Federal University of Goias
Institute of Informatics
Postgraduate Program in Computer Science

Mariana Crisostomo Martins

A Discovery Framework for AI Innovation Projects

Uma abordagem de Discovery para Projetos de IA e Inovação

Goiânia
2026

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [X] Tese [ ] Outro*:___________

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Mariana Crisostomo Martins

3. Título do trabalho

A Discovery Framework for AI Innovation Projects

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

sei! assinatura eletrônica

Documento assinado eletronicamente por **Renato De Freitas Bulcao Neto**, **Professor do Magistério Superior**, em 15/07/2026, às 14:59, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica

Documento assinado eletronicamente por **Mariana Crisostomo Martins**, **Usuário Externo**, em 15/07/2026, às 16:30, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6338472** e o código CRC **9C9E6DF7**.

MARIANA CRISOSTOMO MARTINS

A Discovery Framework for AI Innovation Projects

Uma abordagem de Discovery para Projetos de IA e Inovação

Thesis submitted to the Postgraduate Program in Computer Science of the Institute of Informatics at the Federal University of Goias, as a partial requirement for the degree of Doctor in Computer Science.

Area of concentration: Computer Science.

Research line: Methodologies and Techniques of Computing.

Advisor: Prof. Dr. Renato de Freitas Bulcão Neto

Co-Advisor: Prof. Dr. Marcos Kalinowski

Goiânia
2026

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Martins , Mariana Crisostomo

A Discovery Framework for AI and Innovation Projects [Manuscrito] = Uma abordagem de Discovery para Projetos de IA e Inovação / Mariana Crisostomo Martins. - 2026.

CXCVIII, 198 f.:

Orientador: Prof. Dr. Renato de Freitas Bulcão-Neto ; co-orientador: Dr. Marcos Kalinowski

Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Informática (INF), Programa de Pós-Graduação em Ciência da Computação, Goiânia, 2026.

Anexo.

Apêndice.

Bibliografia.

Inclui: siglas, lista de figuras, lista de tabelas.

1. Requirements Engineering. 2. Discovery. 3. AI Systems. 4. Innovation Projects.
I. Bulcão-Neto, Renato de Freitas , orient. II. Kalinowski, Marcos, co-orient. III. Título.

CDU 004

UFG

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

ATA DE DEFESA DE TESE

Ata nº **12** da sessão de Defesa de Tese de **Mariana Crisostomo Martins**, que confere o título de Doutora em Ciência da Computação, na área de concentração em **Ciência da Computação**.

Aos quinze dias do mês de junho de dois mil e vinte e seis, a partir das catorze horas, no lab. 250 do INF, realizou-se a sessão pública de Defesa de Tese intitulada **"A Discovery Framework for AI Innovation Projects"**. Os trabalhos foram instalados pelo Orientador, Professor Doutor Renato de Freitas Bulcão Neto (INF/UFG) com a participação dos demais membros da Banca Examinadora: Professor Doutor Marcos Kalinowski (PUC-Rio), coorientador; Professora Doutora Carla Taciana Lima Lourenço Silva Schuenemann (CIN/UFPE), membra titular externa; Professor Doutor Awdren de Lima Fontão (FACOM/UFMS), membro titular externo; Professor Doutor Hugo Ricardo Guarín Villamizar (FORTISS GmbH/Alemanha); e Professora Doutora Sofia Larissa da Costa Paiva (INF/UFG), membra titular externa. A participação dos professores Marcos Kalinowski, Carla Taciana Lima Lourenco Silva Schuenemann, Awdren de Lima Fontão e Hugo Guarín Villamizar ocorreu por meio de videoconferência. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese, tendo sido a candidata **aprovada** pelos seus membros. Proclamados os resultados pelo Professor Doutor Renato de Freitas Bulcão Neto, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos quinze dias do mês de junho de dois mil e vinte e seis.

TÍTULO SUGERIDO PELA BANCA

---

sei! assinatura eletrônica — Documento assinado eletronicamente por **Awdren de Lima Fontão**, **Usuário Externo**, em 15/06/2026, às 17:23, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/D10543.htm).

sei! assinatura eletrônica — Documento assinado eletronicamente por **Renato De Freitas Bulcao Neto**, **Professor do Magistério Superior**, em 15/06/2026, às 17:23, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica — Documento assinado eletronicamente por **Sofia Larissa Da Costa Paiva**, **Professora do Magistério Superior**, em 15/06/2026, às 17:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica — Documento assinado eletronicamente por **Carla Taciana Lima Lourenco Silva Schuenemann**, **Usuário Externo**, em 15/06/2026, às 17:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica — Documento assinado eletronicamente por **Hugo Ricardo Guarin Villamizar**, **Usuário Externo**, em 15/06/2026, às 17:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica — Documento assinado eletronicamente por **Mariana Crisostomo Martins**, **Discente**, em 15/06/2026, às 19:06, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica — Documento assinado eletronicamente por **Marcos Kalinowski**, **Usuário Externo**, em 18/06/2026, às 15:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6188161** e o código CRC **561076FE**.

Referência: Processo nº 23070.022461/2026-52 SEI nº 6188161

Mariana Crisostomo Martins

Graduou-se em Ciência da Computação na UFG - Universidade Federal de Goiás. Durante sua graduação, foi monitor no Instituto de Informática para a disciplina de Engenharia de Software e pesquisador do CNPq em um trabalho de iniciação científica ainda na área de Engenharia de Software. Também durante a graduação participou de projetos de extensão, como: organização do CBIS 2016 e desenvolvimento de atividades no Observatório de Dinâmica Humana - “Redes Sociais Online - Você sabia?”. Durante o mestrado desenvolveu um catálogo de padrões de requisitos baseado no Manual de Certificação de Sistemas de Registro de Saúde da Sociedade Brasileira de Informática em Saúde. Atualmente, desenvolve um processo de Engenharia de Requisitos para Sistemas de Inteligência Artificial.

Dedico este trabalho a minha mãe, primeira pessoa que acreditou e me incentivou na vida acadêmica. E às mulheres da minha vida; eu não chegaria aqui sem vocês.

Acknowledgments

Agradeço a Deus pela oportunidade de realizar esse trabalho, sem Ele nada disso seria possível.

Agradeço as instituições que possibilitaram o desenvolvimento do mesmo: CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CEIA - Centro de Exceleência em Inteligência Artificial, ExACTa (PUC-Rio) - Laboratório de Experimentação Ágil e Transformação Digitial, SAIL - Software and AI Engineering Lab, UFG - Universidade Federal de Goiás e INF - Instituto de Informática.

Agradeço ao professor Renato de Freitas Bulcão Neto e a professora Taciana Novo Kudo pela dedicação, orientação, paciência e auxílio em todo o processo de desenvolvimento dessa pesquisa. Agradeço também ao professor Marcos Kalinowski por ter me recebido e me acolhido em seu grupo de pesquisa e ter apoiado o desenvolvimento da minha pesquisa possibilitanto uma nova perspectiva.

Agradeço ao Lucas pelo incentivo para começar essa jornada, me ajudou a continuar, sonhou e me acompanhou em cada etapa. Agradeço pelo apoio, conversas e orientações e motivação em todo o processo, pela dedicação ao meu trabalho e ao background para que eu conseguissem me concentrar no desenvolvimento da pesquisa.

Agradeço aos meus pais, familiares e amigos que me ajudaram nessa caminhada, que me orientaram, me escutaram e entenderam os momentos de ausência enquanto me dedicava a esse trabalho. Gostaria de destacar cada grupo que me apoiou: aos meus irmãos (Thiago e Gustavo) que sempre me ouviam e me incentivavam; a família e amigos que eu ganhei com meu relacionamento (Kamila, tia Bebeire, sogros e Bruno); o Clube do Livro que me fazia enxergar as dificuldades sob uma nova perspectiva (Andressa, Gabi, Giovanna, Iohana e Marina); aos amigos da faculdade que sempre me lembravam que o meu sonho era maior que qualquer adversidade (Altino, Giovanna, Luiz Fernando, Ana Paula e Adriana); as minhas tias (Luzia, Socorro e Raimunda) e minha avó Deusa que dedicaram tudo quanto podiam para que eu chegasse até aqui e sempre me lembraram do quanto minha mãe estaria orgulhosa; aos amigos de escola que sempre se mantiveram firmes e acreditando na minha conquista (deixo um lugar cativo para Amandinha e Iza); a minha amiga de infância Leticia que me viu amadurecer, sonhar e continuou presente mesmo a 600 km. Agradeço ao meu terapeuta Nathan que me acompanhou durante toda a pós-graduação me mostrando mais do que eu conseguia enxergar no momento. Agradeço ainda aos colegas do meu grupo de pesquisa (Ernesto, Lívia, Cinara e Filipe) pelas trocas. E agradeço aos colegas da ExACTa e do SAIL (em especial Lucas, Antônio, Marina, Júlia, Mariana, Clarissa e Ariane) por me receberem, me ensinarem sobre o Rio, sobre a pesquisa e pelas trocas que me fizeram crescer.

Agradeço ainda a banca pela disponibilidade e pelas contribuições para o desenvolvimento deste trabalho. Pelos encontros em cada evento científico e que sempre com tanto carinho e dedicação se dispuseram a me ouvir, a sugerir, a me enxergar uma nova alternativa até esse momento.

Enfim agradeço a todos que diretamente ou indiretamente colaboraram com o desenvolvimento desse trabalho.

Obrigada!

Like all other arts, the Science of Deduction and Analysis is one which can only be acquired by long and patient study, nor is life long enough to allow any mortal to attain the highest possible perfection in it.

Arthur Conan Doyle,
Sherlock Holmes: A Study in Scarlet.

Abstract (Portuguese)

Martins, Mariana. **A Discovery Framework for AI Innovation Projects**. Goiânia, 2026. 198p. PhD Thesis. Postgraduate Program in Computer Science, Institute of Informatics, Federal University of Goias.

As práticas tradicionais de Engenharia de Requisitos (ER) enfrentam limitações ao lidar com o desenvolvimento de sistemas baseados em Inteligência Artificial (IA) centrado em dados e modelos. A literatura revela uma lacuna de suporte metodológico para as atividades de descoberta de problemas (discovery) em projetos de inovação em IA, o que compromete as fases subsequentes de desenvolvimento. Para mitigar esse problema, este trabalho introduz o DIP-AI, um framework de descoberta para guiar a exploração em estágio inicial nessas iniciativas. Ancorado em um estudo terciário, o framework harmoniza processos técnicos das normas ISO/IEC/IEEE 12207 e ISO/IEC 5338 à abordagem ágil do Design Thinking, além de incorporar as matrizes 5W2H e GUT para gerenciamento de riscos e priorização. A validação da proposta deu-se por meio de uma estratégia multi-método: (i) um estudo de caso empírico em um projeto indústria-academia; (ii) uma avaliação controlada com pós-graduandos em Engenharia de Software e IA; e (iii) entrevistas semiestruturadas com coordenadores de projetos de inovação. Embora os resultados confirmem a utilidade e o valor estratégico do DIP-AI para mitigar riscos e alinhar stakeholders, também evidenciam barreiras de sobrecarga cognitiva no manuseio do framework. Em função disso, concebeu-se uma ferramenta assistiva baseada em Modelos de Linguagem de Grande Porte (LLMs), que foi avaliada qualitativamente por meio de um Focus Group com profissionais do mercado. Os resultados demonstram que a abordagem integrada framework-ferramenta reduz a fricção metodológica inicial, acelera a curva de aprendizado e desperta forte intenção de adoção contínua. Como contribuição acadêmica, esta pesquisa preenche um gap metodológico na literatura de ER para IA, fornecendo um modelo conceitual rigoroso e replicável. Para a indústria, a entrega consolida-se como um ativo relevante e aplicável, capacitando organizações e equipes multidisciplinares a estruturarem a descoberta de problemas complexos de IA com maior previsibilidade, reduzindo desperdício de recursos e maximizando o sucesso e a sustentabilidade de produtos.

Keywords

<Engenharia de Requisitos, Descoberta, Sistemas de IA, Projetos de Inovação>

Abstract

Martins, Mariana. **DIP-AI: A Discovery Framework for AI Innovation Projects**. Goiânia, 2026. 198p. PhD Thesis. Postgraduate Program in Computer Science, Institute of Informatics, Federal University of Goias.

Traditional Requirements Engineering (RE) practices face limitations when dealing with the development of data- and model-centric Artificial Intelligence (AI) systems. The literature reveals a lack of methodological support for problem discovery activities in AI innovation projects, which compromises subsequent development phases. To mitigate this issue, this work introduces DIP-AI, a discovery framework designed to guide early-stage exploration in such initiatives. Grounded in a tertiary study, the framework harmonizes technical processes from the ISO/IEC/IEEE 12207 and ISO/IEC 5338 standards with the agile Design Thinking approach, while also incorporating 5W2H and GUT matrices for risk management and prioritization. The proposal was validated through a multi-method strategy: (i) an empirical case study within an industry-academia project; (ii) a controlled evaluation involving graduate students in Software Engineering and AI; and (iii) semi-structured interviews with innovation project coordinators. Although the results confirm the utility and strategic value of DIP-AI in mitigating risks and aligning stakeholders, they also highlight barriers related to cognitive overload when using the framework. Consequently, an assistive tool based on Large Language Models (LLMs) was developed and qualitatively evaluated via a focus group with industry professionals. The results demonstrate that the integrated framework-tool approach reduces initial methodological friction, accelerates the learning curve, and fosters a strong intention for continued adoption. As an academic contribution, this research fills a methodological gap in the RE for AI literature by providing a rigorous and replicable conceptual model. For the industry, this deliverable establishes itself as a relevant and practical asset, enabling organizations and multidisciplinary teams to structure the discovery of complex AI problems with greater predictability, thereby reducing resource waste and maximizing product success and sustainability.

Keywords

<Requirements Engineering, Discovery, AI Systems, Innovation Projects>

Contents

List of Figures

List of Tables

List of Program Listings

CHAPTER 1

Introduction

This chapter begins by presenting the main theme of this research and the justifications for its study. Next, the motivation and the problem to be investigated and the general and specific objectives of this research are described. Finally, the chapter structure of this document is presented.

