Human Interaction and Emerging Technologies (IHIET 2023), Vol. 111, 2023, 609–619
https://doi.org/10.54941/ahfe1004068

Assessing the Transparency and Explainability of AI Algorithms in Planning and Scheduling Tools: A Review of the Literature

Sofia Morandini, Federico Fraboni, Enzo Balatti, Aranka Hackmann, Hannah Brendel, Gabriele Puzzo, Lucia Volpi, Davide Giusino, Marco De Angelis, and Luca Pietrantoni

Alma Mater Studiorum - University of Bologna, Bologna, BO, 40127, Italy

ABSTRACT

As AI technologies enter our working lives at an ever-increasing pace, there is a greater need for AI systems to work synergistically with humans in the workplace. One critical requirement for such synergistic human-AI interaction is that the AI systems' behavior be explainable to the humans in the loop. The performance of decision-making by artificial intelligence has exceeded the capability of human beings in many specific domains. In the AI decision-making process, the inherent black-box algorithms and opaque system information lead to highly correct but incomprehensible results. The need for explainability of intelligent decision-making is becoming urgent and a transparent process can strengthen trust between humans and machines. The TUPLES project, a three-year Horizon Europe R&I project, aims to bridge this gap by developing AI-based planning and scheduling (P&S) tools using a comprehensive, human-centered approach. TUPLES leverages data-driven and knowledge-based symbolic AI methods to provide scalable, transparent, robust, and secure algorithmic planning and scheduling systems solutions. It adopts a use-case-oriented methodology to ensure practical applicability. Use cases are chosen based on input from industry experts, cutting-edge advances, and manageable risks (e.g., manufacturing, aviation, waste management). The EU guidelines for Trustworthy Artificial Intelligence highlight key requirements such as human agency and oversight, transparency, fairness, societal well-being, and accountability. The Assessment List for Trustworthy Artificial Intelligence (ALTAI) is a practical self-assessment tool for businesses and organizations to evaluate their AI systems. Existing AI-based P&S tools only partially meet these criteria, so innovative AI development approaches are necessary. We conducted a literature review to explore current research on AI algorithms' transparency and explainability in P&S, aiming to identify metrics and recommendations. The findings highlighted the importance of Explainable AI (XAI) in AI design and implementation. XAI addresses the black box problem by making AI systems explainable, meaningful, and accurate. It uses pre-modeling, in-modeling, and post-modeling explainability techniques, relying on psychological concepts of human explanation and interpretation for a human-centered approach. The review pinpoints specific XAI methods and offered evidence to guide the selection of XAI tools in planning and scheduling.

Keywords: Human-AI interaction, Trustworthy AI, Decision-making

INTRODUCTION

As the integration of artificial intelligence (AI) technologies into our professional lives accelerates, the need to develop seamless human-AI collaboration systems becomes crucial. A key prerequisite enabling synergistic human-AI partnership is ensuring that AI systems exhibit explainable behaviors to the humans involved in the AI-based decision-making process. Explainability in AI systems, also referred to as interpretable or understandable AI, is an attribute that enables humans to comprehend the rationale underpinning a model's decisions, predictions, or outputs. Dwivedi et al. (2023) differentiate between the following explainability-related concepts: *transparency*, which pertains to an AI system's capacity to offer lucid and comprehensible justifications for its decisions or actions; *trustworthiness*, which denotes the reliability and accuracy of an AI system in its decision-making processes; and *interpretability*, which involves an AI system's capability to elucidate the reasoning behind a specific decision or output.

In numerous specific domains, the decision-making process of AI has surpassed that of humans (Minh et al. 2022). However, the inherent opacity of black-box algorithms and the obscurity of system information often yield indecipherable outcomes (Chromik and Butz, 2021). These indecipherable outcomes pose challenges such as lack of trust, potential biases, and hindered adoption of AI systems in critical decision-making scenarios. Transparency and explainability in AI systems become imperative to address these challenges, fostering trust between humans and machines, and enabling users to identify and mitigate biases in AI-generated decisions (Chatila et al. 2021).

As a matter of fact, research on Explainable AI (XAI) has made considerable strides in acknowledging the necessity of rendering opaque models more transparent (Dwivedi et al. 2023). Although numerous studies have demonstrated that XAI methodologies can enhance users' grasp of black-box models (Adadi and Berrada, 2018; Miller, 2019), others have highlighted the challenges that arise from the disharmony between individuals' cognitive limitations and existing XAI techniques (Bertrand et al. 2022; Bunt, Lount and Lauzon, 2012). Consequently, the need to clarify AI decision-making processes is increasing, and a transparent procedure can bolster trust between humans and machines.

