

Engenharia de Requisitos de Explicabilidade em Sistemas Baseados em Aprendizado de Máquina

Lívia Mancine¹²

¹Instituto de Informática – Universidade Federal de Goiás (INF/UFG)
Goiânia – GO – Brazil

²Instituto Federal Goiano (IF Goiano)
Goiânia – GO – Brazil

livia.mancine@ifgoiano.edu.br

Abstract. *Explainability is the ability to make the predictions of a Machine Learning (ML) model clear and understandable to people. The adoption of ML models in critical decision-making requires explainability to be incorporated as a fundamental requirement in the decision-making process. As a recent topic, explainability in ML-based systems still lacks systematized processes, which opens space for further research. This PhD work aims to develop and contribute with a Requirements Engineering approach for the analysis and specification of explainability requirements in ML-based systems. The Design Science methodology will serve as the model to guide this research.*

Resumo. *A explicabilidade é a habilidade de tornar as previsões de um modelo de Aprendizado de Máquina (AM) claras e compreensíveis para as pessoas. A adoção de modelos de AM em decisões críticas exige que a explicabilidade seja incorporada como um requisito fundamental no processo de tomada de decisão. Por ser um tema recente, a explicabilidade em sistemas baseados em AM ainda carece de processos sistematizados, o que abre espaço para novas pesquisas. Este trabalho de doutorado visa desenvolver e contribuir com uma abordagem de Engenharia de Requisitos para análise e especificação de requisitos de explicabilidade em sistemas baseados em AM. A metodologia Design Science servirá de modelo para conduzir esta pesquisa.*

1. Introdução

O uso emergente de componentes de Inteligência Artificial (IA) em diversos segmentos de software levanta preocupações sobre a confiabilidade dos resultados gerados pela IA. Além disso, surgem dúvidas sobre a responsabilidade em casos de resultados indesejados, bem como a necessidade de compreender os motivos que levaram a resultados bem-sucedidos. O uso intensivo de componentes de IA, em especial de Aprendizado de Máquina (AM), integrados a software tradicional é impulsionada por fatores, como a grande quantidade de dados gerados e o crescente poder computacional [Giray 2021]. Exemplos desses sistemas podem ser observados em domínios como saúde [Jiang et al. 2017] e finanças [Goodell et al. 2021]. Tais domínios, considerados críticos na tomada de decisão, necessitam de explicações dos sistemas baseados em IA nos quais estão envolvidos, pois esses sistemas são classificados como de alto risco devido aos impactos adversos sobre pessoas ou grupos afetados [Senado Federal do Brasil 2023].

Especialmente na área de saúde, os sistemas de AM enfrentam desafios, como vieses em dados clínicos e a opacidade dos modelos [Khattak et al. 2024, Lin et al. 2024]. A confiança, essencial em sistemas críticos [Khattak et al. 2024], é impactada por esses fatores, dificultando a construção de relações seguras entre *stakeholders* e limitando sua adoção prática [Vayena et al. 2018]. Nesse contexto, a transparência torna-se fundamental para fornecer explicações claras e acessíveis sobre as decisões dos algoritmos, fortalecendo a confiança e viabilizando sua adoção [Vayena et al. 2018, Habibullah et al. 2024].

O requisito de explicabilidade tem se tornado uma preocupação emergente em sistemas baseados em AM [Mancine et al. 2024], sendo considerado um requisito não funcional (RNF), que impacta substancialmente a qualidade desses sistemas [Chazette et al. 2022, Umm-E-Habiba 2023]. Assim, aspectos como confiança e transparência devem ser abordados desde as etapas iniciais do ciclo de vida de sistemas de IA/AM [Baldassarre et al. 2024].

Embora a explicabilidade em sistemas baseados em AM tenha despertado interesse crescente na comunidade de ER, *esse recente tema ainda carece de processos sistematizados*. Investigar a sistematização de atividades de processos de ER para abordar explicabilidade torna-se relevante para diferentes *stakeholders*, como especialistas de domínio, usuários finais, desenvolvedores de sistemas e entidades legais, trazendo-lhes explicações adaptadas sobre os resultados dos modelos de AM e, assim, promovendo mais confiança e transparência [Habibullah et al. 2024]. Considere como exemplo um sistema de suporte à decisão clínica baseado em AM, como no diagnóstico de uma doença. Médicos precisam entender as recomendações para avaliar diagnósticos; pacientes necessitam de informações claras para tomar decisões sobre tratamentos; desenvolvedores requerem detalhes técnicos para aprimorar modelos; e entidades regulatórias demandam transparência para garantir conformidade legal. O direito à explicação, previsto no PL 2338/2023 [Senado Federal do Brasil 2023], reforça a importância de explicações acessíveis, especialmente em áreas críticas como saúde, finanças e direito.

