Explainability Requirements Engineering in AI-enabled Systems

Lívia Mancine[1,2][0000−0002−4123−4861], Renata Dutra Braga[2][0000−0002−3448−0343], and Renato F. Bulcão-Neto[2][0000−0001−8604−0019]

[1] Instituto Federal Goiano (IF Goiano). Ceres - GO, Brazil
livia.mancine@ifgoiano.edu.br
[2] Instituto de Informática, Universidade Federal de Goiás. Goiânia - GO, Brazil
{renatadbraga,rbulcao}@ufg.br

Abstract. Explainability is the ability to make the behavior and decisions of AI-enabled Systems understandable to stakeholders and has become essential for trust, transparency, and informed decision-making across application domains. However, Requirements Engineering (RE) still lacks systematic approaches for integrating explainability throughout the development lifecycle in a structured and stakeholder-oriented manner. This doctoral research proposes *OpenUPExp*, a process-oriented RE framework that extends the Open Unified Process by introducing roles, activities, and artifacts to support explainability by design as a non-functional requirement. The research follows the Design Science Research methodology and evaluates the framework through illustrative scenarios and real-world case studies involving AI-enabled Systems. The main contribution is *OpenUPExp*, a process-oriented RE approach for the systematic treatment of explainability from conception through elicitation, analysis, and specification. The framework aims to support stakeholder-centered explainability, improve the traceability and documentation of explainability requirements, and facilitate their alignment with different stakeholder needs and contexts.

Keywords: Requirements Engineering · Explainability · AI-enabled Systems · OpenUP · Design Science Research.

1 Introduction

The increasing use of AI-enabled Systems in software across different types and domains has driven discussions on trust, transparency, and responsibility in the results produced by these systems [6]. This phenomenon is driven by the increasing volume of available data and advances in computing power, which enable the integration of intelligent components into traditional systems in various contexts, such as healthcare [13] and finance [7]. As these systems increasingly support or automate decisions across multiple domains, it becomes essential to understand *why* and *how* these decisions are made.

In this context, explainability emerges as an essential non-functional requirement (NFR) for the quality of AI-enabled Systems [2]. It refers to the ability

Proceedings of the 29th Workshop on Requirements Engineering (WER2026), August 11-13, 2026, La Plata, Argentina.
https://doi.org/10.29327/1884517.29-25

of a system to provide understandable explanations about its behavior and decisions to different stakeholders [10]. Although often associated with high-risk domains, its relevance extends to any system that includes intelligent components, supporting users, managers, and developers in understanding, justifying, and auditing automated decisions [18].

Despite its cross-cutting relevance, explainability still lacks systematic treatment in the field of Requirements Engineering (RE). Recent studies show that elicitation has been the most explored RE activity in this context, while analysis, specification, validation, and management remain underexplored [15]. This gap is critical because, in RE, late changes to requirements tend to increase costs and rework [20]. In AI-enabled Systems, introducing explainability late can directly affect decisions related to model selection, trade-offs between interpretability and performance, and system architecture design.

Given this context, this doctoral research investigates the following problem: *the lack of systematic and integrated RE approaches to treat explainability as a NFR from the early stages of AI-enabled Systems, considering different domains and stakeholder profiles*. To address this problem, a process-oriented RE approach is proposed to support explainability in AI-enabled Systems, enabling its integration from the early stages and covering, in an integrated way, the activities of elicitation, analysis, and specification of explainability requirements.

The proposed approach is developed following the Design Science Research (DSR) methodology. The main resulting artifact is *OpenUPExp*, an extension of the Open Unified Process (OpenUP) focused on the systematic treatment of explainability. Preliminary results indicate the feasibility of the proposal, including the identification of an initial set of elements for the explainability vision, the initial definition of the framework's conceptual architecture, and its partial implementation in a computational environment. These findings provide initial evidence that the systematic treatment of explainability in RE can contribute to more transparent, understandable systems aligned with the expectations of different stakeholders.

This paper is organized as follows: Section 2 presents the research objectives and research question; Section 3 discusses related work; Section 4 describes the research method; Section 5 presents the preliminary results; and finally, Section 6 presents the conclusion.