THE TUPLES PROJECT

The European Commission's (EC) policy on Trustworthy AI emphasizes developing and testing AI systems that are ethical, transparent, and accountable (European Commission, 2021). These systems should respect human rights, democratic values, and the rule of law while prioritizing human agency and oversight. The policy promotes the implementation of AI technologies that benefit societal well-being, ensure privacy, and foster security. Moreover, the EC underscores the importance of robustness, fairness, and non-discrimination in AI systems, along with the need for traceability and clear communication of AI-driven decisions. These principles aim to foster a human-centric AI ecosystem that safeguards public trust and enables sustainable innovation.

The TUPLES project, a three-year research and innovation initiative funded by Horizon Europe, focuses on the development of AI-driven planning P&S tools that emphasize explainability in decision-making processes (TUPLES, 2023). AI-driven P&S systems aim to optimize resource allocation, streamline processes, and enhance decision-making in complex, dynamic environments. Specifically, these systems leverage AI algorithms and techniques to efficiently plan tasks, allocate resources, and schedule activities, considering various constraints, objectives, and uncertainties. By automating the P&S process, AI-driven P&S systems can significantly improve productivity, reduce operational costs, minimize human error, and support organizations in adapting to changing circumstances (Amershi et al. 2019). These systems can be applied in a wide range of industries, amongst others, manufacturing, logistics, transportation, healthcare, and project management (Kotriwala et al. 2021).

At the core of TUPLES project lies a comprehensive, human-centric methodology that adopts a use-case-oriented approach which guarantees practical applicability in real-world scenarios by developing AI-driven P&S. The project selects use cases based on input from industry experts, cutting-edge advances, and manageable risks, with examples of industries including manufacturing, aviation, and waste management. In these various industries, explainable AI in P&S systems can improve transparency and enhance collaboration. For instance, AI-driven workforce scheduling systems, which assign employees to shifts by considering factors such as employee preferences, skill sets, and regulatory requirements, can benefit from XAI. By clarifying the decision-making process, XAI enables managers to adjust schedules based on their understanding of the system's rationale, fostering more effective collaboration.

METHOD

A narrative literature review was conducted to identify techniques and methods for designing and implementing XAI systems in P&S. Additionally, the review sought to identify relevant metrics for evaluating XAI and provide recommendations to ensure the explainability and transparency of AI algorithms employed in P&S contexts. To perform the literature review, we searched extensively using keywords including Explainable AI, XAI, Transparen*, Trustworth*, Planning, and Scheduling in bibliographical databases such as Scopus, Web of Science, and Google Scholar. The search yielded 57 records, which were then assessed for relevance by the research team. Two authors initially skimmed the sources based on title, abstract, and keywords. Then, two other researchers conducted a second skimming based on a more thorough examination of the papers' contents, identifying their relevance to our purpose, outcomes, and recency. This process continued with two more researchers, who further refined the selection based on additional criteria. In total, 28 sources were selected for inclusion in our review.

RESULTS

Techniques and Methods

In the pursuit of enhancing transparency in AI systems, Minh et al. (2022) conducted a comprehensive review, categorizing XAI methods into three primary groups: pre-modeling explainability, interpretable model, and post-modeling explainability. Each approach presents its own advantages and disadvantages from various perspectives. About the first category, the pre-modeling explainability methods seek to improve transparency before the model undergoes training. This approach envisages early identification of potential biases and errors; however, it may necessitate additional resources and constraints model complexity. On the other hand, interpretable models are inherently designed to be transparent, offering advantages such as comprehensibility and trustworthiness. Nevertheless, they may underperform compared to more intricate models and might not be apt for all applications (Mittelstadt, Russell, and Wachteret, 2019). Lastly, post-modeling explainability methods strive to boost transparency after the model has been trained, either through data-driven approaches like artificial neural networks (ANN) or with more human-guided techniques. These types of models, despite providing advantages like flexibility and compatibility with existing models, potentially lack a comprehensive understanding of the model's decision-making process (Minh et al. 2022).