1.1 Context

Artificial Intelligence (AI) systems have become increasingly widespread with advances in computing capabilities. Machine learning (ML), a subfield of AI, aims to acquire knowledge by extracting patterns from data and using this knowledge to solve problems [29]. The ML lifecycle structures the activities required to develop, train, deploy, and maintain ML models [75]. Improvements across this lifecycle can enhance the performance and reliability of ML-based applications [29]. Lifecycle representations are commonly described as iterative processes, reflecting the experimental and evolutionary nature of AI system development.

Despite their potential, the success of AI systems—particularly those based on ML—in real-world settings remains below expectations [43]. Several AI projects fail to progress beyond early stages, with only about 36% of surveyed organizations reporting that ML projects advance beyond pilot phases [38]. Software systems incorporating AI components pose significant challenges for Software Engineering (SE) and, more specifically, Requirements Engineering (RE). These systems typically require intensive communication and collaboration among multidisciplinary teams to address technical, organizational, and domain-specific complexities [71].

A fundamental step in SE is Requirements Engineering (RE), through which stakeholder needs are elicited, analyzed, and validated. According to Pressman [56], the RE phase supports stakeholder identification, specification validation, and conflict resolution, while guiding the entire software lifecycle. Given its importance, inadequate execution of RE raises significant concerns. In practice, requirements are often specified in an informal and descriptive manner, increasing the risk of failures that can impact the entire development process. Poorly executed requirements activities make projects vulnerable to rework, misalignment, and quality issues [52, 65]. Tockey reports that approximately 56% of software defects originate from incomplete documentation, including incorrect, omitted, conflicting, or ambiguous requirements [65].

The AI and ML communities have accumulated extensive experience showing that while developing and deploying ML models can be relatively fast and inexpensive, maintaining AI systems over time is difficult and costly [63]. Consequently, the overall success of ML-based AI applications often falls short of expectations [43]. Alberto [12] identifies nine reasons for ML project failure, including several issues related to project management and requirements definition—areas that can be addressed through a well-structured RE process.

Early-stage activities aimed at understanding problems, stakeholders, data, and feasibility play a critical role. These activities, commonly referred to as discovery, are fundamental to aligning business goals, technical constraints, and data-driven opportunities before formal specification and development take place. However, In AI projects, discovery is often informal, fragmented, and insufficiently supported, which can compromise subsequent development phases [10].

1.2 Problem and Motivation

ER is considered the most difficult activity in AI projects [40]. Requirements in AI scenario are uncertain, new quality properties emerge (such as explainability), goals are ambitious, and development involves a high degree of iterative experimentation. Villamizar et al. [68] pointed out that the state of the art of RE for ML is limited, which makes application difficult. Furthermore, the literature and practice do not clarify issues regarding requirements for this domain, such as specification. Ahmad and Villamizar et al. [6, 71] also highlighted that RE research with intersection in ML focuses on the use of ML techniques to support requirements activities.

The experimental nature of ML and innovation projects presents several challenges, such as understanding the problem, identifying the problem from the data, maintaining effective communication, aligning technical and domain knowledge to manage technological feasibility [9]. In order to assist in the development of AI systems and mitigate development and maintenance problems of these systems, we believe that the adapted and adequate application of the ER process is a viable solution.

The application of an ER process can guide requirements engineers in understanding the data by identifying its characteristics and distribution, which is essential to ensure the quality of the model [73]. The process also helps in understanding the context and domain [39], stakeholder collaboration [40], concerns about model degradation

[72], perception about quality, non-functional requirements and assists in the discussion of trade-offs [71, 5, 34]. In short, the acquisition and understanding of this information with the involvement of all stakeholders helps in the development of higher quality software.

Empirical evidence from industry reinforces this gap. Through a systematic tertiary study and interviews conducted at the Center of Excellence in Artificial Intelligence (CEIA), a unit of the Brazilian Company for Industrial Research and Innovation (EMBRAPII), we identified recurring challenges in discovery activities for AI projects [47]. These challenges include difficulties like:

- concern with non-functional requirements and data requirements;
- traditional elicitation and specification techniques are the most common, for example, *brainstorming* and natural language specification;
- it is not clear which techniques and tools support RE in this domain;
- difficulty in establishing communication with stakeholders and unrealistic expectations of customers and users;
- difficulty in establishing metrics for quality assurance that go beyond the model;
- lack of traceability;
- difficulty in choosing and knowing methods for elicitation and specification appropriate for AI and research, development and innovation (RDI) projects;
- difficulty in aligning data with the project objective and establishing the project objective;
- difficulty in obtaining data, obtaining infrastructure, time management, communication with the team.

Taken together, these findings indicate a clear problem: discovery activities in AI-based innovation projects lack systematic support, methodological guidance, and tailored techniques. This limitation compromises downstream development phases and motivates the need to investigate how an adapted Requirements Engineering (RE) process can better structure and support early-stage problem characterization to improve the sustainability of AI systems.

1.3 Research Question

Given the identified lack of systematic, lightweight RE processes specifically designed for the complexities of Artificial Intelligence within Research, Development, and Innovation (RDI) ecosystems, this research seeks to answer the following central question:

RQ: How does the framework structure and support the requirements discovery phase in AI-based innovation projects?

1.4 Objectives

1.4.1 General Objective

The main objective of this study is to develop and evaluate the effectiveness, efficiency, usability, and usefulness of a Discovery Framework for the development of AI-Based Systems in Research, Development, and Innovation (RDI) projects.

1.4.2 Specific Objectives

As specific objectives in this work, are:

- understand and characterize the problem regarding the state of the art and practice in the context of RE and AI in RDI;
- define the framework with activities, tasks, objective and expected results;
  ISO 12207 technical processes based;
  and double diamond activities formation in Design Thinking;
- evaluate the framework with laboratory, static and dynamic validation [21];
  evaluate the effectiveness and efficiency of using the framework in innovation projects;
  assess the perception of usability and usefulness, this study will employ a questionnaire based on the Technology Acceptance Model (TAM) [21];
- support the technology transfer of the DIP-AI framework at CEIA [54] to empirically validate its effectiveness in improving the requirements discovery phase.;
  share documentation illustrates process use;
  training with coordinators, requirements engineering and students of CEIA.

1.5 Research Methodology

In Easterbrook et al. (2008), the authors present key questions to support the selection of an appropriate research method, ranging from philosophical considerations to practical aspects related to its application. Based on these guidelines, we classified our research within a suitable methodological group. Considering the aspects presented and the classification proposed by Easterbrook et al. (2008), we present Table 1.1, which positions the research activities of this thesis within the corresponding methodological framework.

Aspect	**Classification**
Types of Research Question Exploratory	Description and Design
Philosophical stance	Pragmatism
Nature	Qualitative and Quantitative
Research methods	Case study, survey, and focus group

Table 1.1: *Methodological framework. Source: The author (2025)*

This study adopts an exploratory research approach, as the phenomenon under investigation—discovery activities in AI-based innovation projects—remains insufficiently understood and weakly structured in both research and practice. Exploratory research is appropriate in contexts where existing knowledge is fragmented, theoretical frameworks are immature, and there is a need to investigate how activities are currently performed before proposing or evaluating solutions [25]. In this sense, exploration serves as the overarching research type guiding this study.

Within this exploratory scope, the research incorporates a **descriptive** dimension to systematically characterize how discovery activities are conducted in AI-based Research, Development, and Innovation (RDI) projects. Through literature analysis and empirical investigation in an industrial setting, the study describes existing practices, challenges, stakeholder interactions, and decision-making processes related to early problem understanding and feasibility assessment. This descriptive focus enables an accurate representation of real-world phenomena without manipulating variables or imposing predefined structures. The **design** dimension supports the conception and refinement of a structured approach to discovery, informed by empirical evidence and grounded in practical needs. Together, these dimensions enable the exploration, organization, and purposeful shaping of discovery activities in a complex and emerging domain. This classification is appropriate for our research question, as part of this work investigates the existing contributions to the requirements engineering phase in AI and innovation projects and explores how this activity can be improved to achieve the goal of assisting in the development of these projects.

The philosophical stance adopted in this research is **pragmatism**, in line with the recommendations of Easterbrook et al. (2008) [25] for applied software engineering research. Pragmatism emphasizes the practical consequences of research and values knowledge based on its usefulness in solving real-world problems. This stance is particularly appropriate for this study, as the research is motivated by concrete challenges observed in industrial AI projects and aims to generate actionable knowledge that can improve discovery activities in practice. Given the complex, socio-technical, and data-driven nature of AI-based innovation projects, a pragmatic stance enables the integration of both empirical evidence and practitioner perspectives, supporting the development and evaluation of solutions that are both theoretically grounded and practically relevant.

This research adopts a mixed-methods approach, combining qualitative and quantitative methods, as advocated by Easterbrook et al. (2008) for comprehensive software engineering investigations. The qualitative component is essential for understanding the context, motivations, perceptions, and experiences of stakeholders involved in discovery activities. Tertiary study, qualitative data, collected through focus group, questionnaires and interviews of practices, supports in-depth exploration of how discovery is conducted, why certain challenges arise, and how practitioners interpret and address uncertainty in AI projects. The quantitative component complements this understanding by enabling the measurement and comparison of aspects such as perceived usefulness, usability, effectiveness, and efficiency of structured discovery activities. Quantitative data supports the evaluation of proposed approaches and contributes to building empirical evidence that can be analyzed systematically and compared across contexts. The combination of qualitative and quantitative methods strengthens the validity of the findings by enabling triangulation and providing a richer understanding of the research problem.

1.5.1 Research steps

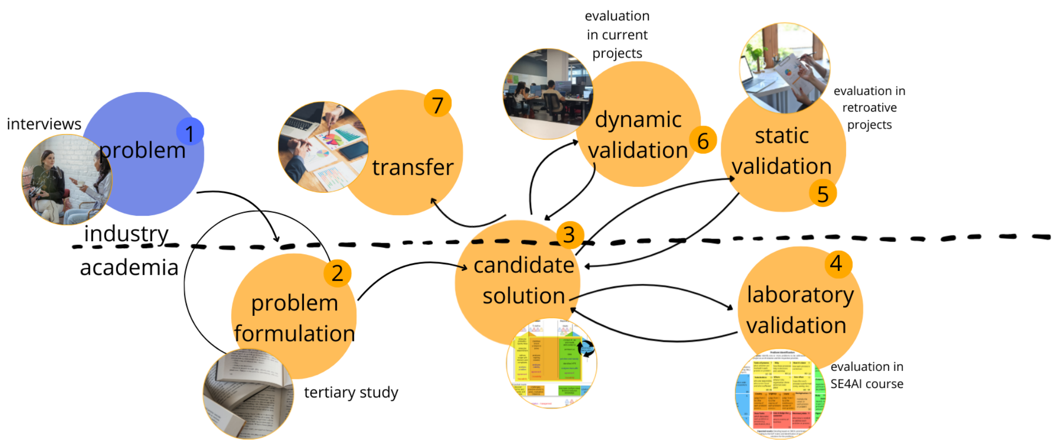

Figure 1.1: *Application of the TTM Method for the proposal of this research.*

In this section, we describe the process adopted to design and evaluate the framework of this research, based on the Technology Transfer Model (TTM) proposed by Gorschek et al [32]. This model was selected because the research aims to connect academic investigation to industrial practice through iterative evaluation in both contexts, which is a central objective of studies oriented towards technology transfer.

The TTM serves as an empirical research method designed to bridge the gap between academic theory and industrial practice through a symbiotic, long-term collaboration. Rather than relying on a prepackaged or rigid prescriptive framework, this model was created in an evolutionary, on-demand manner within real-world industrial software engineering environments to satisfy two concurrent needs: the researchers' requirement to validate new technologies for academic purposes, and the industry's need to minimize operational risks while maximizing business value. Characterized by close cooperation, step-wise collaborative validation, and an iterative structure, the model treats technology transfer not as a passive handover of publications, but as an active, flexible, and continuous process of co-creation. Consequently, it functions as a robust research methodology because it enables the iterative refinement and empirical validation of new frameworks (such as the DIP-AI framework) within authentic, resource-constrained ecosystems, ensuring that the resulting scientific contributions possess both academic rigor and immediate industrial relevance.

The TTM structures the research process around the identification of real industrial needs, the design of candidate solutions, and their progressive evaluation in controlled and industrial environments. This iterative cycle allows for the systematic incorporation of feedback from professionals in the field, increasing the relevance, usability, and practical applicability of the proposed solution.

Figure 2.3 presents the seven stages of the model, which are described sequentially in the following subsections, using the terminology defined by Gorschek et al. Similar uses of this model in requirements engineering research have been reported in the literature [70].

While Design Science Research (DSR) provides a rigorous framework for the conceptual creation and prescription of innovative artifacts, it often lacks explicit, granular mechanisms to navigate the socio-technical friction inherent to industrial adoption. Therefore, this research explicitly adopts Gorschek's TTM as its guiding empirical methodology. The choice of the TTM is justified by its evolutionary, on-demand nature, which structures the validation of the proposed framework. By transitioning from controlled laboratory evaluations to a real-world University-Industry Collaboration (UIC) ecosystem, the TTM establishes a symbiotic partnership between researchers and practitioners. This framework ensures that early-stage requirement barriers—such as data non-determinism and team cognitive overload—are identified and mitigated through continuous feedback loops, ultimately producing a solution that guarantees both academic rigor and uncompromised industrial market relevance.

The application of TTM in this research begins with the identification of a concrete industry need, supported by evidence from the literature. Based on this need, the framework proposed in this thesis was conceived. Subsequently, it was iteratively refined through multiple validation cycles in laboratory, static, and dynamic contexts. Each cycle provided adjustments to the artifact, ensuring alignment with both theoretical foundations and practical constraints. The final phase of the model is defined by technology transfer. In the following sections, we will detail how the model was applied to the conception and evaluation of this research.

