Electronic Markets (2022) 32:2103–2106
https://doi.org/10.1007/s12525-022-00607-2

PREFACE

Explainable and responsible artificial intelligence

Christian Meske[1] · Babak Abedin[2] · Mathias Klier[3] · Fethi Rabhi[4]

Published online: 29 November 2022

Keywords Artificial intelligence · Explainability · Transparency · Responsibility

Introduction

Today's algorithms already reached or even surpassed the task performance of humans in various domains. Especially, Artificial Intelligence (AI) plays a central role for the interaction between organizations and individuals such as their customers, transforming for instance electronic commerce or customer relationship management. However, most AI systems are still "black boxes" that are difficult to comprehend—not only for developers, but also for consumers and decision-makers (Meske et al., 2022). With regards to electronic markets, problems such as trying to manage the risk and ensure regulatory compliance of electronic trading systems based on machine learning stem not only from their data-driven nature and technical complexity, but also from their black-box nature, where the "learning" creates non-transparent dependencies between inputs and outputs (Cliff & Treleaven, 2010). This raises many challenges such as ensuring data quality issues, managing provenance information needed for transparency as well as organizing metadata when combining data from multiple sources (Rabhi et al., 2020). Thus, a responsible and more trustworthy AI is demanded (HLEG-AI, 2019; Thiebes et al., 2021; Schneider et al., 2022).

This is where research on Explainable Artificial Intelligence (XAI) comes in. Also referred to as "interpretable" or "understandable AI", XAI aims to "produce explainable models, while maintaining a high level of learning performance (prediction accuracy); and enable human users to understand, appropriately, trust, and effectively manage the emerging generation of artificially intelligent partners" (Defense Advanced Research Projects Agency (DARPA), 2017). XAI hence refers to "the movement, initiatives, and efforts made in response to AI transparency and trust concerns, more than to a formal technical concept" (Adadi & Berrada, 2018, p. 52,140). XAI is designed in a user-centric fashion so that users are empowered to scrutinize AI (Förster et al., 2020). Overall, XAI objectives are to evaluate, to improve, to learn from, and to justify AI, in order to eventually be able to manage AI (Meske et al., 2022). The need for greater explainability has been recognized in both academic and industry settings as for example tech giants such as Google, Facebook, Microsoft, Amazon and IBM create partnerships with academics and practitioners on platforms such as Partnership on AI (https://www.partnershiponai.org/) to foster public discussions and to improve people's understanding of AI and its consequences (Abedin, 2022).

The special issue

With a focus on the transformation of electronic markets, in this special issue, we explore and extend research on how to establish explainability and responsibility in intelligent black box systems—machine learning-based or not. The submitted

Responsible Editor: Christian Meske

✉ Christian Meske
christian.meske@rub.de

Babak Abedin
Babak.Abedin@mq.edu.au

Mathias Klier
mathias.klier@uni-ulm.de

Fethi Rabhi
f.rabhi@unsw.edu.au

[1] Institute of Work Science, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany

[2] Macquarie Business School, Macquarie University, Balaclava Rd, Macquarie Park, NSW 2109, Australia

[3] University of Ulm, Institute of Business Analytics, Helmholtzstraße 22, 89081 Ulm, Germany

[4] School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

papers in this special issue show a wide range of domains and methods that can be applied to add knowledge in the field of XAI. They also showed that while AI is increasingly being embedded in various information systems (ISs) such as medical information systems, human resource systems, banking and financial systems, and autonomous vehicles, the explainability and understandability aspects of interactions between humans and AI systems remain under-researched (Abedin et al. 2022). However, recent research in IS and other disciplines indicates a substantial increase in research and interest in understanding the black box AI operations and logic and in being able to communicate that with users and other stakeholders.

We are delighted to introduce our special issue on “Explainable and responsible artificial intelligence”. The call was announced in 2021 with April 2022 as the deadline for submissions. Subsequently, Electronic Markets sponsored our second mini-track on "Explainable Artificial Intelligence (XAI)" at the 55th Hawaiian International Conference on Systems Science (HICSS) from which papers were invited to submit extended manuscripts for this special issue.