Com base nos desafios identificados, o problema de pesquisa a ser investigado nesta proposta de doutorado é *a carência de abordagens de ER que tratem explicabilidade como RNF fundamental, desde a etapa de concepção do sistemas baseados em AM*. Nesse cenário, a atividade de ER mais investigada tem sido a elicitación, *com poucos estudos sobre análise, especificação, validação e gerenciamento* [Mancine et al. 2024]. Para esta proposta de doutorado, será elaborada *uma abordagem de ER para as atividades de análise e especificação de explicabilidade para sistemas baseados em AM*.

Este artigo está assim organizado: a Seção 2 apresenta os objetivos e a questão de pesquisa. A Seção 3 descreve os trabalhos relacionados. A Seção 4 relata o método de pesquisa. A Seção 5 apresenta os resultados preliminares. Por fim, a Seção 6 conclui o artigo e descreve as próximas etapas da pesquisa.

2. Questão de Pesquisa e Objetivos

Apesar das pesquisas sobre ER para AM terem crescido recentemente, a especificação de requisitos de AM de forma compreensível para *stakeholders* é pouco explorada, embora seja considerada a mais desafiadora do desenvolvimento de sistemas baseados em AM [Villamizar et al. 2024]. Um estudo sobre ER e explicabilidade em AM [Mancine et al. 2024] apontou que a ER precisa ser mais investigada em todas as suas

atividades. Os autores destacaram que a fase de elicitação foi a mais abordada, seguida da análise, ressaltando a necessidade de aprofundar a especificação. Além disso, a maioria dos estudos se concentra em abordagens teóricas, sem validação em ambientes reais.

Dado o contexto do problema apresentado, este trabalho visa responder a seguinte questão de pesquisa: “*O quanto uma abordagem de ER auxilia na análise e especificação do RNF de explicabilidade em sistemas baseados em AM?*” Assim, o objetivo geral desta pesquisa de doutorado é *desenvolver uma abordagem de ER que contemple as atividades de análise e especificação do RNF de explicabilidade em sistemas baseados em AM.*

Para alcançar esse objetivo geral, definem-se como objetivos específicos:

1. mapear o estado da arte sobre explicabilidade, ER e sistemas baseados em AM [Mancine et al. 2024];
2. coletar pontos de vista de desenvolvedores de sistemas baseados em AM sobre tratamento de explicabilidade;
3. definir técnica de análise de requisitos de explicabilidade;
4. definir técnica de especificação de requisitos de explicabilidade;
5. experimentar a abordagem de análise e especificação elaborada em projetos de sistemas baseados em AM que demandam explicabilidade.

3. Trabalhos Relacionados

Na pesquisa de [Li and Han 2023], é proposta uma estrutura para análise de requisitos de explicabilidade em sistemas de AM, usando *iStar* para modelar um *framework* de explicabilidade. A abordagem baseada em objetivos visa recomendar métodos XAI a desenvolvedores de forma sistemática e automática. A usabilidade do *framework* foi avaliada entre desenvolvedores com diferentes níveis de conhecimento de explicabilidade. O trabalho de [Barrera et al. 2024] utiliza *iStar* como metamodelo para sistemas de AM, permitindo identificar RNF, entre eles, explicabilidade, selecionar algoritmos adequados e analisar fontes de dados alinhadas aos objetivos do projeto. Os autores realizaram um estudo de caso na indústria. Entretanto, os estudos de [Li and Han 2023] e [Barrera et al. 2024] focaram apenas em métodos XAI que fornecem explicações técnicas, sem considerar outros aspectos de explicabilidade para diferentes *stakeholders*.

Com uma abordagem diferente [Guizzardi et al. 2022], a Engenharia de Requisitos Baseada em Ontologia (ObRE) foi proposta para elicitação e análise de requisitos éticos, com foco em explicabilidade e autonomia. A ontologia foi validada quanto à conformidade com os objetivos da União Europeia para sistemas éticos, mas não houve validação por estudos de caso reais, nem avaliação com especialistas.