So far, various techniques that facilitate human understanding of complex AI models and their decision-making processes, thus promoting explainability, exist (Dwivedi et al. 2023). For example, among models that inspect the internal mechanism, the decision trees use a tree-like structure to represent decisions and their potential consequences, enabling intuitive visualization and interpretation (Mahbooba et al. 2021). Fuzzy Logic is a form of reasoning that deals with approximate values rather than binary true or false values, allowing for more human-like reasoning and facilitating the interpretation of the AI's logic (Chimatapu et al. 2018). Bayesian Networks are graphical models that represent probabilistic relationships among variables, offering insights into causal dependencies and probabilistic reasoning (Hennessy, Diz and Reiter, 2020).

In addition, Local Interpretable Model-Agnostic Explanations (LIME) is an XAI method that generates post-hoc explanations for individual predictions made by complex models by approximating their behavior with a simpler, interpretable model in the local vicinity of a particular prediction (Upadhyay et al. 2021). Counterfactuals technique provides insights into AI decision-making by exploring alternative scenarios; they help in understanding the model's decisions by answering "what-if" questions, illustrating how the alternation of certain input features could have led to different outcomes (Byrne, 2019). Lastly, SHAP (SHapley Additive exPlanations) assigns importance values to features in machine learning models, indicating their contribution to predictions. Based on cooperative game theory, SHAP calculates Shapley values, enhancing transparency, interpretability, and trustworthiness by revealing the influence of individual features on model outputs (Chromik, 2020).

Collectively, these techniques aim to enhance transparency, interpretability, and trustworthiness of AI systems across various domains and applications. In P&S, XAI concisely justify plan alterations, clarify constraint unsatisfiability or decision superiority, and elaborate on potential consequences. Table 1 illustrates the diverse techniques and methods, showcasing their application to TUPLES use cases (i.e., pilot assistance and aircraft manufacturing).

Metrics for XAI Evaluation

The literature identifies several essential metrics for assessing the explainability of AI systems that can be adapted in P&S. Sovrano and colleagues (2022) categorize these metrics into three classes: model-based, post-hoc, and subject-based. Model-based metrics encompass accuracy, precision, and recall, while post-hoc metrics involve methods like LIME and SHAP. Subject-based metrics deploy human subjects to assess an AI system's explainability through approaches such as surveys and interviews. The authors maintain

Table 1. Potential of XAI techniques in TUPLES use cases.

XAI Technique	TUPLES use case 1: Pilot assistance	TUPLES use case 2: Aircraft manufacturing
Decision Tree	It helps the flight pilot decide where to land the plane in case of an emergency, based on factors such as weather conditions and airport location.	It evaluates worker skills, material needs, and task order, presenting a step-by-step guide to optimally assign resources and workers to tasks.
Fuzzy Logic	It helps the pilot decide when there are ambiguous or conflicting data inputs, such as the severity of the emergency or the status of the plane's systems.	It assists the manager in deciding where to allocate multi-skilled workers for tasks, while addressing uncertain worker skill levels and fluctuating resource constraints.
Bayesian Networks	It helps the pilot decide by weighing the probabilities of different outcomes, such as the likelihood of a successful landing at a particular airport.	It facilitates managers' decisions by estimating success probabilities of tasks, incorporating uncertainties such as worker experience or resource availability.
LIME	It helps the pilot explain its decision-making process to human pilots and passengers, by highlighting which data inputs were most influential in the decision.	It clarifies AI-recommended worker allocations by highlighting influential factors like experience, task urgency, and resource constraints, in aircraft assembly scheduling.
Counterfactual	It helps the pilot learn from previous decisions by exploring what might have happened if it had chosen a different landing site or approach.	It demonstrates alternative workforce allocations to managers by comparing expected outcomes and delays to the AI-recommended decisions.
SHAP	It assists pilots in understanding how various factors, such as aircraft speed and altitude, contribute to aircraft fuel consumption, by assigning each factor a numerical importance score.	It provides managers with numerical scores that prioritize factors like worker skills and resource availability for workforce allocation decisions in tasks such as avionics installation or wing assembly.

that various metrics may complement each other and serve distinct purposes depending on the context (Sovrano et al. 2022).

Numerous key factors and approaches are prevalent in evaluating an AI system's explainability. *Simplicity* refers to the complexity or straightforwardness of the explanations provided by the AI system, with simpler explanations being generally easier to understand (Miller, 2019). *Fidelity* is the accuracy of the explanation in representing the underlying AI model's behavior, with high fidelity explanations accurately describing the model's decision-making process (Velmurugan et al. 2021). *Consistency* measures the stability of explanations when small changes are made to the input or to the model, with consistent AI systems providing similar explanations despite minor alterations (Hoffman, 2018).