1.6 Summary of Publications

At this stage, we present the publications that have been published as part of the research conducted in this thesis.

1. MARTINS, M. C., CAMPOS, L. M. C., SOARES, J. L. R., KUDO, T. N., & BULCÃO-NETO, R. F. (2025). Requirements Engineering for Machine Learning-Based AI Systems: A Tertiary Study. Journal of Software Engineering Research and Development, 13(2), 13:129 – 13:142. https://doi.org/10.5753/jserd.2025.4892
2. MARTINS, Mariana Crisostomo; KUDO, Taciana Novo; BULCÃO-NETO, Renato F.. A Qualitative Study on Requirements Engineering Practices in an Artificial Intelligence Unit of the Brazilian Industrial Research and Innovation Company. In: Congresso Ibero-Americano em Engenharia de Software (CIBSE), 2024, p. 46-60. DOI: https://doi.org/10.5753/cibse.2024.28438.
3. MARTINS, Mariana Crisostomo; KUDO, Taciana Novo; BULCÃO-NETO, Renato F. A Requirements Engineering Process for Machine Learning Innovation Projects. In: Workshop on Requirements Engineering (WER), 27., 2024, Buenos Aires. Argentina.
4. MARTINS, Mariana Crisostomo; ROCHA, Lucas Elias Cardoso; ROMÃO, Lucas Cordeiro; KUDO, Taciana Novo; KALINOWSKI, Marcos; BULCÃO-NETO, Renato de Freitas. DIP-AI: A Discovery Framework for AI Innovation Projects. In: SIMPÓSIO BRASILEIRO DE QUALIDADE DE SOFTWARE (SBQS), 24. , 2025, São José dos Campos/SP. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025 . p. 131-141. DOI: https://doi.org/10.5753/sbqs.2025.13834.

1.7 Thesis Structure

This thesis is organized into a sequence of chapters that progressively build the theoretical foundation, empirical understanding, and proposed solution for supporting discovery activities in AI-based Research, Development, and Innovation (RDI) projects.

Chapter 2 presents the theoretical foundations necessary to understand the research. It introduces key concepts related to Software and Requirements Engineering, discusses ISO/IEC 12207 as a reference framework, and contextualizes AI systems based on machine learning, which constitute the technological domain of this study. In addition, this chapter presents the Technology Transfer Model that guides the methodological construction of the research.

Chapter 3 reviews related work and existing approaches found in the literature. The chapter analyzes similarities and differences between these approaches and the proposal of this thesis, highlighting gaps, challenges, and limitations that motivate the need for further investigation. This analysis supports the positioning of the research within the current state of the art.

Chapter 4 presents empirical studies that informed the development of the proposal. Reports the results of interviews conducted with coordinators of AI and machine learning projects, identifying how discovery and early engineering activities are performed in practice, as well as the main difficulties and opportunities encountered [47].

Chapter 5 presents a systematic tertiary study of the literature on Requirements Engineering and discovery-related activities in the context of AI and machine learning. The findings of this study synthesize existing evidence and have resulted in scientific publications that have been published or are currently under review.

Chapter 6 describes the proposal of this thesis. The proposed approach aims to structure and support discovery activities for AI-based systems in RDI projects, drawing on concepts from Requirements Engineering and standards such as ISO/IEC 12207. This chapter details the activities involved in the proposal and the framework developed like candidate solution.

Chapter 7 presents the evaluation of the proposed approach through three complementary validation strategies: laboratory studies with students, static or retrospective evaluations with project coordinators, and dynamic evaluations conducted in ongoing projects. These validations aim to assess the feasibility, usefulness, and applicability of the proposed discovery support in different contexts.

Chapter 8 introduces the tool developed to support the use of the proposed framework. The tool leverages Large Language Models (LLMs) to assist practitioners during discovery activities, facilitating documentation, reflection, and knowledge organization.

Chapter 9 presents the final evaluation of the proposal, conducted with industry experts. This chapter discusses expert perceptions regarding the proposed framework, the supporting tool, and the use of LLMs to assist discovery activities in AI and RDI projects.

Finally, Chapter 10 concludes the thesis by summarizing the research, highlighting its main scientific and practical contributions, discussing limitations, and outlining directions for future work.

1.8 Chapter Summary

This chapter provided an introduction to the thesis. First, we presented the context that motivated its development, followed by the underlying problems and motivations for choosing the research topic. We then defined the main and specific objectives and research question, explained the adopted methodology, and summarized the publications produced throughout this work. Finally, we outlined the overall structure of the thesis. In the next chapter, we present the background necessary for understanding this work, as well as the related studies to our research.

CHAPTER 2

Background

This chapter establishes the theoretical foundation and the state of the art necessary to ground this research. The architectural baseline of this study lies at the intersection of **RE (RE)** and the specific engineering challenges imposed by **Artificial Intelligence (AI)** and **Machine Learning (ML)** systems. To discipline this non-deterministic domain, we introduce **ISO/IEC/IEEE 12207** as the core reference framework for software lifecycle processes, specifically focusing on early-stage business analysis and discovery activities. This generic lifecycle standard is further augmented by **ISO/IEC 5338**, which provides specialized processes for AI system lifecycles, and **ISO 31000** to govern data-driven and operational risk management.

To bridge the gap between rigid standards and practical industry needs, this chapter covers the human-centered and agile tools incorporated into our approach. We discuss the **Design Thinking** framework as a vehicle for collaborative problem discovery, complemented by pragmatic elicitation and prioritization tools: the **5W2H** framework for structured problem characterization, and the **GUT Matrix** for systematic severity and urgency analysis. Finally, we describe the **Technology Transfer Model (TTM)**, which establishes the methodological workflow used to build, iteratively refine, and empirically validate the proposed approach within real-world industrial ecosystems.

This chapter presents the key concepts that provide the theoretical foundation for this research. It introduces ISO/IEC 12207 as a reference framework for software lifecycle processes and discusses its relevance to early engineering and discovery activities. The chapter also presents Artificial Intelligence (AI) systems based on machine learning, which constitute the technological domain addressed in this study. Finally, the Technology Transfer Model that guides the construction and validation of the proposed approach is described.

2.1 Requirements Engineering (RE)

The software lifecycle encompasses both development and maintenance phases. Concerning development, the Software Engineering Body of Knowledge (SWEBOK) defines activities such as RE, software design, construction, and testing. Software maintenance, in turn, extends from system delivery until system retirement [14].

RE is concerned with identifying, analyzing, and documenting the requirements necessary for a software system to meet stakeholder needs. Software design refers to the phase in which this knowledge is transformed into models that support developers' understanding of the problem domain and solution structure. Software construction involves the actual implementation of the system and includes activities such as coding and debugging, as well as construction planning, detailed design, unit testing, and integration testing. The testing phase verifies whether the implemented system meets stakeholder expectations. Finally, the maintenance phase comprises all activities related to the economic support of the software, including pre-delivery and post-delivery activities. Pre-delivery activities focus on planning maintenance, whereas post-delivery activities include modification, training, operation, and customer support [14].

A fundamental step in software development is RE, as it provides the basis for understanding stakeholder needs and defining what the system should accomplish. According to SWEBOK, RE is typically structured into several activities. Requirements elicitation focuses on identifying stakeholders, understanding their needs, and uncovering the origins of requirements by analyzing the problem context. Requirements analysis involves classifying and prioritizing requirements, which may include functional and non-functional requirements, as well as categorization by scope, priority, or volatility. Requirements specification results in the production of a document that can be systematically reviewed, evaluated, and approved, commonly referred to as the Software Requirements Specification (SRS). Artifacts produced during RE are subject to validation and verification.

Validation aims to ensure that the documented requirements accurately reflect stakeholder needs, while verification focuses on checking consistency, completeness, clarity, and compliance with adopted standards. In addition, requirements management addresses change control and supports software evolution throughout the lifecycle.

RE impacts the entire software development lifecycle. The outcomes of RE directly influence software design, as requirements serve as the foundation for constructing models that represent the problem and proposed solution. RE activities also affect testing, from unit testing to system and acceptance testing, by defining the criteria against which the system is evaluated [59]. Furthermore, software maintenance is planned based on requirements, which guide the monitoring and management of changes over time.

2.2 Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) seeks to understand and construct systems capable of exhibiting intelligent behavior. According to Russell and Norvig [61], the Turing Test—proposed by Alan Turing in 1950—provides an operational definition of intelligence. A machine is said to pass the test if a human interrogator cannot distinguish between responses generated by a human and those generated by a computer through written interaction. To achieve this, a system must possess capabilities such as natural language processing, knowledge representation, automated reasoning, and machine learning. More advanced forms of intelligence also require perception, such as computer vision, and the ability to act in the physical world through robotics.

Machine learning is a subfield of AI that focuses on the development of algorithms and statistical models that enable systems to learn from data and make predictions or decisions without explicit programming [41]. Machine learning algorithms aim to learn mappings between input and output data by identifying patterns in training data and generalizing this knowledge to new, unseen data. Conceptually, these algorithms can be viewed as searching a large space of candidate models, guided by experience, to optimize a defined performance metric.

Machine learning approaches are commonly categorized into supervised, unsupervised, and reinforcement learning [49]. Supervised learning relies on labeled training data, where models learn mappings between input–output pairs and are later used to infer outputs for new inputs. Unsupervised learning focuses on discovering structure in unlabeled data, often under algebraic or probabilistic assumptions. Reinforcement learning lies between supervised and unsupervised learning, as feedback is provided in the form of rewards or penalties rather than explicit output labels, enabling agents to learn optimal behaviors through interaction with an environment.

2.3 RE and Machine Learning

Organizations across diverse domains are increasingly incorporating machine learning components into software systems to support their operations [70]. In machine learning–enabled systems, data plays a central role in determining system behavior, often replacing or complementing traditional rule-based code.

Machine learning components are now widely used in domains such as healthcare, agribusiness, and commerce [77, 26, 57]. Machine learning–enabled or machine learning–based AI systems are software systems that incorporate ML components as part of their functionality. Although the terminology may vary across contexts, these systems share common characteristics, including data dependency, uncertainty, and iterative development.

Given the challenges associated with developing ML-enabled systems, several studies have identified specific RE issues in this context [71, 5]. Much of the existing research at the intersection of RE and ML focuses on using ML techniques to support RE activities, rather than on adapting RE practices to the characteristics of ML-based systems [71]. Among RE activities, elicitation and design receive greater attention, while activities such as validation and requirements management are often underexplored, potentially affecting system quality and sustainability.

The ML community has demonstrated that developing and deploying ML systems can be relatively fast, whereas maintaining them is often difficult and costly. RE is frequently considered one of the most challenging activities in ML projects [40]. The absence of well-structured early engineering activities can further complicate system maintenance. Alves et al. report that, according to data scientists, understanding the problem is one of the most difficult tasks in ML system development [9].

RE for ML-based systems involves addressing several challenges, including understanding data characteristics and distributions [73], comprehending the application domain and context [39], enabling effective stakeholder collaboration [40], managing concerns related to model degradation [72], and reasoning about quality attributes, non-functional requirements, and trade-offs [5, 69, 34].

These challenges highlight the need for improved support for early-stage activities that aim to align data, objectives, and stakeholder expectations in ML-based AI systems.

2.4 Design Thinking

Design Thinking (DT) is a structured, human-centered approach to problem-solving that has been widely adopted to support innovation and the development of novel products and services [37]. DT emphasizes the exploration of user needs and integrates an agile and flexible environment to address complex, ill-defined, and uncertain problems. It relies on iterative problem reframing, non-technical prototyping, and interdisciplinary collaboration, fostering creativity and supporting improved engineering practices and requirements-related activities [36].

DT is commonly described through three high-level phases: understanding, exploring, and materializing. During the understanding phase, the focus is on developing empathy with users and stakeholders to capture their needs, expectations, and pain points, followed by the definition of the core problem to be addressed. The exploring phase corresponds to ideation, in which alternative solution concepts are generated and discussed. In the materializing phase, ideas are transformed into prototypes, which are subsequently tested and refined through iterative cycles until a satisfactory solution is achieved.

Several models and frameworks have been proposed to structure and communicate DT activities [33]. Among them, the Double Diamond model, introduced by the Design Council, stands out as a widely adopted representation of the DT process [18]. The model provides a standardized yet adaptable structure that can be tailored to projects with specific characteristics, making it particularly suitable for innovation-driven contexts [33].

The Double Diamond model visually represents the alternation between divergent and convergent thinking. In the first diamond, the discovery phase promotes a broad exploration of the problem space, followed by the definition phase, in which insights are synthesized and the problem to be addressed is clearly articulated. The second diamond begins with another divergence, supporting ideation and exploration of alternative solutions, and concludes with convergence through prototyping, testing, and refinement, ultimately leading to a solution that meets stakeholder expectations (Figure 2.1).

DT has been adopted by organizations such as IBM and SAP to support innovation initiatives [35] and has been shown to assist software engineers in understanding user needs and managing uncertainty [53]. In the context of requirements-related activities, DT supports a human-centered perspective that integrates desirability (user needs), feasibility (technical constraints), and viability (business concerns). Prior studies highlight different perspectives for applying DT in requirements-related contexts, including DT as a process, as a toolbox, and as a human-centered meta-approach [35].

In this research, DT is adopted as a process to support discovery activities, particularly those aimed at problem understanding, stakeholder alignment, and exploration of solution spaces in AI-driven innovation projects.

2.5 ISO/IEC/IEEE 12207:2017

ISO/IEC/IEEE 12207:2017 [1] establishes a common framework for software life cycle processes, defining a set of processes, activities, and tasks applicable throughout the acquisition, development, operation, maintenance, and retirement of software systems. Rather than prescribing a specific development methodology, the standard provides a flexible reference model that organizations can tailor to their context and objectives.

