

Interpretable Artificial Intelligence in Information Systems: Status Review and Future Research Directions

Saurabh Sudhakar Umredkar, Swapnil Anil Bagde, Sonu Ramkumar Shahu, Prof Nikita Khanzode

Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, India

Abstract: Efforts to develop black-box artificial intelligence (AI) systems have become a phenomenon of emerging global interest in academia, business, and society, and have led to the development of the XAI research field. With its pluralistic perspective, information systems (IS) research is destined to contribute to this emerging field; thus, it is not surprising that the number of research publications at XAI has increased significantly. This paper aims to provide a comprehensive overview of XAI research in public and electronic markets, specifically using a structured literature review. Based on a literature review of 180 research papers, this work examines the most receptive points, the development of academic debates, and the most important concepts and methodologies. In addition, eight research areas with different levels of maturity in e-markets are identified. Finally, guidelines for the XAI research agenda in IS are presented.

Keywords: Artificial intelligence.

REFERENCES

- [1]. Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI)* (pp. 1–18). <http://dl.acm.org/citation.cfm?doi=3173574.3174156>
- [2]. Abdul, A., Weth, C. von der, Kankanhalli, M., & Lim, B. Y. (2020). COGAM: Measuring and moderating cognitive load in machine learning model explanations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI)* (pp. 1–14). <https://doi.org/10.1145/3313831.3376615>
- [3]. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- [4]. Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2), 427–445. <https://doi.org/10.1007/s12525-020-00414-7>
- [5]. Aghaeipoor, F., Javidi, M. M., & Fernandez, A. (2021). IFC-BD: An interpretable fuzzy classifier for boosting explainable artificial intelligence in big data. *IEEE Transactions on Fuzzy Systems*. Advance online publication. <https://doi.org/10.1109/TFUZZ.2021.3049911>
- [6]. Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60, 102387. <https://doi.org/10.1016/j.ijinfomgt.2021.102387>
- [7]. Akter, S., Hossain, M. A., Lu, Q. S., & Shams, S. R. (2021b). Big data-driven strategic orientation in international marketing. *International Marketing Review*, 38(5), 927–947. <https://doi.org/10.1108/IMR-11-2020-0256>
- [8]. Alam, L., & Mueller, S. (2021). Examining the effect of explanation on satisfaction and trust in AI diagnostic systems. *BMC Medical Informatics and Decision Making*, 21(1), 1–15. <https://doi.org/10.1186/s12911-021-01542-6>
- [9]. 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(1), 1–9. <https://doi.org/10.1186/s12911-020-01332-6>

- [10]. Areosa, I., & Torgo, L. (2019). Visual interpretation of regression error. In P. Moura Oliveira, P. Novais, & L. P. Reis (Eds.), *Lecture notes in computer science. Progress in artificial intelligence* (pp. 473–485). Springer International Publishing. https://doi.org/10.1007/978-3-030-30244-3_39
- [11]. 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>
- [12]. Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T. & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, 22(2). <https://aisel.aisnet.org/jais/vol22/iss2/8>
- [13]. Australian Broadcasting Corporation. (2022). *Robodebt inquiry: Royal commission on unlawful debt scheme begins*. ABC News. https://www.youtube.com/results?search_query=robodebt+royal+commission. Accessed 02 Feb 2023
- [14]. Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, 45(1). <https://doi.org/10.25300/MISQ/2021/15882>
- [15]. Balijepally, V., Mangalaraj, G., & Iyengar, K. (2011). Are we wielding this hammer correctly? A reflective review of the application of cluster analysis in information systems research. *Journal of the Association for Information Systems*, 12(5), 375–413. <https://doi.org/10.17705/1jais.00266>
- [16]. Bandara, W., Miskon, S., & Fieft, E. (2011). A systematic, tool-supported method for conducting literature reviews in information systems. *Proceedings of the 19th European Conference on Information Systems (ECIS 2011)* (p. 221). Helsinki, Finland. <https://eprints.qut.edu.au/42184/1/42184c.pdf>
- [17]. Barakat, N. H., Bradley, A. P., & Barakat, M. N. H. (2010). Intelligible support vector machines for diagnosis of diabetes mellitus. *IEEE Transactions on Information Technology in Biomedicine*, 14(4), 1114–1120. <https://doi.org/10.1109/TITB.2009.2039485>
- [18]. Barda, A. J., Horvat, C. M., & Hochheiser, H. (2020). A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. *BMC Medical Informatics and Decision Making*, 20(1), 1–16. <https://doi.org/10.1186/s12911-020-01276-x>