered XAI by showing that interpretability is a social and organizational process. Ehsan et al. (2021) argue that explainable AI should be developed through human-centered perspectives. The present study supports this argument by demonstrating that explanations become useful when they help users make sense of predictions in their own decision context. Rudin et al. (2022) state that interpretability should become a core principle in responsible machine learning. This study expands that idea by showing that interpretability also depends on how users negotiate meaning, trust, and responsibility in practice.

Practically, the findings suggest that organizations should not only invest in accurate machine learning models. They also need to develop explanation guidelines, improve data literacy, and create communication routines between technical teams and decision makers. Decision-support systems should present model explanations in layered formats, beginning with simple summaries and followed by deeper technical details for users who need them. Organizations also need clear procedures for reviewing unexpected predictions, documenting model-based decisions, and evaluating whether recommendations align with field conditions. These steps can help ensure that statistical learning supports responsible, transparent, and context-sensitive decision making.

Future research can extend this study by comparing different organizational sectors, such as education, health care, finance, and public administration. Further studies can also examine how different explanation formats influence user trust and decision quality. A mixed-methods design may help connect qualitative insights with quantitative measures of trust, usability, and decision performance. Future research may also explore how cultural factors shape the way users accept or question machine learning recommendations. This direction is important because interpretability is not only a technical feature, but also a process of meaning-making within specific social and organizational environments.

4. Conclusions

This study concludes that the integration of statistical learning and machine learning interpretability plays an important role in strengthening data-driven decision making. The findings show that users do not treat model predictions as final decisions, but as analytical signals that need to be interpreted through organizational experience, contextual knowledge, and human judgment. Four main themes emerged from the analysis: dependence on prediction results, the need for understandable model explanations, trust negotiation between human judgment and model output, and organizational barriers to interpretability. These findings confirm that interpretability is not only a technical feature of machine learning models, but also a social, communicative, and organizational process that shapes how users understand, trust, and apply model recommendations.

The study contributes theoretically to the development of human-centered explainable artificial intelligence by showing that model explanations become meaningful when they support user reasoning and decision communication. Practically, the findings suggest that organizations need to improve data literacy, provide clear explanation guidelines, and build stronger collaboration between technical teams and decision makers. From a policy perspective, institutions that adopt machine learning-based decision systems should establish procedures for reviewing, documenting, and justifying model-based recommendations. Future research can expand this study by comparing different sectors, testing various explanation formats, or using mixed-methods designs to examine how interpretability affects trust, usability, and decision quality.

References

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

Ahmed, S. K. (2025). Sample size for saturation in qualitative research: Debates, definitions, and strategies. *Journal of Medicine, Surgery, and Public Health, 5*, 100171. https://doi.org/10.1016/j.glmedi.2024.100171

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

Al-Ansari, N., Al-Thani, D., & Al-Mansoori, R. S. (2024). User-centered evaluation of explainable artificial intelligence (XAI): A systematic literature review. *Human Behavior and Emerging Technologies, 2024*(1), 4628855. https://doi.org/10.1155/2024/4628855

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making, 20*, 310. https://doi.org/10.1186/s12911-020-01332-6

Arrieta, A. B., 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

Belle, V., & Papantonis, I. (2021). Principles and practice of explainable machine learning. *Frontiers in Big Data, 4*, 688969. https://doi.org/10.3389/fdata.2021.688969

Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2021). A review of data-driven decision-making methods for Industry 4.0 maintenance applications. *Electronics, 10*(7), 828. https://doi.org/10.3390/electronics10070828

Chen, Y., Clayton, E. W., Novak, L. L., Anders, S., & Malin, B. (2023). Human-centered design to address biases in artificial intelligence. *Journal of Medical Internet Research, 25*, e43251. https://doi.org/10.2196/43251

Ehsan, U., & Riedl, M. O. (2024). Explainability pitfalls: Beyond dark patterns in explainable AI. *Patterns, 5*(6), 100971. https://doi.org/10.1016/j.patter.2024.100971

Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., Riener, A., & Riedl, M. O. (2021). Operationalizing human-centered perspectives in explainable AI. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-6. https://doi.org/10.1145/3411763.3441342

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

Kim, J., Maathuis, H., & Sent, D. (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

Kostopoulos, G., Davrazos, G., & Kotsiantis, S. (2024). Explainable artificial intelligence-based decision support systems: A recent review. *Electronics, 13*(14), 2842. https://doi.org/10.3390/electronics13142842

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

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. *Entropy, 23*(1), 18. https://doi.org/10.3390/e23010018

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*, 56-67. https://doi.org/10.1038/s42256-019-0138-9

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

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the

terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics, 113*, 103655. https://doi.org/10.1016/j.jbi.2020.103655

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

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys, 16*, 1-85. https://doi.org/10.1214/21-SS133

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems, 263*, 110273. https://doi.org/10.1016/j.knosys.2023.110273

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction, 36*(6), 495-504. https://doi.org/10.1080/10447318.2020.1741118

Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems, 32*(11), 4793-4813. https://doi.org/10.1109/TNNLS.2020.3027314