ISO 12207 groups life cycle activities into four main process groups: agreement processes, project-enabling organizational processes, technical management processes, and technical processes. This research focuses on the technical processes that are most

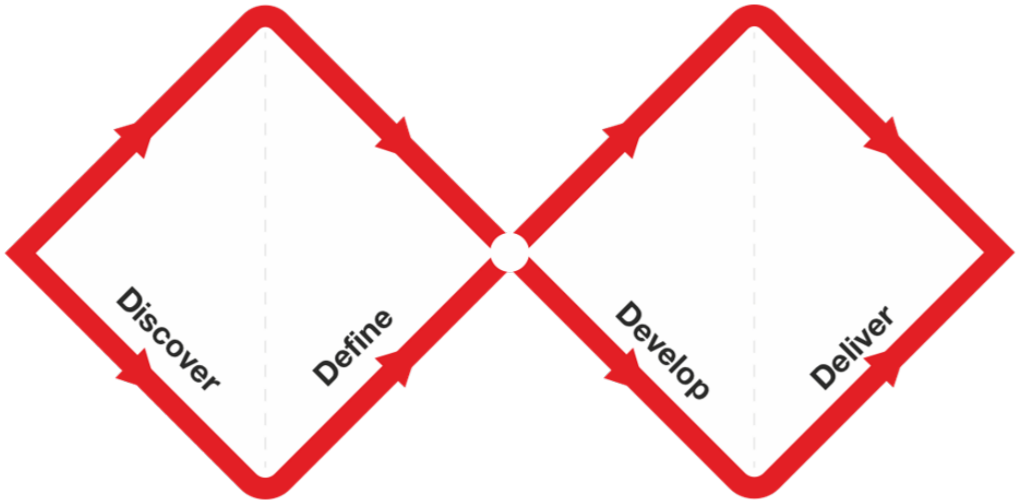

Figure 2.1: *Design Thinking like Double Diamond concepts [18].*

relevant to early-stage discovery activities, particularly those concerned with problem understanding, stakeholder needs, and solution exploration.

Within the technical process group, several processes relate to discovery-oriented concerns, including Business or Mission Analysis, Stakeholder Needs and Requirements Definition, System or Software Requirements Definition, and System Analysis. Collectively, these processes support the identification and characterization of the problem domain, the solution space, relevant stakeholders, usage context, constraints, and key quality attributes.

The Business or Mission Analysis process aims to define the business problem or opportunity, characterize the solution space, and identify alternative solution classes. Its outcomes include the definition of the problem domain, preliminary operational concepts, evaluation of alternative solutions, and establishment of traceability. These outcomes directly support discovery by framing the problem and clarifying decision boundaries.

The Stakeholder Needs and Requirements Definition process focuses on identifying stakeholders, understanding their needs, and defining usage contexts and constraints. Although traditionally associated with RE, its early activities are closely aligned with discovery, as they emphasize stakeholder identification, needs elicitation, prioritization, and scenario exploration.

The System or Software Requirements Definition process transforms stakeholder needs into a technical perspective of the solution. While later stages of this process extend beyond discovery, its initial activities contribute to refining the solution space and clarifying system boundaries and constraints.

The System Analysis process provides analytical support for decision-making by evaluating alternatives, assumptions, and trade-offs. In discovery contexts, this process contributes to assessing feasibility, risks, and implications of alternative solution paths.

Taken together, these ISO 12207 processes highlight the complexity and multi-

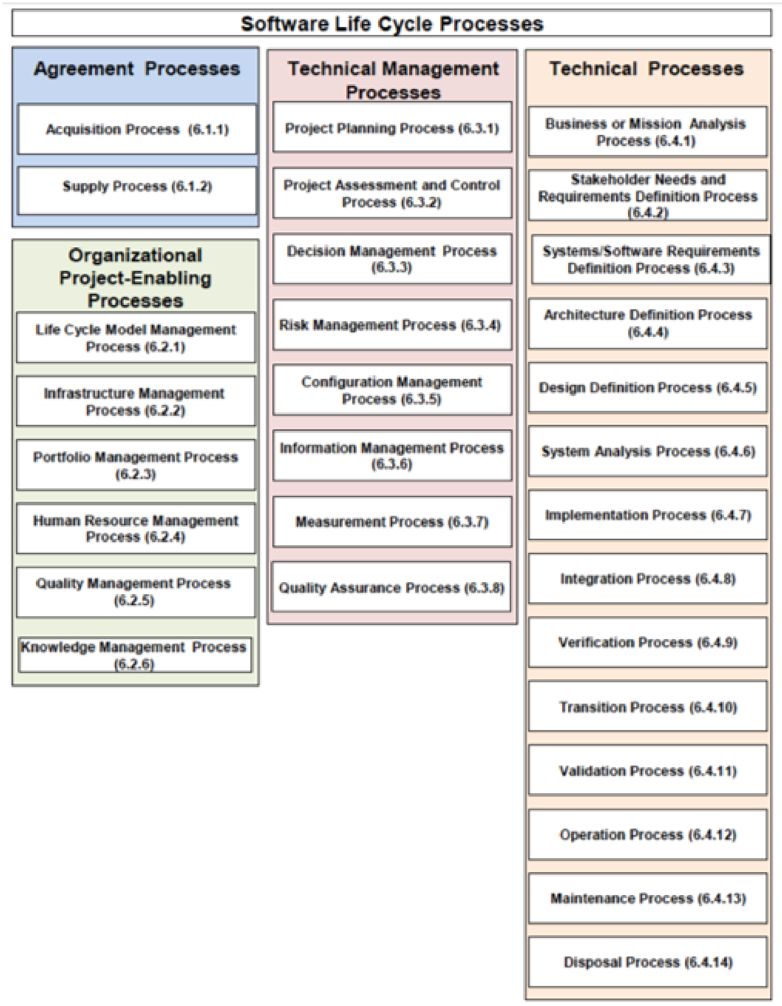

Figure 2.2: *Process of ISO 12207.*

disciplinary nature of early-stage discovery activities. In this research, ISO 12207 is used not as a prescriptive RE process, but as a conceptual and technical reference to inform and structure discovery activities in AI-driven innovation projects. Given the experimental nature of AI systems, the strong dependency on data, and the inherent uncertainty in RDI contexts, an adapted and discovery-oriented interpretation of ISO 12207 provides a systematic foundation for addressing challenges such as problem framing, stakeholder alignment, expectation management, and early decision-making.

Agreement processes are those that apply outside the project's life cycle and are used by organizations that act as acquirers and suppliers of a system. Project-enabling organizational processes provide resources to enable the project to meet the needs and expectations of the organization's stakeholders and are related to the strategic level of the organization that provides human and financial resources, defines and monitors quality measures, and establishes processes and models. Technical management processes are responsible for managing the resources and assets allocated by management and using them to meet agreements related to costs, deadlines, and commitments. To do so, they verify actions, performance targets, and identify and select corrective actions. Finally, technical processes include the technical actions of the life cycle that transform needs into products.

In the group of technical processes, we have three processes that can correspond to the RE phase. They are the business analysis, definition of stakeholder needs and requirements, and definition of system or software requirements processes.

The Business Analysis or Mission Analysis process is the first technical process related to ER. Its purpose is to define the business problem or opportunity, characterize the solution space, and determine the potential classes of solutions that can solve the problem or take advantage of an opportunity. Its expected results are: definition of the problem domain, characterization of the solution domain, definition of preliminary operational concepts, identification and analysis of alternative candidate solutions, selection of preferred candidate solutions, provision of enabling systems, and establishment of traceability. To this end, this process begins with preparation, that is, analysis of problems and opportunities, strategic objectives, capabilities, systems, and products of the organization. Followed by the definition of the problem and solution space and identification of requirements and interfaces of enabling systems and services. Subsequently, the understanding of the scope and definition of context and key parameters for the solution are defined. Next, the main stakeholder groups and alternative solution classes are identified. Then, each alternative solution class is evaluated and the preferred solution class(es) is selected. Finally, it maintains traceability and provides the obtained artifacts and information.

The Stakeholder Needs and Requirements Definition process aims to define the requirements of a system that can provide the necessary capacity to users and other stakeholders in a defined environment. As expected results, this process seeks to: identify system stakeholders, characteristics, context of use, system constraints, define stakeholder needs, define critical performance measures, obtain agreement and enabling systems or services, and ensure traceability. To this end, the process begins with the identification of stakeholders and the strategy to obtain the needs of each stakeholder is established, identifying and acquiring the enabling systems and services of the process. Next, the context of use of the system is defined, the needs of stakeholders are identified, and the needs are prioritized and classified. Subsequently, a representative set of scenarios is defined and factors that affect the interaction between users are identified. Next, the needs are transformed into requirements with the identification of solution constraints, identification of requirements and roles of stakeholders related to critical quality characteristics, and these requirements are defined. Next, the complete set of requirements is analyzed and critical performance measures are defined to assess technical progress, feedback is provided, and questions about the requirements are resolved. Finally, explicit agreement is obtained to confirm that the requirements are expressed correctly and are understandable, traceability is maintained, and artifacts are provided.

The purpose of the System or Software Requirements Definition process is to transform the capabilities desired by the stakeholder, a user-oriented view, into a technical view of a solution that meets the user's operational needs. This process has the following expected results: description of the system or element, interfaces, functions and boundaries for the solution, definition of system requirements and design constraints, definition of critical performance measures, analysis of requirements, obtaining enabling services and maintenance of traceability. To this end, the process begins with the definition of the functional boundary of the system, which includes inputs and responses to users and external systems, analysis and description of interactions, properties and interface constraints, such as: flows, procedures and calls. Next, it defines the strategy for defining the requirements and identifying and obtaining the enabling services. Next, it defines the system requirements by defining each function to be performed by the system, identification of states and modes of operation, definition of implementation constraints, identification of requirements related to risks and quality characteristics, defines requirements with: data elements, structures and formats; interfaces and documentation; interfaces with other systems; non-functional functions and characteristics such as quality and cost; migration approach, schedule, software installation, product acceptance; and requirements attributes such as priority, traceability, test cases, risk, baselines, and verification methods. The set of requirements is then analyzed, critical performance measures are defined to assess technical achievements and provide feedback to stakeholders, problems and conflicts are identified and resolved in the set. Finally, agreement is obtained and traceability is maintained, and the artifacts are provided.

The System Analysis process aims to provide a rigorous basis of data and information for technical understanding to assist decision-making throughout the life cycle. Its expected results are: identification of system analysis needs, validation of analysis premises and results, provision of analysis results for decision-making, provision of enabling services and maintenance of traceability. The process begins with the identification of the problem or issue that requires analysis, which includes technical and functional objectives, quality characteristics, technological maturity and technical risks; identification of stakeholders, selection of analysis methods, obtaining data for analysis and selection and obtaining of enabling services for analysis. After, analysis is carried out by identifying contexts and premises, application of analysis methods, critical analysis of the results according to quality and validity, establishment of conclusions and recommendations and recording of the analysis. Key artifacts are then provided and traceability is maintained.

Given the presentation of these four technical processes linked to the ER phase, it is possible to understand the complexity of this activity. These processes enable the identification of the problem and solution domain, identification of stakeholders, context of use, constraints, identification of needs, prioritization, critical performance measures and the description of the system, interfaces, functions, requirements definition, requirements analysis and traceability is developed. The traditional ER process that comprises the phases of elicitation, analysis, specification, validation and management seeks to cover the activities established in ISO 12207.

ISO 12207 in the context of this research will guide the RE process for the development of AI systems in innovation projects following the DT process. Therefore, ISO 12207 will serve as a technical guide for defining fundamental activities and tasks for the development of AI systems. Thus, we will have a guide of the objectives and expected results of each phase related to RE for AI systems in innovation projects. Since RDI projects in AI systems add many challenges, such as: experimental nature, identification of the objective, alignment of objective with data, management of stakeholder expectations, difficulty in establishing communication and in establishing quality metrics, we believe that an adapted process based on ISO 12207 can be an alternative for addressing these challenges.

2.6 ISO/IEC 5338: AI System Lifecycle Processes

With the rapid adoption of AI and ML in critical software ecosystems, traditional software engineering standards became insufficient to manage the distinct complexities of probabilistic, data-driven systems. To fill this gap, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) established **ISO/IEC 5338** [3]. This international standard provides a comprehensive reference model for AI system lifecycle processes, explicitly defining a structured set of processes, activities, and tasks required to manage the inception, development, deployment, maintenance, and retirement of AI-based applications.

The primary objective of ISO/IEC 5338 is to discipline the unique characteristics of AI components, such as non-determinism, continuous performance degradation (data drift), and severe data dependencies. Unlike traditional, rule-based software where developers explicitly code the application logic, ML-based systems learn patterns directly from data. This fundamental paradigm shift introduces highly specialized workflows that are not explicitly covered in traditional software lifecycle standards. Consequently, ISO/IEC 5338 proposes dedicated process frameworks for:

- **Data Engineering:** Managing data collection, curation, labeling, and preprocessing as continuous, high-quality engineering pipelines.

- **AI Model Training and Evaluation:** Establishing systematic mechanisms for model selection, training, verification, and hyperparameter tuning against strict validation sets.
- **Operational Monitoring and Retraining:** Defining ongoing processes to monitor algorithmic performance in production, detect model drift, and trigger automated or manual retraining loops.

By formalizing these phases, ISO/IEC 5338 provides organizations with a robust framework to address technical debt, transparency, and traceability across the entire AI engineering lifecycle. Within this research, this standard provides the core baseline required to understand how early requirements discovery seamlessly interfaces with specialized downstream data and model engineering tasks.

2.7 Alignment between ISO 12207 and ISO 5338

Despite these necessary differences in downstream implementation and operational maintenance, a closer examination of the early lifecycle stages reveals a critical convergence. Within the technical processes group, foundational engineering activities remain unchanged. Specifically, the **Business Analysis process**, the **Stakeholder Needs and Requirements Definition process**, and the **System or Software Requirements Definition process** do not exhibit structural differentiation between ISO/IEC 5338 and ISO/IEC/IEEE 12207 [8].

This lack of differentiation underscores a fundamental principle of RE: regardless of whether the target system is deterministic or driven by machine learning, the high-level boundaries of problem discovery must be established identically. Both standards dictate that understanding the organizational context, eliciting user expectations, and defining boundaries remain independent of the underlying technological stack [8]. As detailed in Table 2.1, the foundational objectives of these three early technical processes are shared across both frameworks.