2 Research Questions and Objectives

Despite recent advances in RE research for AI-enabled Systems, the systematic treatment of explainability remains limited, and most proposed approaches are still at a conceptual or theoretical level, with limited validation in real-world contexts [15]. Given these gaps, especially in integrating elicitation, analysis, and specification activities throughout the development lifecycle, this research aims to answer the following question: "*How can a process-oriented RE approach support the systematic treatment of explainability in elicitation, analysis, and specification activities in AI-enabled Systems?*"

Thus, the main objective of this doctoral research is to *develop and evaluate a process-oriented RE approach for treating explainability requirements in AI-enabled Systems from the early stages, covering elicitation, analysis, and specification activities.* To achieve this objective, the following specific objectives are defined:

1. map the state of the art on explainability, RE, and AI-enabled Systems [15];
2. investigate challenges and needs related to the treatment of explainability in AI-enabled Systems with specialists;
3. define mechanisms to incorporate explainability from the early stages in AI-enabled Systems;
4. define a technique for eliciting explainability requirements oriented to different stakeholder profiles;
5. structure a technique for analyzing explainability requirements, considering context, users, and decision goals;
6. establish a technique for specifying explainability requirements oriented to different stakeholder needs;
7. develop OpenUPExp as a process-oriented approach to support the systematic treatment of explainability;
8. evaluate the proposed approach through illustrative scenarios and real-world applications.

3 Related Work

Studies have investigated explainability in AI-enabled Systems from a RE perspective [15]. Li and Han [14] propose a framework for analyzing explainability requirements in Machine Learning (ML) systems using *iStar* to recommend Explainable AI (XAI) methods. Similarly, Barrera et al. [1] employ *iStar* as a metamodel for ML systems to support the identification of NFRs, including explainability, and the selection of suitable algorithms. Both approaches focus on technical explainability and ML model analysis.

Other studies emphasize stakeholder concerns and user-centered explanations. Habiba et al. [9] propose a framework centered on user needs for explainability requirements, while Schoonderwoerd et al. [19] introduce the DoReMi approach, which explores explainability requirements and user interface design for clinical decision-support systems. Their results highlight the importance of adapting explanations to domain experts and end users.

From another perspective, Guizzardi et al. [8] propose an Ontology-Based RE (ObRE) approach for eliciting and analyzing ethical requirements, including explainability and autonomy. Crook et al. [3] discuss trade-offs among explainability, performance, and development constraints, proposing a framework to support RE decisions regarding ML models. More recently, Quah et al. [17] proposed a framework for analyzing personalized explainability requirements based on stakeholder and domain characteristics, highlighting the need to tailor explanations to different users. However, the approach focuses mainly on requirements analysis and does not address integration into RE processes.

In addition, Chazette et al. [2] discuss explainability as a NFR and provide conceptual foundations and recommendations for incorporating explainability into software systems. While their work establishes important theoretical foundations, it is not specifically focused on AI-enabled Systems nor on the operationalization of explainability through a process-oriented RE approach.

Given this scenario, a gap can be identified in the literature regarding process-oriented approaches that address explainability throughout the development lifecycle of AI-enabled systems. Existing studies typically focus on isolated RE activities or specific concerns, such as XAI method selection, ethical requirements, or personalized explanations, providing limited support for integrating conception, elicitation, analysis, and specification. To address this gap, *OpenUPExp* incorporates Explainability by Design (EbD) principles into a structured process framework, defining roles, tasks, artifacts, and workflows to support these activities. The approach is currently being evaluated through illustrative scenarios and real-world contexts.

4 Research Methodology

This research adopts the DSR paradigm, widely used in Software Engineering to guide investigations focused on the construction and evaluation of artifacts [12,5]. The goal of DSR is to produce prescriptive knowledge through the development of innovative solutions to relevant problems, combining scientific rigor with practical applicability [12]. Figure 1 presents an overview of the research method, structured based on the model proposed by [21]. The research is organized into three main stages: problem understanding, solution design, and artifact evaluation, integrating theoretical foundations, exploratory empirical evidence, and systematic evaluation in terms of relevance, rigor, and innovation.

The problem understanding stage aims to investigate how explainability is treated as a NFR in AI-enabled Systems. It is supported by two systematic literature reviews [16,15], the identification of explainability requirements in different intelligent system contexts, and an exploratory empirical investigation conducted with professionals involved in the development of these systems. This investigation was carried out through semi-structured interviews with specialists affiliated with the Center of Excellence in AI at the Federal University of Goiás (CEIA/UFG), enabling the identification of challenges, needs, and practices related to explainability in RE. The study was approved by the Research Ethics Committee of UFG under approval number 7.416.401 (CAAE 85638924.3.0000.5083).