Moreover, *transparency* refers to the insight the AI system provides into its internal workings, with more transparent systems offering greater visibility into their decision-making processes (Felzmann et al. 2020). *Local explanations* describe the decision-making process for a specific instance, while *global explanations* provide a broader understanding of the AI system's behavior across multiple instances (Setzu et al. 2021). Lastly, *actionability* measures how useful the explanations are for users to make informed decisions or take actions based on the AI system's output, with actionable explanations guiding users on modifying input features to achieve desired outcomes (Joshi et al. 2019).

The evaluation of AI systems' explainability in the context of P&S requires consideration of a range of metrics. Each metric serves a distinct purpose, and their combination provides a comprehensive assessment of the system's ability to convey its decision-making process effectively (Sovrano et al. 2022) Selecting the appropriate metrics based on the specific use case and user needs is essential to ensure meaningful and practical insights, ultimately fostering trust and collaboration between humans and AI systems.

Benefits of Explainability

Users who understand how an AI system makes decisions are more likely to trust and accept its outputs (Chatila et al. 2021). This fosters confidence in technology and its adoption across various industries. Adadi and Berrada (2018) noted that XAI holds significant potential for enhancing trust and transparency in AI-based systems. This is particularly important in healthcare, finance, and law enforcement, where AI-driven decisions can substantially affect individuals and society. Kotriwala et al. (2021) emphasized that transparent explanations provided by XAI tools are crucial for building human-AI trust in these environments. However, the application of XAI in the process industry presents challenges due to the diverse requirements arising from a wide range of AI end-users and application cases (Kotriwala et al. 2021).

When stakeholders can understand the rationale behind AI-generated insights, they can make better-informed decisions, leading to improved outcomes. Cartovloni and colleagues (2022) discussed the ethical, legal, and social considerations of AI-based medical decision-support tools. They observed that XAI could provide transparent explanations of the AI system's actions and decisions, enabling patients to understand the decision-making process better and assess whether it aligns with ethical and regulatory standards. By providing explanations for algorithmic decisions, XAI systems help users understand the underlying factors influencing the outcomes and enable efficient debugging. This is particularly beneficial in complex fields like P&S, where errors can have significant consequences (Schmid and Wrede, 2022).

Trade-Off Between Accuracy and Explainability

The trade-off between accuracy and explainability in AI presents a significant challenge as it involves balancing a model's predictive performance with its interpretability. Gunning and Aha (2019) identified an inherent tension between AIs' predictive accuracy and explainability. They noted that the most accurate methods are frequently the least explainable, while the most explainable methods often exhibit lower accuracy.

This accuracy-explainability trade-off manifests itself in two distinct categories. Firstly, complex models, such as deep neural networks, are characterized by high accuracy and low explainability and offer impressive predictive accuracy, while their inner workings are often challenging to interpret, making it difficult for users to comprehend the rationale behind their predictions. Secondly, simpler models, including decision trees, exhibit low accuracy but high explainability. These models are more easily understood by humans as they provide a transparent decision-making process and can be visually represented. However, due to their simplicity, they may fail to capture intricate relationships within the data, resulting in lower predictive accuracy than their more complex counterparts (Gunning and Aha, 2019).

Consequently, striking the right balance between accuracy and explainability is crucial for developing AI systems that can effectively support human decision-making while maintaining trust and transparency in their operations.

Human-Centered Approaches

Human-centered approaches to XAI can result in more adaptive and flexible systems that cater to users' diverse needs and preferences. Chromik and Butz (2021) argue that explanation user interfaces should be designed to enable users to adjust the level of detail and complexity of explanations based on their requirements. User's needs and preferences for explanations vary depending on their skills, expertise, and context.

It is important to tailor AI explanations to users, which can enhance the usefulness of explanatory agents for a broader range of users. Miller (2019) discusses the significance of explaining decisions and actions to human observers and how examining human explanations can inform the development of AI explanations. He suggests that explainable AI can benefit from existing models of how people define, generate, select, present, and evaluate explanations. Thus, developers can create more transparent AI algorithms by studying human explanations. Human explanations have several characteristics such as they address why-questions in a contrastive way, are selectively biased, have a social dimension, and causal links are more important than probabilities (Miller, 2019). Incorporating these ideas into AI models can improve explanatory agents, although it may not be feasible for all applications.