A proposta deste artigo visa explorar as lacunas de pesquisa existentes quanto ao requisito de explicabilidade. Há uma oportunidade de desenvolver abordagens que contemple a análise e a especificação da explicabilidade para sistemas baseados em AM desde a concepção do sistema, considerando as necessidades dos diferentes *stakeholders* e validando a proposta por meio de estudos de caso.

4. Metodologia de Pesquisa

A metodologia a ser seguida nesta pesquisa é a do paradigma de *Design Science*, utilizado para conduzir e comunicar pesquisas aplicadas, como na Engenharia de Software

[Engström et al. 2020]. O objetivo das pesquisas baseadas em *Design Science* é gerar conhecimento prescritivo para profissionais, além de compartilhar *insights* empíricos obtidos em investigações aplicadas a contextos específicos [Engström et al. 2020]. Na Figura 1 é descrito o método de pesquisa deste trabalho de doutorado. A estrutura utilizada foi baseada na proposta de [Storey et al. 2017], e apresenta elementos como: teoria proposta, contribuição empírica do estudo em termos da instância de solução de problema e os ciclos de design e validação correspondentes, suporte para a avaliação do valor do conhecimento produzido em termos de relevância, rigor e novidade.

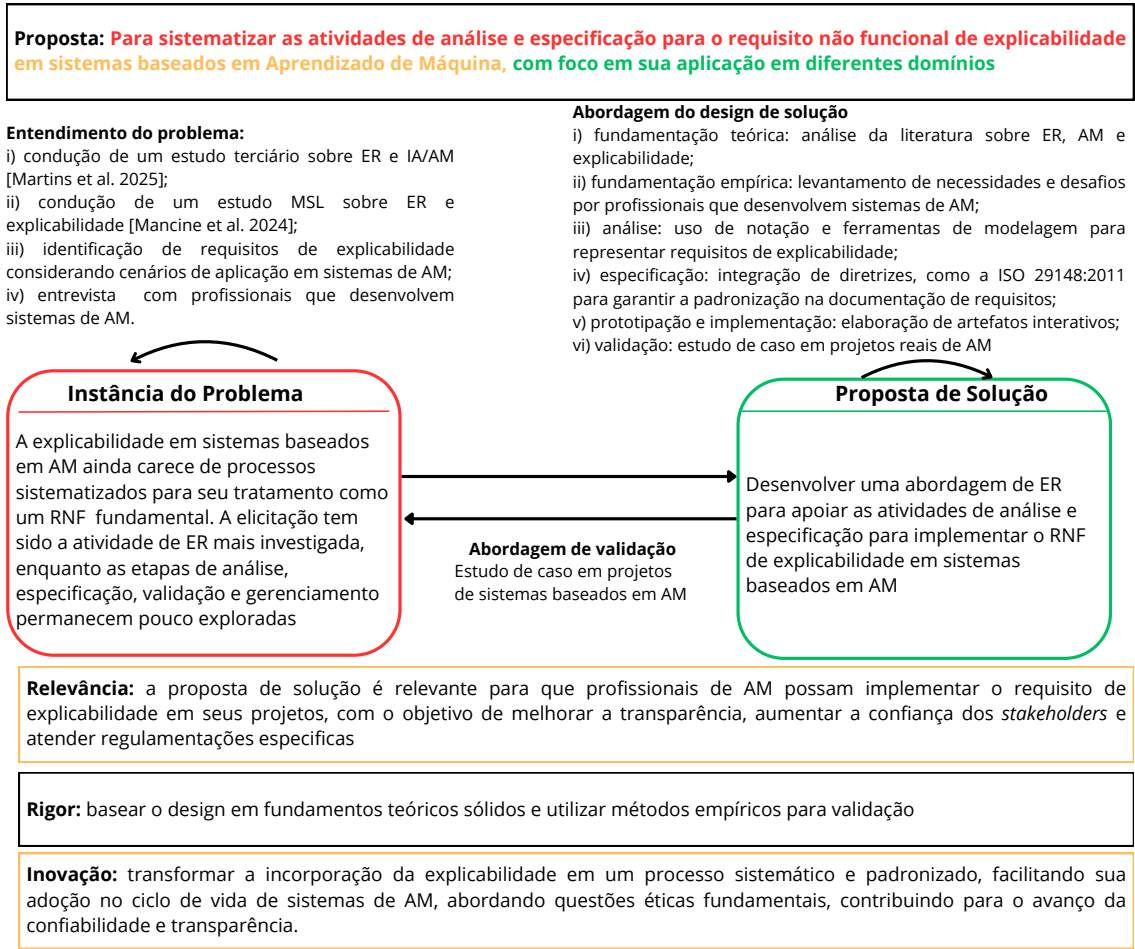


Figura 1. Método de Pesquisa deste doutorado baseado em Design Science.