Based on this understanding, the solution design was structured, supporting the development of *OpenUPExp*, which is derived from OpenUP. OpenUP was selected as the foundation process because its iterative and incremental nature aligns with the progressive refinement of explainability requirements. In addition, its explicit definition of roles, tasks, and artifacts provides a suitable structure for incorporating explainability into RE activities. Its lightweight and

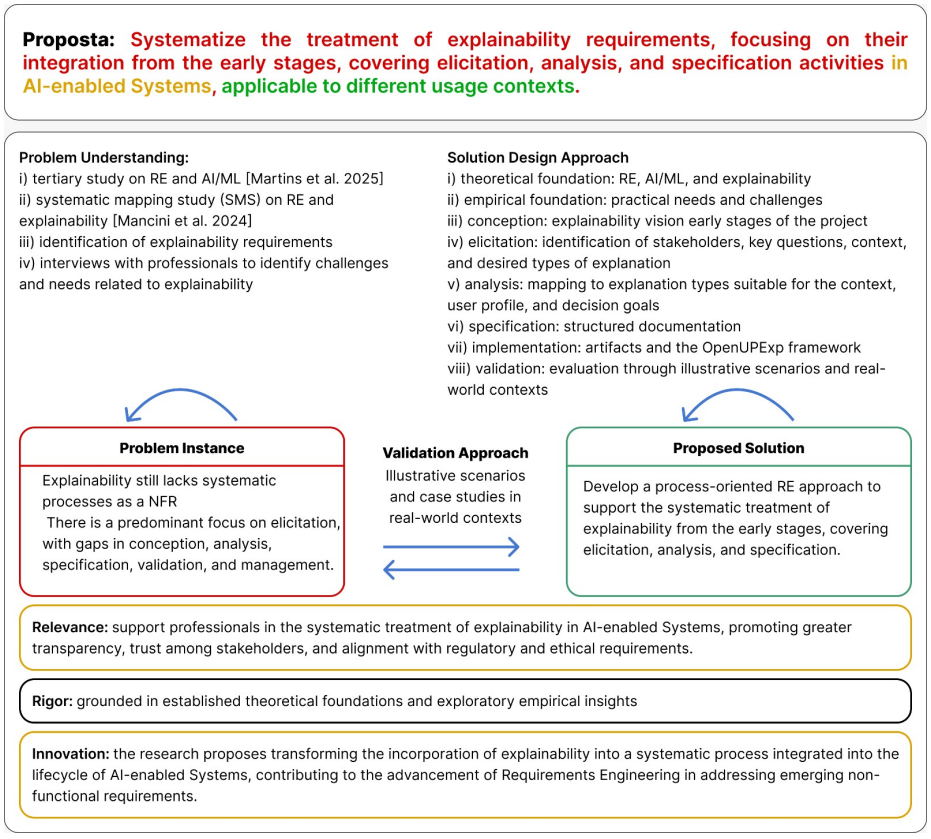

Fig. 1. Overview of the methodolody.

extensible nature enables the systematic integration of explainability concerns without introducing excessive process complexity.

The proposal integrates theoretical and empirical foundations to structure the treatment of explainability from conception through elicitation, analysis, and specification activities. During conception, the explainability vision is defined. In the elaboration phase, elicitation identifies stakeholders, key questions, context, and desired explanation types; analysis maps requirements to appropriate explanations according to user profiles and decision goals; and specification documents these requirements in a structured way. The implementation stage involves the development of artifacts within the *OpenUPExp* framework, promoting traceability and consistency throughout the system lifecycle.

The proposal will be validated through two complementary approaches: (i) illustrative scenarios for conceptual evaluation and incremental artifact refinement, and (ii) case studies in real-world AI-enabled Systems to assess usefulness, applicability, and feasibility. The evaluation will use constructs adapted from the Technology Acceptance Model (TAM) [4], covering Job Relevance, Output Qual-

ity, Result Demonstrability, and Anticipated Satisfaction. Additionally, NASA-TLX [11] will be employed to measure perceived cognitive load, enabling the assessment of both usefulness and cognitive effort associated with the approach. The quality of the generated explainability requirements will also be assessed using adapted ISO/IEC/IEEE 29148 criteria, considering completeness, correctness, understandability, and verifiability. As an expected result, *OpenUPExp* is the central artifact of this research. By integrating theoretical foundations, exploratory empirical insights, and evaluation in real-world contexts, the methodology seeks to ensure scientific rigor and practical relevance, contributing to the advancement of RE for AI-enabled Systems.