This alignment justifies the core architectural design of the DIP-AI framework. By anchoring our discovery phase in these specific technical processes, we leverage a standardized baseline common to both general software engineering and specialized AI lifecycles [8]. While AI introduces unique downstream technical debts, the initial phase of defining the business problem, bounding stakeholder expectations, and formalizing software requirements remains a shared, undifferentiated methodology between ISO/IEC/IEEE 12207 and ISO/IEC 5338 [8].

Table 2.1: *Alignment of Early Technical Processes between ISO 12207 and ISO 5338*

Technical Process	ISO/IEC/IEEE 12207 Focus	ISO/IEC 5338 Focus
Business Analysis	Review organizational problems, scope solutions, and define business cases.	Identical scope; establishes the business viability of an AI intervention [8].
Stakeholder Needs and Requirements Definition	Elicit stakeholder needs, identify constraints, and establish agreement.	Identical scope; defines the non-technical expectations of users and data owners [8].
System/Software Requirements Definition	Specify technical boundaries, operational modes, and system capabilities.	Identical scope; establishes what the software must achieve before modeling begins [8].

2.8 ISO 31000:2018

ISO 31000:2018 [2] is an international standard that provides guidelines for risk management in organizations. Several studies presented by Febiyanti et al. [27] discuss, extend, and critically analyze this standard, highlighting its relevance across different organizational and industrial contexts. ISO 31000 supports organizations in systematically identifying, analyzing, evaluating, monitoring, and controlling risks that may affect the achievement of their objectives. The standard is generic and can be applied to any type of risk and across different organizational domains.

Rather than prescribing mandatory procedures, ISO 31000 offers recommendations for establishing a structured and continuous risk management process. Following the framework proposed by the standard, risk management activities typically include: risk identification, risk analysis to support decision-making regarding mitigation strategies, risk evaluation to determine priorities and treatment needs, risk treatment through appropriate techniques, and continuous monitoring and critical review of risks over time.

In the context of this research, ISO 31000 provides a conceptual foundation for addressing uncertainty and risk during early-stage discovery activities. Its principles support systematic reflection on technical, organizational, and contextual risks that emerge when exploring problem and solution spaces in AI-driven innovation projects.

2.9 ISO/IEC 23894

As Artificial Intelligence (AI) systems introduce unique socioeconomic, technical, and regulatory challenges, conventional risk management frameworks became ill-equipped to address threats such as algorithmic bias, model hallucination, opacity, and data privacy vulnerabilities. To bridge this gap, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) developed **ISO/IEC 23894** [4]. This international standard provides specific guidance on risk management for organizations developing, providing, or using AI technologies. It establishes a set of principles, workflows, and organizational guidelines tailored to manage the distinct uncertainties inherent to the non-deterministic nature of machine learning models.

While ISO/IEC 23894 is fundamentally rooted in the high-level architecture of **ISO 31000** (the generic international standard for risk management), it introduces critical differentiators tailored to the AI domain. Traditional risk management under ISO 31000 focuses on static, broad organizational and operational risks, assuming predictable system boundaries. In contrast, ISO/IEC 23894 operationalizes these principles specifically for the AI lifecycle, detailing risk assessment methodologies that account for continuous data variations, model drift, lack of explainability, and evolving ethical compliance. It moves past generic risk identification to address technical specificities, such as assessing risks associated with the quality of training datasets, security vulnerabilities unique to adversarial attacks, and the societal impact of automated decision-making.

Despite these advanced technical specializations in downstream implementation, a critical realization emerges during the earliest stages of software engineering. The initial identification and characterization of fundamental project risks during the **discovery phase** do not strictly require the granular, AI-specific definitions articulated in ISO/IEC 23894. At the inception of an innovation project, risk management focuses heavily on high-level feasibility, strategic alignment, and basic data availability.

During early RE, characterizing foundational constraints—such as whether data exists, who owns it, or what the core business objective is—can be completely and effectively supported by the broader principles of ISO 31000. Because the discovery phase is concerned with bounding the problem domain rather than optimizing algorithmic hyper-parameters or auditing fully trained neural networks, the generic risk architecture remains entirely sufficient. Therefore, while ISO/IEC 23894 is vital for governance during the development and operational phases of an AI system, early-stage problem discovery and requirement scoping remain agile and robust when anchored simply in foundational risk frameworks.

2.10 GUT Matrix

To support the prioritization of identified problems, the GUT Matrix—based on Gravity, Urgency, and Trend—was applied. This matrix enables a structured assessment of problems according to their severity, immediacy, and likelihood of worsening if left unresolved. As a result, it facilitates the definition of priorities and supports more informed decision-making.

The GUT Matrix is a decision-support tool used to compare and rank problems or situations based on three criteria [28]:

- Gravity, which represents the severity or impact of the problem;
- Urgency, which reflects the time sensitivity for addressing the problem; and
- Trend, which indicates the tendency of the problem to escalate over time.

The matrix is typically organized as a table, in which each problem is listed and evaluated against these three dimensions. A numerical scale, commonly ranging from 1 to 5—where 5 represents the highest level—is used to score each criterion. The scores assigned to Gravity, Urgency, and Trend are then multiplied to produce a single GUT value for each problem. Higher values indicate higher priority for intervention [28].

In this research, the GUT Matrix supports discovery activities by helping stakeholders prioritize problems, risks, and improvement opportunities identified during exploratory phases.

2.11 5W2H

Nakagawa [51] states that the 5W2H tool is so obvious, simple and widely used that there is no agreement on who developed it. It is a management and planning tool that helps define and execute actions in a clear and objective manner [28]. Each letter of the 5W2H represents a question in English, which are as follows:

- What: Refers to the description of the task or action to be performed, that is, what will be done.
- Why: Indicates the justification or reason why the task is being performed, that is, why it will be done.
- Who: Determines the person or people responsible for executing the task, that is, who will do it.
- When: Establish the deadline or deadline to complete the task, that is, when it will be done.
- Where: Defines the place or location where the task will be performed, i.e., where it will be done.
- How: Refers to the method or way in which the task will be performed, i.e., how it will be done.
- How much: Indicates the cost or amount of resources required to perform the task, i.e., how much it will cost.

The application of the 5W2H is useful for making the planning and execution of projects, tasks, and actions clearer, ensuring that all relevant information is provided, assigned, and understood by everyone involved. By answering the 5W2H questions, a team or individual can establish an action plan to be followed and facilitate monitoring and control of the progress of activities [28].

This tool also originated in Japan after World War 2, designed by professionals in the automotive industry with the aim of assisting in the planning stage. This type of tool, because it is simple and objective for developing action plans, is commonly used in project management; business analysis; development of business plans, strategic planning, among other activities related to management.

The 5W2H tool is composed of seven steps?

1. What = What will be done?;
2. Why = Why will it be done?;
3. Where = Where will it be done?;
4. When = When will it be done?;
5. Who = By whom will it be done?;
6. How = How will it be done?;
7. How much = How much will it cost?).

Due to its versatility and dynamism, this tool is used in situations such as quality planning; planning of mergers and/or acquisitions of organizations; planning of human resources areas; planning of product development stages; risk planning, among others.

In this research, 5W2H is used as a complementary tool to support discovery and early planning activities, translating identified problems and insights into actionable and well-defined initiatives. By doing so, it helps bridge the gap between exploratory analysis and structured execution.

2.12 Business Model Canvas and AI Model Canvas

The Business Model Canvas (BMC) was employed in this study as a tool to support the understanding and structuring of value creation in innovation contexts. A business model is defined as the rationale through which an organization creates, delivers, and captures value. The BMC captures this rationale by focusing on three key aspects:

- how core components are integrated to deliver value to customers;
- how these components are interconnected through stakeholder networks; and
- how value is generated through these interconnections.

In the context of AI-driven innovation projects, the Business Model Canvas—and its extensions, such as the AI Model Canvas—supports discovery by aligning technical possibilities with business objectives, stakeholders, and value propositions. These canvases help structure discussions about assumptions, constraints, and dependencies early in the project lifecycle.

2.13 Technology Transfer Model

According to Pfleeger [54], the introduction of new techniques and tools into industrial practice often faces significant challenges, and innovative ideas may take years to become accepted as standard practice. In software engineering, technologies—including methods, models, languages, and tools—are continuously proposed to improve quality and productivity, making effective technology transfer a critical concern.

Successful technology transfer depends on close collaboration between researchers and practitioners. It requires iterative observation, evaluation, and validation of proposed solutions in both academic and industrial settings. Building on earlier work by Pfleeger [54], Gorshek et al. proposed a Technology Transfer Model (TTM) designed to support the development and industrial adoption of research-based solutions in RE.

The model proposed by Gorshek et al. [32], illustrated in Figure 2.3, consists of seven steps. These steps include: identifying industrial needs and improvement areas; formulating research problems; developing candidate solutions in collaboration with industry; performing academic or laboratory validation; conducting static validation through methods such as interviews; executing dynamic validation in industrial contexts, such as pilot projects; and finally, releasing the solution while allowing for incremental refinement.

This model has been successfully applied in previous RE research [70]. In this study, the Technology Transfer Model is adopted as a research method to guide the development, validation, and evolution of the proposed solution. By applying the model, the research aims to ensure that the proposed approach addresses real industrial needs, is empirically validated in multiple contexts, and can be effectively transferred to practice.

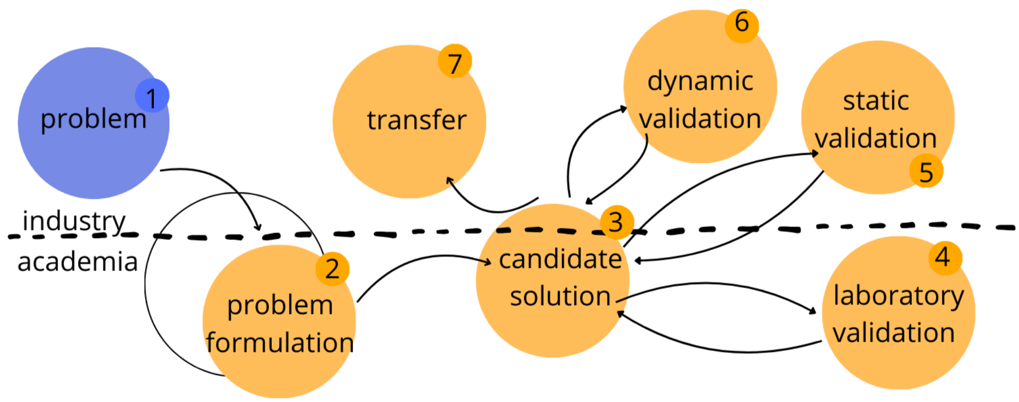

Figure 2.3: *Transfer Technology Model [32].*

2.14 Final Considerations

ISO/IEC 12207 provides a comprehensive set of processes to guide the software life cycle. However, the development of AI-based systems introduces additional challenges, such as high uncertainty, strong data dependency, and an experimental nature, particularly in innovation-driven projects. In this context, discovery activities play a critical role in problem understanding, stakeholder alignment, and exploration of solution spaces.

This research adopts a combination of Design Thinking principles, structured risk management approaches, decision-support tools, and a technology transfer perspective to support discovery in AI-driven projects. The Technology Transfer Model is used to guide the research process itself, ensuring that the proposed solution evolves through iterative validation and remains aligned with industrial needs. Ultimately, the goal is to contribute a discovery-oriented approach that is both theoretically grounded and practically applicable.

CHAPTER 3

Related Work

This chapter presents studies related to the development approach. The focus of this chapter is to show similarities and differences of the approaches found in the literature compared to the proposal of this work. Understanding these studies helps to design approaches and results found in these studies as well as gaps and challenges.

3.1 PerSpecML

PerSpecML [69] is a perspective-based approach to specifying ML-based systems that encompasses 59 concerns related to typical tasks of practitioners involved in defining and structuring these systems. PerSpecML was developed so that stakeholders are able to: increase clarity about goals, concerns, and requirements so that everyone has an understanding of the system; promote collaboration and communication among stakeholders; identify trade-offs with conflicting goals and requirements; improve decision-making by comparing and understanding solutions; and ensure completeness by considering multiple concerns. The concerns are grouped into five perspectives: system goals, user experience, infrastructure, model, and data. These perspectives align activities among business owners, domain experts, designers, software engineers, and data scientists, i.e., the main stakeholders involved in the development of ML systems.

In this approach, the authors described 28 tasks that group the concerns associated with the development [68]. Relationships representing influences and constraints between concerns were also identified. The diagram of tasks and concerns based on perspective is presented in Figure 3.1. The figure also presents the relationships between tasks and highlights the involvement of different stakeholders.