The overarching goal of this special issue was to stress the need for conceptualizing and empirically studying explainability of AI and to invite research for exploring, theorizing, and testing guidelines for upholding and implementing good XAI practices with regards to individual, group/team, organizational, and societal level of analysis and with a focus on electronic markets domains. The special issue had a broad target audience, including academics, practitioners, and policy makers who could demonstrate innovative strategies and resolutions to the described tensions, risks, and opportunities.

Accepted papers

In the first paper, “Explainable product backorder prediction exploiting CNN: Introducing explainable models in businesses”, the authors Md Shajalal, Alexander Boden and Gunnar Stevens (Shajalal et al., 2022) explore how intelligent predictive models could be made explainable to the stakeholders in strategic inventory management. It proposes a new convolutional neural network (CNN)-based explainable predictive model for product backorder prediction. The paper employs Shapley Additive Explanations to explain the overall model’s priority in decision making. Besides that, the paper introduces locally interpretable surrogate models that can explain any individual prediction. The experimental results demonstrate effectiveness in predicting backorders in terms of standard evaluation metrics. In addition, the proposed methods and explanation generation techniques can apply to other predictive tasks in the business domain.

The second paper, “Global reconstruction of language models with linguistic rules – Explainable AI for online consumer reviews”, is written by Markus Binder, Bernd Heinrich, Marcus Hopf, and Alexander Schiller (Binder et al., 2022). The authors address problems caused by “black box” language models such as BERT which are unclear on how they arrive at their predictions. Yet, many eCommerce applications require checks and justifications by means of global reconstruction of their predictions, since the decisions based thereon can have large impacts or are even mandatory due to regulations such as the GDPR. To this end, the paper proposes a novel global XAI approach by means of linguistic rules based on NLP building blocks (e.g., part-of-speech) and analyze it on different datasets of online consumer reviews and NLP tasks. Since this approach allows for different setups, this paper is the first to analyze the trade-off between comprehensibility and fidelity of global reconstructions of language model predictions. With respect to this trade-off, the paper finds that this approach indeed allows for balanced setups for global reconstructions of BERT’s predictions. Thus, this approach paves the way for a thorough understanding of language model predictions in text analytics. In practice, this approach can assist businesses in their decision-making and supports compliance with regulatory requirements.

The third paper is authored by Jana Gerlach, Paul Hoppe, Sarah Jagels, Luisa Licker, and Michael Breitner (Gerlach et al., 2022), and entitled “Decision support for efficient XAI services – A morphological analysis, business model archetypes, and a decision tree”. The market volume of XAI services has grown significantly and XAI services are applied to explore relationships in data, improve AI methods, justify AI decisions, and control AI technologies. In their paper, the authors contribute to theory and practice by deducing XAI archetypes and developing a user-centric decision support framework to identify the XAI services most suitable for the requirements of relevant stakeholders including managers, regulators, users of XAI models, developers, and consumers. The proposed decision tree is founded on a literature-based morphological box and a classification of real-world XAI services.

In the fourth paper, “Designing a feature selection method based on explainable artificial intelligence”, the authors Jan Zacharias, Moritz von Zahn, Johannes Chen and Oliver Hinz (Zacharias et al., 2022) argue that AI system's pre-processing stage has been unjustifiably neglected and should receive greater attention in current efforts to establish explainability. The authors focus on introducing explainability to an integral part of the pre-processing stage: feature selection, and build upon design science research to develop a design framework for explainable feature selection. They instantiate the design framework in a running software artifact and evaluate it in two expert workshops. The achieved artifact helps organizations to persuasively justify feature selection to stakeholders and, thus, comply with upcoming AI legislation. The authors further provide researchers and practitioners with a design framework consisting of meta-requirements and design principles for explainable feature selection.