Bertrand et al. (2022) conducted a systematic review of 37 papers on cognitive biases in XAI systems, revealing that these biases can affect the transparency and accuracy of AI decision-making. The authors identified several techniques for mitigating the effects of cognitive biases, such as using multiple explanation methods and involving users in the design process. Based on these findings, they constructed a heuristic map to guide the development of future XAI systems that better align with people's cognitive processes (Bertrand et al. 2022). Overall, the study underscores the importance of considering human factors in the development of XAI systems to ensure their effectiveness and acceptance.

Interdisciplinary Approach

Interdisciplinary collaboration is crucial for advancing XAI research, as it brings together expertise from various fields such as data science, human-computer interaction, cognitive psychology, and philosophy (Adadi and Berrada, 2018). Considering human factors during design and development is essential, as cognitive biases in XAI systems can affect transparency and accuracy (Bertrand et al., 2022). Schmid and Wrede (2022) also highlight the significance of interdisciplinary approaches and linguistic expertise in developing dialog modeling approaches for contingent multi-modal interaction, emphasizing the need for explanations to be adaptive to individual user needs and consider human cognitive processes.

Chromik and Butz (2021) explore the interdisciplinary nature of XAI and its aim to enhance human understanding of black-box machine-learning models through explanation-generating methods. They identify "flexibility" as a key design principle for interactive explanation user interfaces in XAI and argue that these interfaces should allow users to adjust the level of detail and complexity of explanations according to their needs and preferences, ultimately facilitating informed decision-making based on their understanding of the AI system's behavior.

CONCLUSION

This paper provided a comprehensive overview of XAI systems in the P&S domain, emphasizing the crucial role of transparency, explainability, and interdisciplinary collaboration in advancing AI research. The review identified specific XAI methods and presented evidence to guide the selection of suitable XAI tools in P&S, uncovering numerous opportunities to enhance system effectiveness and user acceptance.

A key opportunity lies in improving trust and transparency, which are vital elements for the uninterrupted progression of AI systems (Adadi and Berrada, 2018). In domains such as healthcare, where AI-based decisions significantly impact patient treatments, XAI can enable users to make informed decisions by helping them understand the AI system's decision-making processes.

As a result, trust and transparency are increased, leading to more effective decision-making and improved confidence in AI-driven recommendations.

Several recommendations emerge from the findings to ensure the successful implementation of XAI in P&S. First, it is essential to prioritize human-centered design by involving end-users in the development process, ensuring actionable and informative explanations (Chromik and Butz, 2021). Second, interdisciplinary collaboration should be encouraged, leveraging expertise from fields like data science, cognitive psychology, and philosophy (Schmid & Wrede, 2022). Third, it is crucial to promote the identification and standardization of evaluation metrics for XAI in P&S systems to objectively assess their impact. Lastly, research should focus on mitigating cognitive biases in XAI systems to enhance transparency and accuracy (Chromik and Butz, 2021).

Future research should explore adaptive XAI systems and assess the impact of XAI on user trust and adoption across various P&S scenarios. By aligning AI systems with human cognitive processes and tailoring explanations to users, trust can be fostered, decision-making improved, and ethical concerns addressed. An approach that aligns AI systems with human cognitive processes and tailors explanations to users will ultimately pave the way for the widespread adoption of AI in the P&S domain.

ACKNOWLEDGMENT

This work has received funding from the European Union's Horizon Europe Research and Innovation program under the grant agreement TUPLES N° 101070149.

REFERENCES

Adadi, A., Berrada, M. (2018) "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)". IEEE access, Volume 6, pp. 52138–52160.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, Suh, J., Iqbal, et al. (2019) "Guidelines for human-AI interaction", Proceedings of the CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK.

Bertrand, A., Belloum, R., Eagan, J. R. and Maxwell, W. (2022) "How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review", Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES'22), Oxford, UK.

Bunt, A., Lount, M., and Lauzon, C. (2012) "Are explanations always important? A study of deployed, low-cost intelligent interactive systems", Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, Lisbon, Portugal.

Byrne, R. M. (2019) "Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning", Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), Macao, PRC.

Čartolovni, A., Tomičić, A., and Mosler, E. L. (2022) "Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review", International Journal of Medical Informatics, Volume 161, no. 104738.

Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., and Yeung, K., eds. (2021). Trustworthy AI, Reflections on Artificial Intelligence for Humanity, Lecture Notes in Computer Science. Springer, Cham. pp. 13–39.

Chimatapu, R., Hagras, H., Starkey, A., and Owusu, G. (2018), "Explainable AI and fuzzy logic systems", Proceeding of Theory and Practice of Natural Computing: 7th International Conference, Dublin, Ireland, Springer International Publishing, pp. 3–20.