From the diagram, a specification model [70] was developed that provides a standardized format for documenting and organizing the concerns involved in the development of ML systems. The specification model consists of a set of questions that guide the exploration and evaluation of concerns. By analyzing the questions, stakeholders will have a comprehensive view of the project's development requirements. A section of

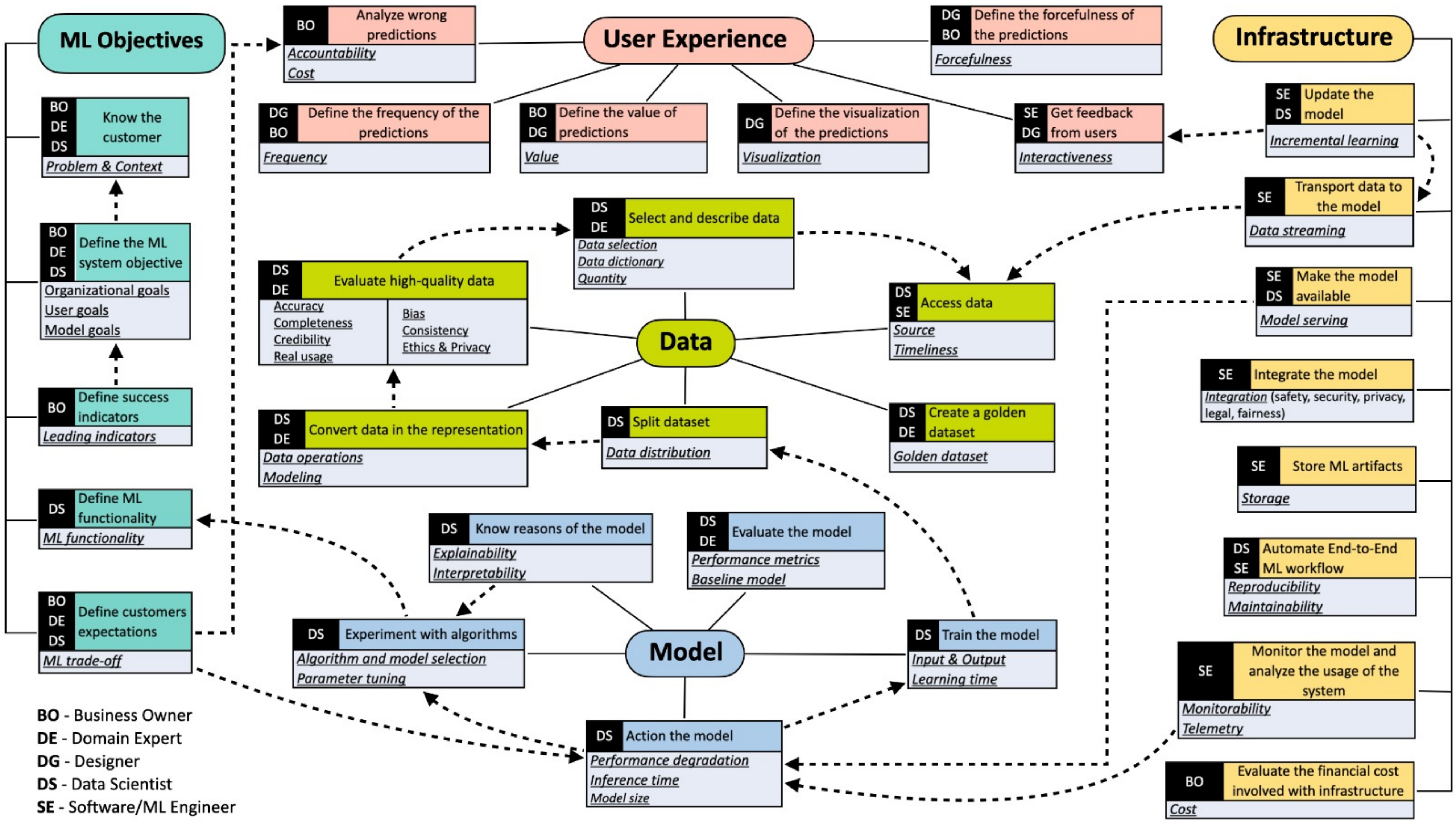

Figure 3.1: *Concerns Diagram of ML Persceptive-Based [70]*

the model is presented in Figure 3.2 that shows the user experience and infrastructure perspectives and some of the issues related to each.

The authors have used the approach in the ExACTa initiative at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), where they carry out agile experimentation and co-creation for digital transformation. The initiative was created by four professors from the main staff of the Department of Computer Science, researchers in the areas of Data Science, Software Engineering, Optimization, and Human-Computer Interaction. The goal of the initiative is to strengthen ties between research and development, allowing the application of research results to trigger innovation and operational excellence in partner companies.

Using ExACTa, the authors performed validation with 53 students from Software Engineering and Data Science courses at Loggi and 15 Computer Science students at PUC-Rio. Participants used PerSpecML to analyze relevance, integrity, ease of use, usefulness, and intended use. Based on the results that were positive regarding the aspects analyzed, they made improvements and performed static validation. Static validation was performed with ExACTa professionals working on Petrobras projects to characterize industrial relevance, ease of use, usefulness, and intended use of the approach. Three data scientists, two developers, and two project leaders used the approach for retroactive projects. A focus group and questionnaire were conducted to collect data for evaluation. With feedback from users, the authors proceeded to dynamic validation. Dynamic validation was performed with ExACTa professionals working on Americanas projects to characterize specification quality, ease of use, usefulness, and intended use. The validation was performed with two groups for different products, one was for product classification using natural language processing and the other for sales using a recommendation system. Each group had four developers, one scrum master and two data scientists with different backgrounds and experiences. For this validation, they used requirements engineering workshops and interviews, and the feedbacks helped in the improvements for the published version of PerSpecML. As a result, the approach was evaluated as useful, easy to use and with intended industrial use.

Although PerSpecML offers a comprehensive, perspective-based architecture its operational scope introduces a lifecycle limitation that motivates the creation of the DIP-AI framework. The key point of divergence lies in the baseline assumptions of both approaches: PerSpecML is fundamentally designed as a specification model, meaning it relies on a standardized, question-driven format that assumes the core business problem has already been well-defined and bounded by stakeholders. It lacks explicit methodological support for the highly fluid, early-stage activities where the problem itself is still unknown or ambiguous. In contrast, the DIP-AI framework is specifically engineered to support the upstream *discovery* and exploration phases of innovation

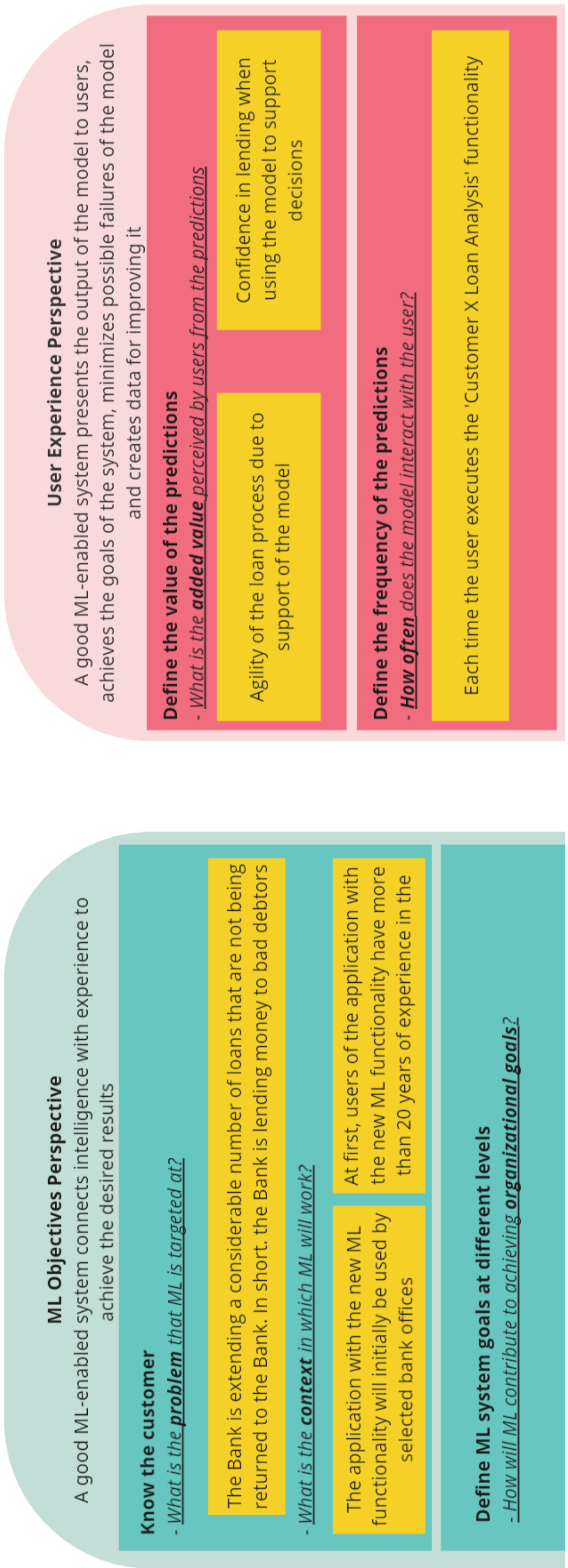

Figure 3.2: *Specification Model for EM-Based Systems - A Perspective-Based Approach*

projects. While PerSpecML excels at documenting and organizing specific technical constraints once the project boundaries are set, DIP-AI bridges the initial gap by guiding multidisciplinary teams through the complex process of problem space exploration, feasibility assessment, and strategic alignment before specification takes place.

3.2 ML Canvas

With the aim of having a canvas that would provide an overview of the complexity of developing a ML system, Louis Dorard [22] developed the ML Canvas. It is a framework based on the *Lean Canvas* and *Business Model Canvas* in order to assist in the development of ML systems by facilitating collaboration between stakeholders. The author argues that the canvas can be used to: *refine* ideas by describing how the ML system will transform predictions into value for users, what data it will learn from, and how to ensure that it will work as expected; *collaborate* when building high-value systems that involve different stakeholders concerned with product, business, data science, etc. in a way that aids communication; and *plan* in advance, since the canvas allows you to anticipate costs, identify bottlenecks, specify requirements, and create a roadmap.

Structurally, the ML Canvas architecture is bifurcated into two specialized functional regions that map out the operational mechanics of the system. The left-hand side of the framework is strictly dedicated to *Predictions*, providing four interconnected blocks to define how machine learning models generate value from input data prior to and immediately following deployment. Within this region, the **ML Task** block specifies the core algorithmic approach (e.g., classification or regression) along with its precise input-output boundaries, while the **Decisions** block articulates how these resulting predictions are translated into actionable value for the end-user. Furthermore, the **Making Predictions** block defines the operational constraints, such as the timing and latency requirements of new inferences, which is complemented by the **Offline Evaluation** block to establish the metrics and methods required to assess model performance before live deployment. Conversely, the right-hand side of the canvas focuses on *Learning from Data*, capturing the backend engineering pipeline through four distinct components: the **Data Sources** and **Collecting Data** blocks, which map available raw inputs and the mechanisms to continuously capture new training data, alongside the **Features** block to outline the specific input representations extracted from these sources. Finally, this learning lifecycle is bounded by the **Building Models** block, which details the scheduling, constraints, and time allocation required to train and iteratively update the underlying machine learning models with new data.

The ML Canvas is a framework that allows you to describe the value proposition in a central block, describing what, why and who will use the system. The sub-frame on

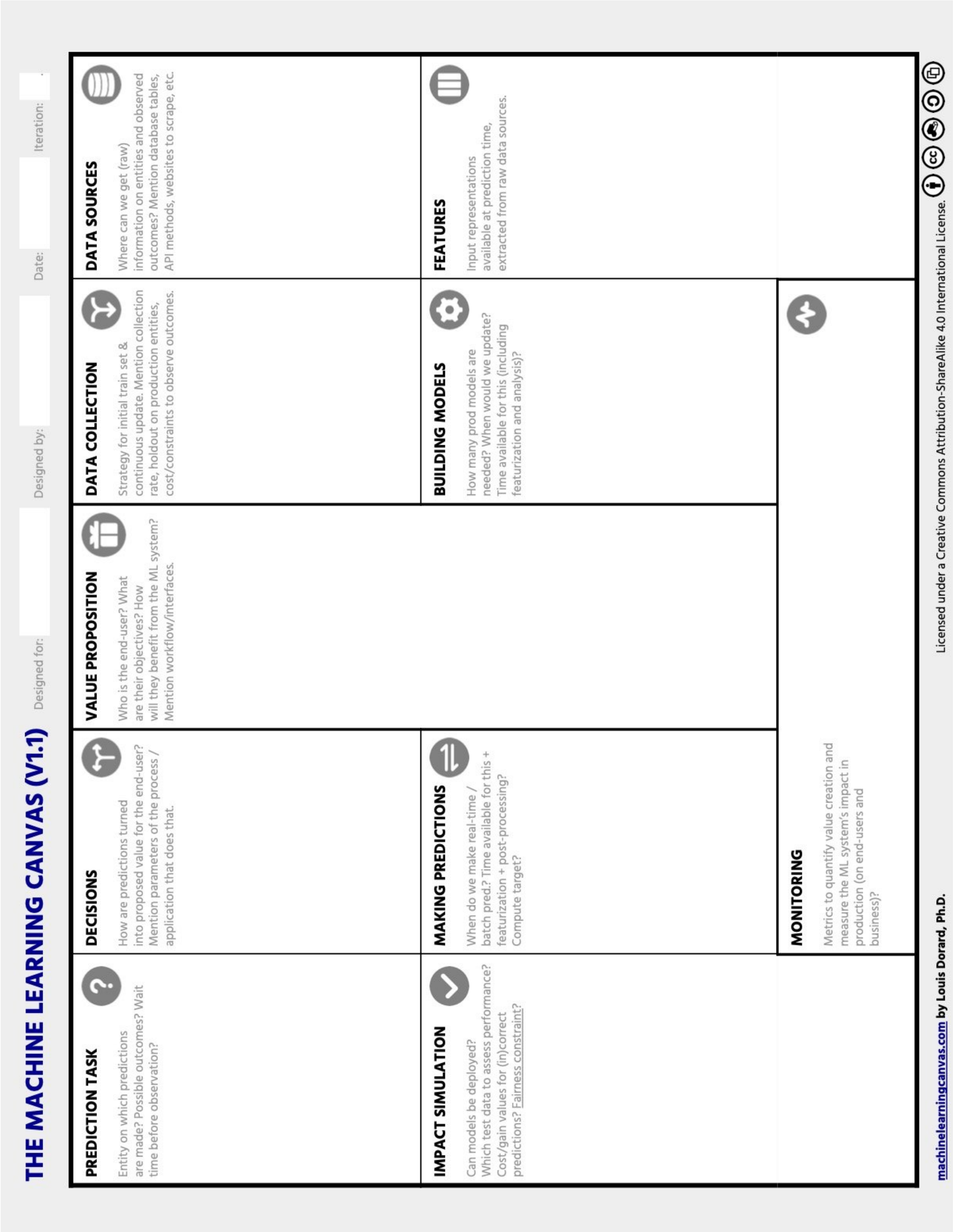

Figure 3.3: *ML Canvas - ML Project Board*

the left is for describing predictions and the frame on the right is dedicated to learning. Regarding predictions, information about data sources, data collection, resources and model construction will be filled in. The learning sub-frame allows you to describe the task, decisions, when predictions will be made and the time available for this, methods and metrics for evaluation. Finally, the lower sub-frame is dedicated to measuring how well the system works, that is, methods and metrics for evaluating the system after implementation. Figure 3.3 shows the ML Canvas framework and its sub-frames.