The fifth paper, titled "A nascent design theory for explainable intelligent systems", written by Lukas-Valentin Herm, Theresa Steinbach, Jonas Wanner and Christian Janiesch (Herm et al., 2022), focuses on the challenge, that the complexity of intelligent systems renders the user hardly able to comprehend the inherent decision logic of the underlying machine learning model. As a result, so they argue, the adoption of this technology, especially for high-stake scenarios, is hampered. In this context, XAI offers numerous starting points for making the inherent logic explainable to people. While research manifests the necessity for incorporating XAI into intelligent systems, there is still a lack of knowledge about how to socio-technically design these systems to address acceptance barriers among different user groups. In response, the authors have derived and evaluated a nascent design theory for explainable intelligent systems based on a structured literature review, two qualitative expert studies, a real-world use case application, and quantitative research. Their design theory includes design requirements, design principles, and design features covering the topics of global explainability, local explainability, personalized interface design, and psychological/emotional factors.

In the sixth paper, "Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring", the authors Lennart Hofeditz, Sünje Clausen, Alexander Rieß, Milad Mirbabaie and Stefan Stieglitz (Hofeditz et al., 2022) examine the proliferation of application of AI-based systems in the human resource field and particularly in candidate selections. In particular, the study focuses on the final hiring decision, which often remains with humans and is prone to human biases. The study investigates the impact of an AI-based system's candidate recommendations on humans' hiring decisions and how this relation could be moderated by an XAI approach. A self-developed platform was used to conduct an online experiment with 194 participants. The quantitative and qualitative findings showed that the recommendations of an AI-based system can reduce discrimination against older and female candidates but appear to cause fewer selections of foreign-race candidates. Contrary to expectations, the same XAI approach moderated these effects differently depending on the context.

The seventh and last paper, "Explanation matters: An experimental study on explainable AI", is written by Pascal Hamm, Michael Klesel, Patricia Coberger and H. Felix Wittmann (Hamm et al., forthcoming). While previous literature has already addressed the technological benefits of XAI, there has been little research comparing XAI and AI from the user's perspective. Building upon the theory of trust, the authors theorize that post-hoc explainability (using Shapley Additive Explanations) has a significant impact on use-related variables in this context. To test their model, they conduct an experimental task asking participants to compare signatures and detect forged signatures. Surprisingly, their study shows that XAI has only a small but significant impact on perceived explainability. Nevertheless, they demonstrate that a high level of perceived explainability has a strong impact on important constructs including trust and perceived usefulness. A post-hoc analysis shows that hedonic factors are significantly related to perceived explainability and require more attention in future research. They conclude with important directions for academia and for organizations.

Conclusion and directions for future research

In various ways, the accepted papers provide scientific innovations and advance our understanding of how explainable and responsible artificial intelligence can be established. On a more abstract level, Gerlach et al. (2022) introduced a user-centric decision tree applicable for different stakeholders to support them in identifying the most suitable XAI services for their individual requirements. Herm et al. (2022) established a nascent design theory for explainable intelligent systems, including design requirements, principles, and features for such artefacts. New XAI methods were proposed, such as a novel global XAI approach by means of linguistic rules based on NLP building block (Binder et al., 2022), while Zacharias et al. (2022) suggested an innovation regarding the pre-processing stage of AI by a new feature selection method based on XAI. Regarding the impact of XAI, it has been shown that transparent AI can be a very effective tool for decision support in different business domains such as inventory management (Shajalal et al., 2022) or human resources (Hofeditz et al., 2022), and others. For instance, the latter study highlighted that with XAI, bias and discrimination in decision making can be mitigated. Finally, Hamm et al. (forthcoming) showed that explanations matter for users, and that perceived explainability has an important influence on the individual's trust towards AI.