Chromik, M. (2020) "Reshape: A framework for interactive explanations in xai based on shap", Proceedings of 18th European Conference on Computer-Supported Cooperative Work, Siegen, Germany, European Society for Socially Embedded Technologies, Volume 4, no. 2, pp 1–5.

Chromik, M., Butz, A. (2021) "Human-xai interaction: A review and design principles for explanation user interfaces", in Proceedings of 13° International Conference Human-Computer Interaction, Bari, Italy, Springer International Publishing, pp. 279–298.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Quian, B., Wen, Z., et al. (2023) "Explainable AI (XAI): Core ideas, techniques, and solutions", ACM Computing Surveys, Volume 55, no. 9, pp. 1–33.

European Commission, High-level expert group on artificial intelligence (2021). Deliverable 1: Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., and Tamò-Larrieux, A. (2020) "Towards transparency by design for artificial intelligence", Science and Engineering Ethics, Volume 26, no. 6, pp. 3333–3361.

Gunning, D., and Aha, D. W. (2019) "DARPA's explainable artificial intelligence (XAI) program", AI Magazine, Volume 40, no. 2, pp. 44–58.

Hennessy, C., Diz, A. B. and Reiter, E. (2020) "Explaining Bayesian Networks in natural language: state of the art and challenges" 2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence, pp. 28–33.

Hoffman, R. R., Miller, T., Mueller, S. T., Klein, G., and Clancey, W. J. (2018) "Explaining explanation, Part 4: A deep dive on deep nets", IEEE Intelligent Systems, Volume 33, no. 3, pp. 87–95.

Joshi, S., Koyejo, O., Vijitbenjaronk, W., Kim B. and Ghosh, J. (2019) "Towards realistic individual recourse and actionable explanations in black-box decision making systems" arXiv:1907.09615: https://arxiv.org/abs/1907.09615

Kotriwala, A., Klöpper, B., Dix, M., Gopalakrishnan, G., Ziobro, D. and Potschka, A. (2021) "XAI for Operations in the Process Industry-Applications, Theses, and Research Directions", Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering, California, USA, Volume 2845, no. 26.

Mahbooba, B., Timilsina, M., Sahal, R. and Serrano, M. (2021) "Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model" Complexity, pp. 1–11.

Miller, T. (2019) "Explanation in artificial intelligence: Insights from the social sciences", Artificial Intelligence Volume 267, pp. 1–38.

Minh, D., Wang, H. X., Li, Y. F. and Nguyen, T. N. (2022) "Explainable artificial intelligence: a comprehensive review" Artificial Intelligence Review, Volume 55, pp. 1–66.

Mittelstadt, B., Russell, C. and Wachter, S. (2019) "Explaining explanations in AI' in Proceedings Conference on Fairness, Accountability, and Transparency", Atlanta, GA, USA. ACM, New York: NY, USA, pp. 279–288.

Ribeiro, M. T., Singh, S., and Guestrin, C. "Why should I trust you? Explaining the predictions of any classifier" Proceedings of the CHI Conference on Human Factors in Computing Systems, Yokohama Japan, pp. 1135–1144.

Setzu, M., Guidotti, R., Monreale, A., Turini, F., Pedreschi, D., and Giannotti, F. (2021) "Glocalx-from local to global explanations of black box ai models" Artificial Intelligence, Volume. 294, no. 103457.

Schmid, U. and Wrede, B. (2022) "What is Missing in XAI So Far? An Interdisciplinary Perspective" KI-Künstliche Intelligenz, pp. 1–13.

Sovrano, F., Sapienza, S., Palmirani, M., and Vitali, F. (2022) "Metrics, explainability and the European AI act proposal" Multidisciplinary Scientific Journal, Volume 5, no. 1, pp. 126–138.

TUPLES 2023, "Building trustworthy planning and scheduling systems", TUPLES Project Website: https://tuples.ai.

Upadhyay, S., Isahagian, V., Muthusamy, V., and Rizk, Y. (2021) "Extending LIME for business process automation", arXiv:2108.04371v: https://arxiv.org/abs/2108.04371

Velmurugan, M., Ouyang, C., Moreira, C., and Sindhgatta, R. (2021) "Evaluating fidelity of explainable methods for predictive process analytics" Proceedings of Intelligent Information Systems: CAiSE Forum 2021, Melbourne, Australia, Cham: Springer International Publishing, pp. 64–72.