Dorard shows that the approach has been used since 2016 by many teams around the world and highlights AWS, ESADE Business School, Hitachi, UCL, Deutsche Telekom. However, the author does not show when, by whom it was used or the results of its use. It also does not present an evaluation of the approach as a study.

Similarly to PerSpecML, the ML Canvas is inherently built upon the prerequisite assumption that the core business problem has already been identified and validated. While Dorard's framework provides a highly collaborative, high-level blueprint to map out how machine learning predictions translate into user value and how data pipelines operate, it expects stakeholders to approach the canvas with a predefined objective. It lacks the upstream, structured mechanisms required to explore an ambiguous problem space, assess alternative non-AI solutions, or systematically weigh initial business risks. This shared limitation across the literature—where both conceptual canvases like the ML Canvas and rich specification models like PerSpecML focus heavily on the *how* of the ML implementation rather than the *why* of the business problem—serves as the primary motivation behind the development of the DIP-AI framework. DIP-AI was specifically engineered to fill this methodological void, shifting the focus to early-stage problem discovery and open exploration. By providing teams with the tools to dissect and boundary the problem before committing to an architectural stack, DIP-AI acts as a necessary precursor to frameworks like the ML Canvas, ensuring that teams do not prematurely optimize complex machine learning models for ill-defined or non-viable business problems.

3.3 Lean Design Thinking Methodology for Data Projects

Ahmed et al. [7] presents a software development methodology - *Software Development Methodology* - (SDM) given that the *Cross-Industry Standard Process for Data Mining* (CRISP-DM) does not meet the challenges of working with current technologies such as ML [55]. Given the scarce literature on the use of methodologies, *frameworks* and processes that could allow teams to complete projects with more resources and success

[62].

Given the above, the authors combined the strengths of CRISP-DM, DT and *Lean*. As a result, the authors proposed a three-step approach: business, data and product. To reiterate, LDTM works by combining: *Design Thinking* (to understand the customer/user and discover the business need), with *Lean Startup* (to evolve the model/solution) and CRISP-DM (to develop the algorithmic/technical elements of the model/solution) according to Figure 3.4. The goal of LDTM is to allow development teams the opportunity to routinely improve a data model, acting iteratively and incrementally based on accuracy statistics and *user feedback* by combining experimentation and greater understanding of requirements in a single approach.

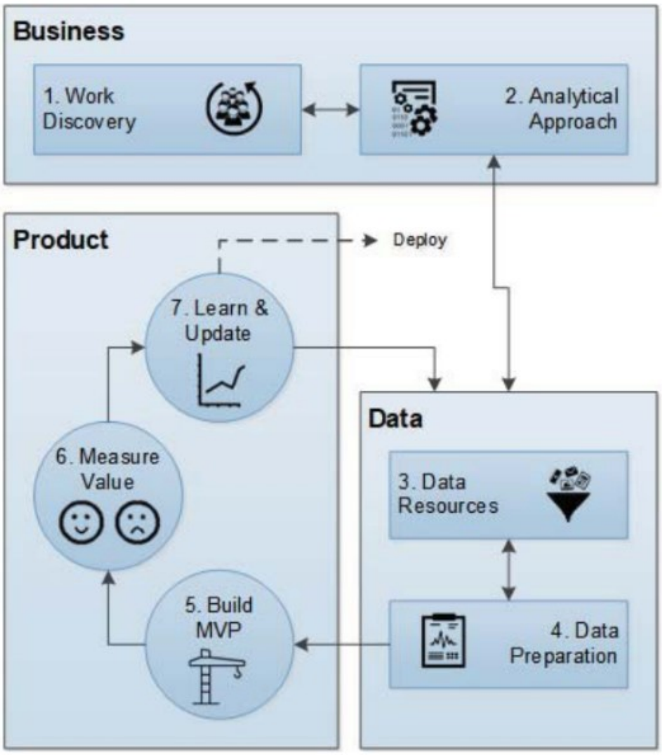

Figure 3.4: *LDTM Process for data project [7].*

Design thinking is an approach to problem-solving that is carried out through ideation. Ideas are converted into prototypes before being tested and evaluated. Lean Startup is a model for how to manage a startup by helping to find market fit and delivering a minimum viable product (MVP) through the build, measure, and learn cycle. When building products, development teams start by building a minimum set of features that satisfy some users. They then test hypotheses made about these features and measure feedback obtained from experiments. In this way, decisions are made based on evidence that directs the progression of product development for subsequent iterations. This process is repeated in continuous cycles until product-market fit is achieved. CRISP-DM is a comprehensive methodology for conducting data mining projects divided into six phases:

business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Efforts to update it emerged in the late 2000s in light of its limitations, but no updates have yet been identified [46].

The proposed approach is divided into six phases: work discovery - where teams get to know the client and user and understand their problems; analytical approach - where the team articulates the work discovery in the context of statistical and ML techniques and already discusses possible solutions; data resources - where the team identifies and collects data resources relevant to the problem; data preparation - where the team builds the dataset; MVP construction - where the team builds a solution in the simplest version that can be implemented and tested quickly; value measurement - MVP is implemented and *feedback* is collected; learning and updating - the last phase where the team learns from the delivery of the MVP and makes decisions about changes and makes those changes.

This approach does not address specific aspects of the requirements, targets an MVP that may not be the final product of an innovation project, and does not define how each of the phases will be developed. Typically, professionals do not know which methods or tools can support development, and this knowledge is not clear in the literature [47, 71].

3.4 Comparison of Related Works

When analyzing the related works presented in this chapter, we summarize the different aspects in the Table 3.1. There are different approaches regarding some stages of RE applied to the development of ML systems. However, there is still a gap in an RE process for this type of system. Since RE is considered the most difficult activity in the development of ML systems [40], we believe that an adapted process can help improve the quality of development and the products delivered. Among the proposals analyzed, only two were used in the practice of projects in partnership with the industry [70, 22]. And only one of the proposals reports its evaluation method and results [70]. In view of this, we sought to evaluate the process to report on the use of the process that can allow teams to complete projects with more resources and success, addressing the gap in the state of the art [71, 62].

When analyzing the related works presented in this chapter, we summarized the different construction aspects in the Table 3.2. Therefore, we did not identify a proposal that adopts ISO 12207 for the development of ML or AI systems. We believe that ISO 12207 serves as a technical guide for defining fundamental activities and tasks for the development of ML systems in innovation projects. As RDI projects in ML Systems add many challenges, such as: experimental nature, identification of the objective, alignment of objective with data, management of stakeholder expectations, difficulty in establishing communication and in establishing quality metrics. We believe that an adapted process based on ISO 12207 can be an alternative for addressing these challenges.

Among the proposals analyzed, we did not find an adapted ER process that assists in all RE phases for the development of ML systems. As already presented, RE is considered the most difficult phase of the ML systems development cycle [40]. Therefore, we believe that adequate treatment of all RE can assist in higher quality development and delivery [52, 65].

We also did not identify an approach that helps identify the problem. Problem identification is a crucial part and the biggest challenge identified in the state of the art in AI systems development. Problem identification is usually performed before the proposals are implemented. Therefore, when implementing the proposals, it is expected that the problem has already been defined. However, when dealing with AI systems development in innovation projects, it is often not clear what the problem is or what data is related to that problem, as well as whether this data is available. We believe that the application of ISO 12207, which establishes a business analysis process where the problem is identified, is one of the distinguishing features of this proposal.

Finally, most of the proposals did not perform or did not report the evaluation. To discuss the usefulness of the RE process for AI systems in innovation projects, we sought to evaluate our approach with professionals and students involved in innovation projects at CEIA. We will seek to assess whether the process helps in development, whether it is easy to use, useful, and whether the evaluation participants express a desire to continue using this process.

Thus, the proposal of this research is the most complete regarding the RE process for AI in innovation projects. We will use ISO 12207 to direct the essential activities and tasks to be addressed in RE, seeking to go through all phases of the RE cycle, we will assist in identifying the problem, the biggest challenge pointed out by the state of the art and we will carry out the evaluation in 3 instances following the MTT. Some of the related works do not deal with processes, but rather approaches that cover some activities of the RE phases. We identified only one process based on ISO 12207, however, it is defined for the development of IoT and does not address aspects of model construction and maintenance or aspects of data collection and processing.

3.5 Final Considerations

We identified some works related to our proposal. However, none of the works combines the ISO 12207 approaches and covers all phases of requirements engineering. In addition, some works are not specific to ML or even to data. Given the above, we

Propose	Whats is it?	Why is it necessary?	Where was it used?	When was it used?	Wh
PerSpecML	Approach to elicitation and specification	Helps in elicitation and specification adressing SE concerns beyond the model and data	ExACTa in partnership with Loggi, Americanas e Petrobrás	Since 2023	Stud prac
ML Canvas	Framework for elicitation and specification	Assists in elicitation and specification adressing aspects of the model and data	AWS, ESADE, Hitachi, UCL e Deutsche Telekom	Since 2016	Expe
LDTM	Methodology for developing and improving data model iteratively and incrementally	Helps elicitation and validation of RE and lifecycle of ML	-	-	-
DIP-AI	Process Helps in all RE phases, uses ISO 12207 e DT	Helps in all RE phases in innovation projects in AI systems	Used in projects in CEIA and ExACTa	From 2024	stud expe

Table 3.1: *Comparative Table of Related Works - Part 1*

	Adopts ISO 12207	RE Phases	Identify Problem	Performed evaluation
PerSpecML	No	elicitation and specification	No	Yes
ML Canvas	Not	elicitation and specification	No	Not
LDTM	No	elicitation and validation	Yes	Not
DIP-AI	Yes	elicitation, analysis, specification, validation and management	Yes	Yes

Table 3.2: *Comparative Table of Related Works - Part 2 (evaluated)*

believe that our proposal is significant for the state of the art and practice of requirements engineering for innovation projects involving AI.

CHAPTER 4

Interviews with Experts

This chapter presents the empirical study conducted to ground the development of our proposal. This chapter 4 details the results of interviews held with AI project coordinators, aiming to characterize RE practices, as well as to identify recurring difficulties and opportunities [47].

4.1 Planning

The primary objective is to investigate the state of the practice of RE in AI-based Research, Development, and Innovation (RDI) projects within the EMBRAPII Unit. We applied the GQM (Goal-Question-Metric) method [66] to define the following research parameters:

- **Object of study:** RE practices in CEIA innovation projects.
- **Objective:** To characterize the current state of RE practices in AI innovation projects.
- **Quality Focus:** Process characteristics, methodologies, and artifacts related to the RE phase.
- **Point of view:** Innovation project leaders and researchers.
- **Environment:** The Center of Excellence in Artificial Intelligence (CEIA-EMBRAPII-UFG), Brazil.

4.1.1 Research questions

1. How is requirements elicitation performed in AI projects?
2. How is requirements analysis performed in AI projects?
3. How is requirements specification performed in AI projects?
4. How is requirements validation performed in AI projects?
5. How is requirements management performed in AI projects?
6. How are non-functional requirements handled in AI projects?

7. Are the best practices of RE in projects achieved due to their leaders having more experience or greater knowledge of RE?
8. Are the best practices of RE in projects achieved due to their teams having more experience or greater knowledge of RE?
9. What challenges in RE are faced in these projects?
10. What was generally neglected in the RE phase in these projects?

This work was submitted to the Ethics Committee on 28/02/2023 and approved under CAAE: *67531823.4.0000.5083*.

4.1.2 Data Collection

Initially, some AI systems project coordinators were invited to respond to the interview. More details about the interview setup are available in the survey link at Zenodo[1]. During the interview, the context of the research, presentation of the researchers and application of the Free and Informed Consent Form (FICF) were explained. The questions organized in a script were asked based on the participant's answers. The interview was then conducted with authorized recording.

4.1.3 Data Analysis

Data analysis was performed using the Grounded Theory (GT) method. Grounded Theory has proven to be useful in several fields, the method described by Glaser and Strauss [30] aims to generate theories rather than test or validate existing theory. Widely used in qualitative data analysis and more recently widely used in software engineering research [64].

We followed the Straussian approach to GT [31]. In this approach, the research question is open-ended and defined based on the literature. Literature is consulted throughout the process to understand concepts and information that may be useful in data analysis. The coding process is determined by the open, axial and selective phases, which generate categories and relationships. Finally, the central category and links are defined.

The interviews began in mid-June and were completed in mid-December. Some interviews had to be split into two parts due to the participants' time constraints. During the interviews, we transcribed and coded the data. A follow-up interview was conducted with the coordinator after he had taken on more projects and gained more experience. For each interview record, we uploaded them to the **Reshape** [58] tool, which automatically transcribes and edits texts with transcription issues. We downloaded the documents and uploaded them to the **Atlas.TI** [13] tool.

[1] https://zenodo.org/records/10580470

Coord.	Team	AI Team	TRL	Domain	Learning	Task	SE expert	RE expert
C1	17	25%	3	health	S	P	yes	no
C2	22	40%	6	health, industry	U and SS	P and D	yes	yes
C3	10	20%	6	business, HR	S	P	no	no
C4	20	80%	5	health, legal, energy	S, U and SS	P, D, RS and SA	no	no
C5	11	45%	7	health	S	IR	yes	yes
C6	18	80%	4	robotics	S	P, IR, RN	no	no
C7	10	30%	5	business	S and U	P, D, RS and SA	yes	yes
C8	13	60%	3	health	S and U	IR and RN	no	no
C9	12	50%	7	agribusiness and market	S, U and R	P and RS	no	no
C10	14	35%	7	industry	S	P	yes	yes

Table 4.1: *Project characterization. Note: Learning paradigms: S = Supervised, U = Unsupervised, SS = Semi-supervised. Task types: P = Prediction, D = Detection, RN = Recommendation, RS = Regression, SA = Sentiment Analysis, IR = Information Retrieval.*

Each document related to an interview was coded. The document was read in full, marking items in the answers that referred to a specific question. The idea was to identify similarities between the coordinators in order to group the coded items and establish what RE practice was like in these projects, testing the hypotheses defined for this research.