The contributions indicate that explainability is a prerequisite for responsibility of AI usage as we believe moving forward, XAI will receive even more attention from scholars and practitioners especially as ethical and rules and principles shape and become part of compliance expectations and Government law and rules. These contributions also illustrate the challenges associated with adopting XAI within organizations from ensuring user centricity to presenting information in a way that is comprehensible to experts from different domains like HR or eCommerce. We expect to see more of the XAI design theories to be translated into XAI practices, extending the notion of explainability to other trust-based considerations as well as embedding them into emerging frameworks that take a holistic approach to the development and management of AI systems. We also expect more research from socio-technical and non-technical perspectives into further theorization of XAI and its interactions with end users and other stakeholders as well as examination of ways that such interactions influence AI design and development practices and outcomes.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2021). Designing and managing human-AI interactions. *Information Systems Frontiers, 24*(3), 691–697. https://doi.org/10.1007/s10796-022-10313-1

Abedin, B. (2022). Managing the tension between opposing effects of explainability of artificial intelligence: A contingency theory perspective. *Internet Research., 32*(2), 425–453. https://doi.org/10.1145/3479645.3479709

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

Binder, M., Heinrich, B., Hopf, M., & Schiller, A. (2022). Global reconstruction of language models with linguistic rules – Explainable AI for online consumer reviews. *Electronic Markets, 32*(4). https://doi.org/10.1007/s12525-022-00612-5

Cliff D., & Treleaven, P. (2010). *Technology trends in the financial markets: A 2020 vision.* UK Government Office for science's foresight driver review on the future of computer trading in financial Markets – DR 3, October 2010.

Defense Advanced Research Projects Agency (DARPA) (2017). *Explainable artificial intelligence (XAI).* https://www.darpa.mil/program/explainableartificial-intelligence. Accessed 7 April 2021.

Förster, M., Klier, M., Kluge, K., & Sigler, I. (2020). Fostering human agency: A process for the design of user-centric XAI systems. *Proceedings of the 41st International Conference on Information Systems (ICIS).* https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/12.

Gerlach, J., Hoppe, P., Jagels, S., Licker, L., & Breitner, M. H. (2022). Decision support for efficient XAI services – A morphological analysis, business model archetypes, and a Decision Tree. *Electronic Markets, 32*(4). https://doi.org/10.1007/s12525-022-00603-6

Herm, L. V., Steinbach, T., Wanner, J., & Janiesch, C. (2022). A nascent design theory for explainable intelligent systems. *Electronic Markets, 32*(4). https://doi.org/10.1007/s12525-022-00606-3

Hofeditz, L., Clausen, S., Rieß, A., Mirbabaie, M., & Stieglitz, S. (2022). Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring. *Electronic Markets, 32*(4). https://doi.org/10.1007/s12525-022-00600-9

HLEG-AI. (2019). *Ethics guidelines for trustworthy artificial intelligence*. Brussels: Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. Accessed 25/09/2022 https://eskillsalliancecms.gov.mt/en/news/Documents/2019/AIDefinition.pdf

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

Rabhi, F. A., Mehandjiev, N., & Baghdadi, A. (2020). State-of-the-art in applying machine learning to electronic trading. In *International Workshop on Enterprise Applications, Markets and Services in the Finance Industry* (pp. 3–20). *Springer Lecture Notes in Business Information Processing*, vol 401. https://doi.org/10.1007/978-3-030-64466-6_1.

Shajalal, M., Boden, A., & Stevens, G. (2022). Explainable product backorder prediction exploiting CNN: introducing explainable models in businesses. *Electronic Markets, 32*(4). https://doi.org/10.1007/s12525-022-00599-z

Schneider, J., Abraham, R., Meske, C., & vom Brocke, J. (2022). Artificial intelligence governance for businesses. *Information Systems Management*, pp. 1–21. https://doi.org/10.1080/10580530.2022.2085825

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets, 31*(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

Zacharias, J., von Zahn, M., Chen, J., & Hinz, O. (2022). Designing a feature selection method based on explainable artificial intelligence. *Electronic Markets, 32*(4). https://doi.org/10.1007/s12525-022-00608-1