Considerando a Figura 1, o entendimento do problema envolve a compreensão detalhada do seu contexto, causas e consequências, fornecendo uma base para a proposição de soluções. Nesta proposta, esse entendimento está sendo construído por meio de estudos sistemáticos da literatura sobre o tema [Martins et al. 2025] e [Mancine et al. 2024]; identificação de requisitos de explicabilidade com base em regulamentações e diretrizes internacionais e nacionais, e pela condução de uma entrevista semiestruturada com profissionais da área. Essa entrevista visa oferecer *insights* sobre o estado da prática ao tema investigado. A proposta parte desse conhecimento sobre o problema e seu contexto, alinhando-se ao objetivo geral da pesquisa descrito na Seção 2. Com base nisso, a abordagem de design de solução permite definir uma solução para o problema apresentado e planejar sua validação. A validação será conduzida por meio de estudos de caso em

sistemas baseados em AM.

Espera-se como solução uma abordagem de ER para apoiar a análise e especificação de explicabilidade em projetos de AM, viabilizada por artefatos interativos que promovam a colaboração entre os *stakeholders*. Esses artefatos têm o potencial de facilitar a comunicação, a colaboração e o engajamento dos *stakeholders*, além disso, ao integrar IA na criação desses artefatos, é possível elevar a qualidade e a eficiência das soluções propostas [Bogucka et al. 2024]. Por fim, a metodologia adotada ressalta o rigor, a relevância e a inovação da proposta apresentada.

Como contribuição, este trabalho busca avançar a ER para sistemas baseados em AM, fornecendo um processo sistemático para análise e especificação de requisitos de explicabilidade. Para isso, há atividades essenciais para a condução desta pesquisa, como as descritas a seguir.

- i) Fundamentação teórica:
 - Foi conduzido um estudo terciário que sintetizou a relação entre a ER para AM [Martins et al. 2025], no qual foram discutidas lacunas e apresentada uma agenda de pesquisa. O estudo foi submetido para avaliação em um periódico especializado e, após uma análise inicial pelos revisores, respondemos às suas dúvidas e aguardamos um retorno;
 - Realizou-se um estudo secundário, um Mapeamento Sistemático da Literatura (MSL), focado na ER e no requisito de explicabilidade no contexto de AM, publicado em [Mancine et al. 2024]. Com base nesse estudo, foram identificadas lacunas específicas relacionados à ER para a explicabilidade;
 - Revisão de diretrizes e regulamentações para IA.
- ii) Fundamentação empírica:
 - Com o objetivo de alinhar os estudos sistemáticos ao estado da prática, está em andamento uma entrevista semiestruturada com pesquisadores e profissionais da área de AM, incluindo aqueles que atuam no Centro de Excelência em Inteligência Artificial¹ (CEIA) da UFG, além de possíveis parcerias com outros centros de IA no Brasil. A entrevista vai investigar a percepção da explicabilidade no contexto da ER em sistemas baseados em AM. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da UFG sob o parecer de número 7.416.401, vinculado ao Certificado de Apresentação para Apreciação Ética (CAAE) 85638924.3.0000.5083.
- iii) Abordagem de análise:
 - Uso de notações e linguagens de modelagem da ER para modelar o requisito de explicabilidade;
 - Emprego de técnicas da ER, como cenários e casos de uso para compreender a influência da explicabilidade na interação dos *stakeholders*, permitindo a classificação de diferentes aspectos, como nível de detalhe e público-alvo.
- iv) Abordagem de especificação:
 - Definição de um modelo estruturado para especificar o requisito de explicabilidade baseados em diretrizes e normas, como a ISO/IEC/IEEE 29148:2011, para descrever requisitos de explicabilidade de forma consistente e compreensível;
 - Especificação de diferentes níveis de explicabilidade conforme o público-alvo, detalhando o que deve ser explicado e para quem.