5 Preliminary Results

This research started in 2023 with a tertiary study on RE for AI/ML-based systems [16], which identified explainability as an underexplored requirement, particularly regarding its systematic integration into RE. The study also highlighted that different stakeholders require different levels of explanation, emphasizing the need for structured mechanisms to define what should be explained, to whom, and in which context.

Motivated by these findings, a Systematic Mapping Study (SMS) on RE and explainability in ML was conducted [15]. The results indicate that explainability is treated as an emerging NFR, with most studies focusing on elicitation and analysis, while specification, validation, and management remain less explored. The scarcity of applied empirical studies further reinforces the need for approaches validated in real-world contexts. To complement the literature findings, semi-structured interviews were conducted with AI specialists from different domains. The results indicate that explainability is still treated in an ad hoc manner, mainly during elicitation, with limited support for analysis and specification activities. Participants also reported difficulties in defining appropriate explanation levels for different stakeholders and a lack of structured artifacts to support documentation.

Based on these findings, the initial conceptual structure of *OpenUPExp* was defined. The framework extends OpenUP by incorporating explainability into the conception and elaboration phases, covering vision, elicitation, analysis, and specification activities. As a preliminary result, the conception phase process was modeled using BPMN, representing the activities, roles, and artifacts associated with the Explainability Vision. Figures 2 and 3 present an overview of the framework and the corresponding BPMN process.

In addition to modeling, an implementation of the Explainability Vision artifact was developed in a computational environment, using the LangFlow platform as infrastructure for prototyping the workers responsible for operationalizing this stage. This implementation enables the structuring and automation of interactions among the components of the approach, supporting its practical application and evaluation.

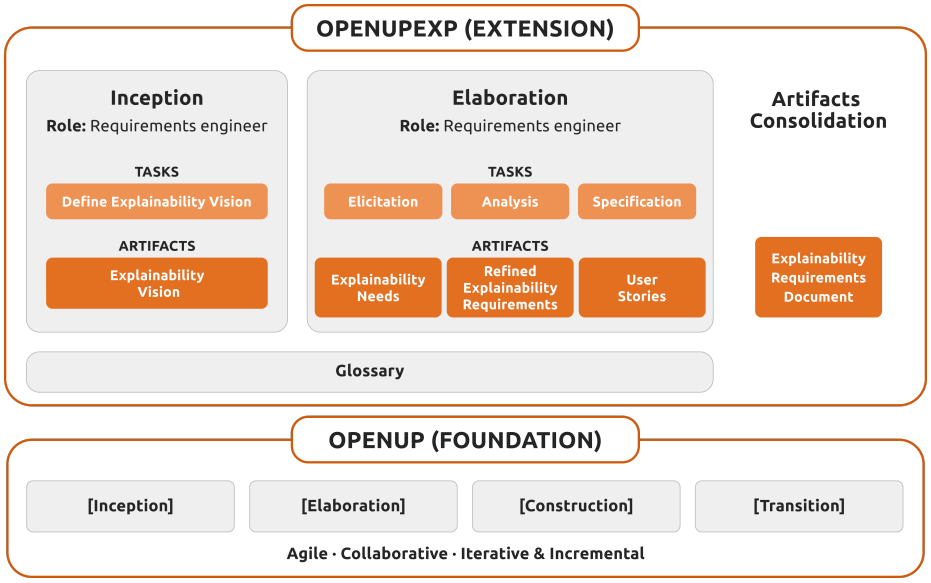

Fig. 2. Overview of the extended OpenUPExp and its main features.

As an initial evaluation, the Explainability Vision artifact was analyzed through an illustrative scenario with two specialists. The participants provided feedback using a questionnaire adapted from TAM, covering four dimensions: Job Relevance, Output Quality, Result Demonstrability, and Anticipated Satisfaction, using a five-point Likert scale. In addition, an open-ended question was included to collect qualitative observations.

The results suggest that *OpenUPExp* is perceived as a promising approach with potential for practical application, although it still requires refinements to improve its maturity, especially regarding communicability, technical detail, and validation robustness. These findings provide preliminary indications that the framework may support the systematic structuring of explainability requirements, contributing to integrating explainability as a central concern in RE for AI-enabled Systems. However, these results are limited to an initial assessment involving two specialists and cannot yet be generalized. The findings have been consolidated in a scientific paper that has been accepted for publication and is currently undergoing final revisions prior to publication.

At the current stage, the evaluation strategy has also been defined, structured into two complementary approaches: (i) illustrative scenarios, for the incremental validation of artifact fragments, and (ii) case studies in real-world contexts, for analyzing practical applicability. Both approaches will be evaluated using constructs adapted from TAM and NASA-TLX. As next steps, the research aims to complete the full implementation of *OpenUPExp*, conduct the evaluation through illustrative scenarios in the elaboration phase, carry out case studies in real-world environments, and consolidate the empirical results to support the

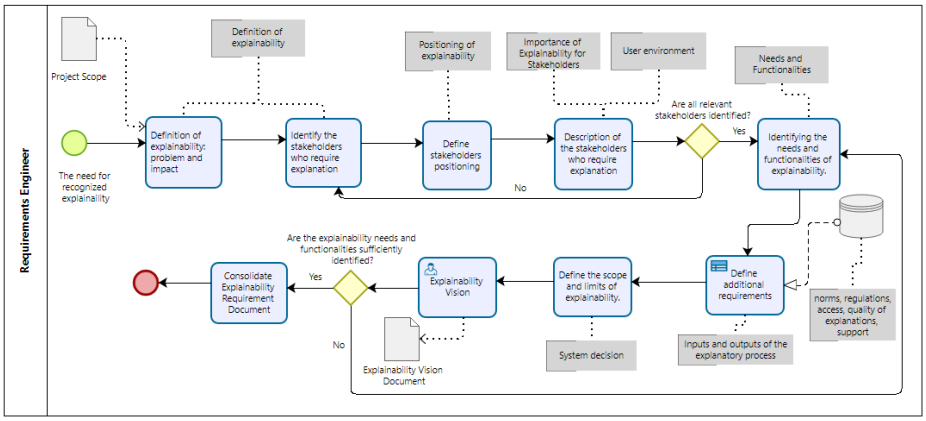

Fig. 3. Explainability Vision Process Modeling using BPMN

final refinement of the framework and validation of the thesis. The following schedule presents the remaining activities for completing the research.

Table 1. Schedule for the Next Research Phases

Stage	**Start Date**	**End Date**
Solution Proposal – Elaboration Phase (Process)	04/2026	05/2026
Framework Implementation	02/2026	08/2026
Framework Evaluation – Illustrative Scenarios	03/2026	08/2026
Analysis of Illustrative Scenario Results	04/2026	10/2026
Framework Evaluation – Real-World Case Study	10/2026	01/2027
Analysis of Evaluation Results	02/2027	03/2027
Thesis Writing	06/2026	05/2027
Thesis Defense	06/2027	06/2027

6 Conclusion

This paper presented ongoing doctoral research aimed at developing a process-oriented RE approach for the systematic treatment of explainability in AI-enabled Systems, from the early stages through elicitation, analysis, and specification activities. The proposal is materialized in OpenUPExp, an extension of OpenUP that integrates explainability as a NFR into the development lifecycle, supported by structured artifacts and mechanisms that promote collaboration among stakeholders.

As part of the progress achieved, systematic literature studies were conducted to identify gaps in the treatment of explainability in RE, along with an empirical

investigation with specialists. In addition, preliminary results were obtained in the modeling of the conception phase process, the definition of the conceptual architecture of OpenUPExp, and the initial implementation of the Explainability Vision artifact using LangFlow.

The next steps focus on completing the elaboration phase of *OpenUPExp* and the full implementation of the framework, including elicitation, analysis, and specification activities for explainability requirements. Future work also includes the consolidation of explainability requirement quality criteria, based on ISO/IEC/IEEE 29148:2018, into the framework checklists. In parallel, case studies will be conducted in real-world contexts to evaluate the applicability of *OpenUPExp* across different domains and development teams. These activities will contribute to the progressive refinement of the framework and provide further insights into its ability to support the systematic treatment of explainability and align system behavior with stakeholder needs and expectations.