¹<https://ceia.ufg.br/>

- Artefatos interativos implementados em uma ferramenta de software de apoio às atividades de análise e especificação de explicabilidade
- v) Validação:
 - Estudo de caso para implementação do processo sistematizado para análise e especificação do requisito de explicabilidade em sistema baseado em AM.

Dessa forma, a metodologia adotada busca garantir que a abordagem proposta seja efetiva na análise e especificação da explicabilidade em sistemas baseados em AM. A combinação de artefatos interativos e estudos de caso permitirá validar a abordagem em cenários reais, alinhando os requisitos às necessidades dos diferentes *stakeholders*. A partir das etapas descritas, espera-se que os resultados contribuam para a ER de sistemas baseados em AM, oferecendo um processo sistemático e aplicável na prática.

5. Resultados Preliminares

Esta pesquisa iniciou-se em junho de 2023 com a realização de uma revisão sistemática sobre ER e sistemas baseado em AM [Martins et al. 2025]. Embora abrangente, a revisão destacou a explicabilidade como uma lacuna para pesquisas futuras. Caracterizado como um estudo terciário, o trabalho evidenciou a necessidade de abordar a explicabilidade sob a perspectiva da ER, considerando que os *stakeholders* possuem diferentes níveis de compreensão e exigem explicações em distintos níveis de detalhamento. Assim, a documentação de requisitos deve especificar o que deve ser explicado e para quem, garantido que as informações sejam adequadas ao público-alvo.

O resultado dessa revisão, impulsionou a condução de um MSL para identificar características específicas da ER em relação a explicabilidade em sistemas baseados em AM. Nesse MSL, utilizamos pesquisas automáticas em seis bases de dados. Dos 200 artigos retornados, após aplicar os critérios de seleção, analisamos e relatamos os resultados de 27 artigos. Os estudos da literatura mostram que a explicabilidade é um requisito de qualidade emergente em sistemas baseados em AM, desafiando os paradigmas clássicos da ER, especialmente nas fases de análise e especificação. Embora a elicitação e a análise sejam as atividades mais exploradas, há uma carência de estudos empíricos. Como RNF, a explicabilidade pode contribuir para requisitos como conformidade legal, justiça e ética. Portanto, sua abordagem na ER deve ser ampliada para todas as atividades, incluindo a adaptação de técnicas clássicas ao contexto de AM.

Para identificar aspectos de explicabilidade, está sendo desenvolvido um documento de especificação de requisitos de explicabilidade (ERS), seguindo as diretrizes da ISO/IEC/IEEE 29148:2011. A saúde, como um sistema crítico de AM, é utilizada como exemplo, com base em regulamentações e diretrizes nacionais e internacionais, incluindo recomendações da Organização Mundial da Saúde (OMS), o Projeto de Lei n. 2338/2023 [Senado Federal do Brasil 2023], que regulamenta a IA no Brasil, e o estudo sobre IA na saúde do Núcleo de Informação e Coordenação do Ponto BR² (NIC.br).

6. Considerações Finais

Este artigo apresentou uma pesquisa de doutorado em andamento que tem como objetivo desenvolver uma abordagem de ER que contemple as atividades de análise e especificação

²<https://www.nic.br/noticia/na-midia/inteligencia-artificial-pode-aprimorar-a-saude-no-brasil-aponta-estudo-inedito/>

do RNF de explicabilidade em sistemas baseados em AM. A proposta de criar um artefato interativo dessa abordagem visa permitir que diferentes *stakeholders* utilizem esse requisito em projetos de AM.

Foram realizados estudos sistemáticos, atualmente, está sendo conduzida uma entrevista com desenvolvedores de AM para verificar sua percepção sobre o requisito de explicabilidade. As próximas atividades, como a definição de técnicas de análise e especificação serão conduzidas por meio de estudo teóricos e práticos, incluindo a aplicação da abordagem em projetos reais de AM. Essa aplicação ocorrerá no CEIA. Paralelamente, está prevista a produção de artigos científicos com os resultados obtidos, visando à publicação em eventos e periódicos das áreas de Engenharia de Software, ER, AM e nos domínios de aplicação, como a saúde, destacando o caráter multidisciplinar da pesquisa. A qualificação deste doutorado ocorrerá em agosto de 2025 e sua defesa está programada para março de 2027.

A implementação da abordagem proposta pode enfrentar desafios como a adaptação da explicabilidade para diferentes *stakeholders*, a validação em cenários reais e a aceitação pelos desenvolvedores. Para mitigar esses riscos, serão desenvolvidos artefatos interativos ajustáveis, validados por estudos de caso no CEIA e refinados com *feedback* de especialistas. Além disso, a abordagem será construída de forma colaborativa para garantir alinhamento com as necessidades práticas dos *stakeholders*.

Referências

- Baldassarre, M. T., Gigante, D., Kalinowski, M., Ragone, A., and Tibidò, S. (2024). Trustworthy ai in practice: an analysis of practitioners' needs and challenges. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*, pages 293–302.
- Barrera, J. M., Reina-Reina, A., Lavallo, A., Maté, A., and Trujillo, J. (2024). An extension of istar for machine learning requirements by following the prise methodology. *Computer Standards & Interfaces*, 88:103806.
- Bogucka, E., Constantinides, M., Scepanovic, S., and Quercia, D. (2024). Ai design: A responsible ai framework for pre-filling impact assessment reports. *IEEE Computer Society*.
- Chazette, L., Brunotte, W., and Speith, T. (2022). Explainable software systems: from requirements analysis to system evaluation. *Requirements Engineering*, 27(4):457–487.
- Engström, E., Storey, M.-A., Runeson, P., Höst, M., and Baldassarre, M. T. (2020). How software engineering research aligns with design science: a review. *Empirical Software Engineering*, 25:2630–2660.
- Giray, G. (2021). A software engineering perspective on engineering machine learning systems: State of the art and challenges. *Journal of Systems and Software*, 180:111031.
- Goodell, J. W. et al. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, 32:100577.

- Guizzardi, R., Amaral, G., Guizzardi, G., and Mylopoulos, J. (2022). Eliciting ethicality requirements using the ontology-based requirements engineering method. In *International Conference on Business Process Modeling, Development and Support*, pages 221–236. Springer.
- Habibullah, K. M. et al. (2024). Explainable ai: A diverse stakeholder perspective. In *2024 IEEE 32nd International Requirements Engineering Conference (RE)*, pages 494–495. IEEE.
- Jiang, F. et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4).
- Khattak, F. K., Subasri, V., Krishnan, A., Pou-Prom, C., Akinli-Kocak, S., Dolatabadi, E., Pandya, D., Seyyed-Kalantari, L., and Rudzicz, F. (2024). Mlhops: Machine learning health operations. *IEEE Access*.
- Li, T. and Han, L. (2023). Dealing with explainability requirements for machine learning systems. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1203–1208. IEEE.
- Lin, H., Han, J., Wu, P., Wang, J., Tu, J., Tang, H., and Zhu, L. (2024). Machine learning and human-machine trust in healthcare: A systematic survey. *CAAI Transactions on Intelligence Technology*, 9(2):286–302.
- Mancine, L., Soares, J. L., Kudo, T., and Bulcão-Neto, R. (2024). 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*, pages 143–158, Porto Alegre, RS, Brasil. SBC.
- Martins, M. C., Campos, L. M. C., Soares, J. L. R., Kudo, T. N., and Bulcão-Neto, R. F. (2025). Requirements engineering for machine learning-based ai systems: A tertiary study. *Journal of Software Engineering Research and Development (JSERD)*. Submitted and under review.
- Senado Federal do Brasil (2023). Projeto de Lei Nº 2338, de 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 28 jan. 2025.
- Storey, M.-A., Engstrom, E., Höst, M., Runeson, P., and Bjarnason, E. (2017). Using a visual abstract as a lens for communicating and promoting design science research in software engineering. In *2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 181–186. IEEE.
- Umm-E-Habiba (2023). Requirements engineering for explainable ai. In *2023 IEEE 31st International Requirements Engineering Conference (RE)*, pages 376–380.
- Vayena, E., Blasimme, A., and Cohen, I. G. (2018). Machine learning in medicine: addressing ethical challenges. *PLoS medicine*, 15(11):e1002689.
- Villamizar, H., Kalinowski, M., Lopes, H., and Mendez, D. (2024). Identifying concerns when specifying machine learning-enabled systems: a perspective-based approach. *Journal of Systems and Software*, 213:112053.