Key challenges include adapting explainability to different stakeholder profiles, validating the approach in diverse real-world contexts, and promoting its adoption by industry practitioners. To address these challenges, the research follows an incremental evaluation strategy that combines continuous feedback from specialists with iterative artifact refinement. Ultimately, the expected outcome is a process-oriented RE framework that supports the systematic treatment of explainability and contributes to the development of more transparent and stakeholder-aligned AI-enabled Systems.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Barrera, J.M., et al.: An extension of istar for machine learning requirements by following the prise methodology. Computer Standards & Interfaces **88**, 103806 (2024). https://doi.org/10.1016/j.csi.2023.103806
2. Chazette, L., et al.: Explainable software systems: from requirements analysis to system evaluation. Requirements Engineering **27**(4), 457–487 (2022). https://doi.org/10.1007/s00766-022-00393-5
3. Crook, B., et al.: Revisiting the performance-explainability trade-off in explainable artificial intelligence (xai). In: IEEE 31st International Requirements Engineering Conference Workshops. pp. 316–324. IEEE (2023). https://doi.org/10.1109/REW57809.2023.00060
4. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems: Theory and results. Ph.D. thesis, Massachusetts Institute of Technology (1985)
5. Engström, E., et al.: How software engineering research aligns with design science: a review. Empirical Software Engineering **25**, 2630–2660 (2020). https://doi.org/10.1007/s10664-020-09818-7
6. Giray, G.: A software engineering perspective on engineering machine learning systems: State of the art and challenges. Journal of Systems and Software **180**, 111031 (2021). https://doi.org/10.1016/j.jss.2021.111031

7. Goodell, J.W., et al.: Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance **32**, 100577 (2021). https://doi.org/10.1016/j.jbef.2021.100577
8. Guizzardi, R., et al.: Eliciting ethicality requirements using the ontology-based requirements engineering method. In: International Conference on Business Process Modeling, Development and Support. pp. 221–236. Springer (2022). https://doi.org/10.1007/978-3-031-07475-2_15
9. Habiba, U.E., et al.: Can requirements engineering support explainable artificial intelligence? towards a user-centric approach for explainability requirements. In: IEEE 30th International Requirements Engineering Conference Workshops. pp. 162–165. IEEE (2022). https://doi.org/10.1109/REW56159.2022.00038
10. Habibullah, K.M., et al.: Explainable ai: A diverse stakeholder perspective. In: 2024 IEEE 32nd International Requirements Engineering Conference. pp. 494–495. IEEE (2024). https://doi.org/10.1109/RE59067.2024.00060
11. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Advances in psychology, vol. 52, pp. 139–183. Elsevier (1988)
12. Hevner, A.R., et al.: Design science in information systems research. MIS quarterly **28**(1), 75–106 (2004). https://doi.org/10.2307/25148625
13. Jiang, F., et al.: Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology **2**(4) (2017). https://doi.org/10.1136/svn-2017-000101
14. Li, T., Han, L.: Dealing with explainability requirements for machine learning systems. In: IEEE 47th Annual Computers, Software, and Applications Conference. pp. 1203–1208. IEEE (2023). https://doi.org/10.1109/COMPSAC57700.2023.00182
15. Mancine, L., et al.: Estado da arte sobre engenharia de requisitos e explicabilidade em sistemas baseados em aprendizado de máquina. In: Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web. pp. 143–158. SBC, Porto Alegre, RS, Brasil (2024). https://doi.org/10.5753/webmedia_estendido.2024.243944
16. Martins, M.C., et al.: Requirements engineering for machine learning-based ai systems: A tertiary study. Journal of Software Engineering Research and Development (2025). https://doi.org/10.5753/jserd.2025.4892
17. Quah, J.K., et al.: Personalized explainability requirements analysis framework for ai-enabled systems. Malaysian Journal of Computer Science **38**(1), 55–80 (2025). https://doi.org/10.22452/mjcs.vol38no1.3
18. Rosenfeld, A., Richardson, A.: Explainability in human–agent systems. Autonomous agents and multi-agent systems **33**(6), 673–705 (2019). https://doi.org/10.1007/s10458-019-09408-y
19. Schoonderwoerd, T.A.J., et al.: Human-centered xai: Developing design patterns for explanations of clinical decision support systems. International Journal of Human-Computer Studies **154**, 102684 (2021). https://doi.org/10.1016/j.ijhcs.2021.102684
20. Sommerville, I.: Software engineering 9th edition. ISBN-10 **137035152**, 18 (2011)
21. Storey, M.A., et al.: Using a visual abstract as a lens for communicating and promoting design science research in software engineering. In: ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. pp. 181–186. IEEE (2017). https://doi.org/10.1109/ESEM.2017.28