We conducted the interviews with ten coordinators and then transcribed them using the Reshape tool. A total of 555 minutes were recorded. The interviews, on average, lasted just under sixty minutes.

4.2 Results

4.2.1 Characterization

To characterize the unit that conducts research in AI for RDI projects, we sought to collect as much information as possible. To evaluate and relate the research results to the profile, we asked about the projects and also about the coordinator's training.

Project

Regarding the project, we distinguished teams with varying numbers of members. Teams of six to twenty-five people with varying degrees of training and expertise comprised the groups. The contributions made by the AI teams also varied; in some teams, only 16% of the team worked on AI-related tasks, while in others, 100% of the team was directly involved with the AI solution. It is imperative to emphasize that teams exhibit volatility throughout a project, depending on the requirements of each team, the phases of the project, and the number of participants.

Technology Maturity Levels (TRL) are a type of measurement system used to determine the degree of maturity of a specific technology [24]. The TRL levels of the consulted projects ranged from 3 to 7, with level 6 being the most typical. The lowest TRL is 1 and the highest TRL is 9. Technical coordinators, who supervised smaller teams tasked with completing specific project tasks, assisted the coordinators interviewed. Each project had one to three technical coordinators. Regarding the agile approach used in the research. Most coordinators said they use modified Scrum, modified Kanban, and a combination of agile approaches in their projects. No coordinator acknowledged the use of non-agile methods. Although most of the projects led by the coordinators interviewed were related to health, we also had research on business, industrial management, process improvement in organizations, law, and agribusiness.

Most of the projects consulted used supervised and unsupervised learning approaches. We also found examples of reinforcement and semi-supervised learning. Most of the tasks assigned to these projects were related to prediction. We also found systems for recognition, diagnosis, regression, signal processing, robot navigation, prescription, and recommendation. There are not many participants in most of the projects who are experts in requirements engineering. Most experts in requirements or software engineering are project coordinators or technical coordinators.

Coordinator

ID	SE Knowledge	RE Knowledge	CP	AI Coord.	AI Industry	SE Exp.
C1	very understanding	understand	1	1	2	Since 2005
C2	very understanding	very understanding	2	1	-	Since 2005
C3	understand	understand	3	-	15	
C4	little understanding	middle understanding	5	13	18	
C5	very understanding	very understanding	2	1	-	Since 2018
C6	understand	middle understanding	1	13	14	Since 2009
C7	understand	understand	1	1	4	Since 2013
C8	little understanding	little understanding	2	10	2	Since 2010
C9	do not understand	little understanding	2	10	-	
C10	middle understanding	middle understanding	1	-	-	Since 2006

Table 4.2: *Profile of the coordinators interviewed.*

Most coordinators work on one or two projects (see in Table 4.2 in cloumn CP or Coordinated Projects). There are 20 projects in total for the 10 coordinators surveyed. Half of the coordinators have previous experience working on AI projects in industry, ranging from two to eighteen years of experience. Three of the coordinators reported one to thirteen years of experience coordinating previous AI projects.

4.2.2 RE Practice

All coordinators performed the elicitation phase, seven coordinators performed analysis, five declared performing specification and validation. And only one coordinator responded regarding the practice of requirements management in his/her project.

4.2.3 RE Practices in AI Projects

The interviews revealed a significant disparity in the adoption of RE phases across projects. While all ten coordinators ($n = 10$) performed elicitation, the subsequent stages showed a decline in formal adoption: seven conducted analysis, five performed specification and validation, and only one reported formal requirements management.

Elicitation

Elicitation practices are primarily centered on stakeholder communication. The most prevalent technique was meetings ($n = 8$), followed by brainstorming sessions ($n = 5$). Other methods included interviews, prototyping, business analysis, and document analysis. Regarding stakeholders, coordinators primarily consulted clients ($n = 4$), users ($n = 2$), and domain experts ($n = 1$). Collaborative tools such as Google Meet, Google Docs, and Figma were frequently cited as essential support for these activities.

Analysis and Specification

The analysis phase is characterized by a lack of standardization. Practices identified include requirement classification ($n = 2$), prioritization, modeling, and grouping. Tools such as flowcharts and use cases were mentioned as means to facilitate communication between the team and the client.

Regarding specification, there is no consensus on documentation standards. Approaches varied from natural language and User Stories ($n = 2$) to the use of the ISO/IEC 29148 standard ($n = 1$). One participant mentioned the use of *ML Canvas* for AI-specific needs. Specification artifacts typically include task descriptions, author records, and unique identifiers (IDs). Management tools like Clickcopy, Gitscrum, and Google Sheets are often repurposed to serve as ad-hoc requirement repositories.

Validation and Management

Validation is heavily reliant on prototyping ($n = 5$), with simulation and peer review being less frequent. Six coordinators emphasized the importance of client consultation during this stage. In contrast, requirements management is the most neglected phase; only one coordinator reported systematic management, using Google Sheets to track requirements throughout the lifecycle.

ID	Elicitation	Analysis	Specification	Validation	Management	Agile M.	NFR	Quality evaluation	Progress evaluation	Changing manag.	Lessons learned
C1	brainstorming, prototype and meet client	priorization, classification and modeling	user stories in docs & figma; ISO/IEC 29148 based	review, prototype	-	scrum	usability	usability and acceptance criteria	big commit	docs, gitscrum, weekly meet	yes
C2	documental analysis and meet client	priorization and classification	user stories in docs & Jira	review	-	scrum and kanban	usability	validation with client	work plan	weekly meet	no
C3	brainstorming, documental analysis and process analysis	-	in ML Canva, gitscrum	-	-	scrum	data exploration	proof of concept	weekly meet and big commit	-	no
C4	-	-	-	-	-	scrum and kanban	RGPD and data security	accuracy	weekly meet and big commit	weekly meet	no
C5	interview, meet client and brainstorming	fluxogram	-	prototype	-	scrum		-	-	weekly meet	no
C6	meet client, documental analysis	priorization	-	simulation, review	-	kanban	accuracy	accuracy	weekly meet	weekly meet	no
C7	meet client	use case	clickcopy	prototype	spreadsheets	scrum	model quality	satisfaction questionnaire	weekly meet	weekly meet	yes
C8	observation, brainstorming, paper and meet client	priorization	in spreadsheet and gitscrum	prototype	-	scrum	accuracy and image quality	accuracy	big cocmmit	-	yes
C9	brainstorming and meet clint	-	-	review, prototype	-	scrum	accuracy, scalability and sustainability	accuracy and response time	big commit	-	no
C10	interview, brainstorming and prototype	-	-	-	-	scrum	performance	paper metrics	big commit	weekly meet	yes

Table 4.3: *RE Practice in the CEIA projects.*

Transversal Activities and Quality Assessment

Project alignment is mostly achieved through periodic meetings; six coordinators hold change management meetings, and three maintain weekly agendas for team synchronization. Progress assessment is predominantly based on macro-deliverables ($n = 7$), though some coordinators ($n = 3$) utilize model accuracy as a proxy for project progress.

A critical gap was identified in the documentation of lessons learned: while four coordinators document them, none reported revisiting or sharing these documents. Furthermore, quality assessment remains informal. Activities are often limited to customer acceptance criteria and satisfaction questionnaires. Notably, two coordinators admitted to not performing any quality assessment due to a lack of knowledge regarding software quality metrics.

Non-Functional Requirements (NFRs)

The coordinators demonstrated a clear focus on technical NFRs, specifically **model accuracy** and **performance**. Other mentioned requirements included response time, data security, privacy, and usability. Emerging AI-specific concerns—such as sustainability, scalability, and data exploration—were mentioned only sporadically. This suggests that while technical performance is prioritized, broader software quality attributes are often secondary.

4.3 RE4AI and Challenges

Proposed Mitigation Strategies

To address the aforementioned challenges, the coordinators suggested a series of interventions focused on process adaptation and structured documentation. These suggestions provide a foundation for the requirements framework proposed in this thesis:

- **Process and Method Extension:** A consensus emerged regarding the need to extend existing RE methods or develop new ones tailored for AI. Specifically, coordinators highlighted the importance of incorporating artifacts like *Model Cards* to document model behavior and limitations.
- **Contextual Adaptation (RDI and TRL):** Participants emphasized that any RE process for AI must be dynamic and adaptable to the project's *Technology Readiness Level* (TRL) and the experimental nature of Research, Development, and Innovation (RDI).

Coord.	Better elicitation and spec.	Better data req. and RNF	Difficulty in project	Reduce difficulties
C1	yes, extension process and methods	yes, establish metrics	quality commit and discovery process to develop	process RE4ML
C2	yes, do not know how maybe extension process and methods and knowledge RE and patterns	not sure know according TRL	obtaining data and lack of experience	process RE4ML, specification forms
C3	no, adapted for RDI	no	team and infrastructure and RDI nature	document for communication
C4	yes, adapted for agile maybe extension process and other RE methods	yes aligning data to objective	team and infrastructure and obtaining data	RE agile and process RE4ML
C5	yes, other RE methods	yes, efficiency model and aligning data to objective	obtaining data	-
C6	yes, extension process	yes aligning data to objective	team and infrastructure and expectations clients	RE agile
C7	yes, use model cards	yes, model cards	financial and expectations clients and obtaining data	document for communication
C8	not sure know, adapted for agile	not sure know simplified method	time management and communication with team and RDI nature	process RE4AI
C9	yes, adapted for agile and extension of methods	yes, simplified method	obtaining data and expectations clients	document for communication and validation forms
C10	yes, knowledge RE and extension of methods	yes aligning data to objective	team, obtaining data, expectations clients	document for communication

Table 4.4: *RE aspects that can be improved in ML projects at CEIA.*

- **Structured Documentation as a Solution:** Most coordinators ($n > 8$) pointed out that implementing a formal process or standardized documentation is the primary way to reduce technical debt and improve team-client alignment.
- **Enhanced Elicitation and NFR Treatment:** Nearly all interviewees acknowledged that current elicitation and specification techniques—particularly for Non-Functional Requirements (NFR)—are insufficient and require domain-specific refinement for AI.

4.3.1 Synthesis of Interview Findings

The empirical data collected through these interviews confirms a significant gap between traditional Requirements Engineering and the practical needs of AI innovation projects. The difficulties identified—ranging from the "discovery"of a suitable development process to the complexities of data acquisition—stem from the lack of a structured approach that accounts for the RDI nature of these projects.

Almost all participants expressed a belief that elicitation, specification, and NFR management can be significantly improved through the adoption of agile yet rigorous RE practices. These findings validate the motivation for this research, suggesting that a tailored RE framework for AI can mitigate common pitfalls, streamline team communication, and better manage stakeholder expectations. Thus, the results presented in this chapter serve as the empirical baseline for the proposal described in Chapter 6.

4.4 Threats to Validity

Following the guidelines by Wohlin et al. [76], we identified and mitigated potential threats to the validity of the interviews across four dimensions:

- **Descriptive Validity:** To mitigate the risk of incomplete data gathering or transcription errors, all interviews were video-recorded, stored securely, and meticulously transcribed. The transcripts were subsequently annotated and reviewed using the Reshape tool. To minimize structural bias, a precise, semi-structured protocol was followed. This protocol underwent prior pilot testing with an innovation project coordinator to ensure clarity; questions were solely reordered and enumerated for better scannability without altering their core substance.
- **Interpretation Validity:** To prevent mutual misunderstandings regarding core domains, the research objectives and fundamental Requirements Engineering (RE) concepts were explicitly communicated to the participants at the beginning of each session. Furthermore, the interview guide was iteratively peer-reviewed to enhance its reliability and alignment with the study's scope.

- **Researcher Bias and Theoretical Validity:** To prevent the risk of forcing data to conform to an initial theory, the researchers maintained a strictly neutral stance. As an exploratory investigation into this domain, there was no predefined methodology being promoted, ensuring high openness to emergent, unbiased qualitative outcomes.
- **Reactivity and External Validity:** While participant reactivity due to the researcher's presence cannot be entirely eliminated, it was mitigated by emphasizing the anonymity, non-obligatory nature, and direct benefits of the study—specifically, identifying real organizational needs to drive operational improvement. Finally, a threat to external validity is that the study was bounded to a single Brazilian innovation unit due to feasibility constraints. We mitigate this by advocating for future replications across diverse industrial settings to broaden the generalizability of our findings.

4.5 Final Considerations

The findings presented in this chapter establish a critical bridge between the theoretical gaps identified in the literature and the practical challenges faced by experts in the field. By synthesizing the state of the art with the empirical data from the interviews, we have outlined the core requirements and constraints that define the scope of this research.

The literature review indicates that while specific RE proposals for AI are emerging, many traditional stages remain neglected or poorly adapted. Crucially, there is a notable absence of an RE process specifically designed to navigate the high-uncertainty environment of **AI-driven innovation projects**. The interviews with project coordinators corroborated this deficiency, revealing that the primary obstacles—such as data acquisition, stakeholder expectation management, and the experimental nature of RDI—often stem from the lack of a structured yet flexible methodology.

Furthermore, the data underscores that the "experience gap"in AI project coordination can be mitigated by standardized processes and artifacts (e.g., *Model Cards* and TRL-aligned specifications). It is evident that an adapted RE framework tailored for the AI innovation lifecycle can significantly enhance project predictability, facilitate team-client communication, and ultimately improve the quality of project deliverables.

In light of these considerations, the following chapter introduces a tertiary study designed to identify specific challenges and research opportunities in to requirements and innovation in AI state of art.

CHAPTER 5

Tertiary Study

In Chapter 5, we present a systematic tertiary study of the RE literature for AI. The primary objective of this chapter is twofold: first, to establish a rigorous and replicable research protocol that details the planning, execution, and selection criteria used to map existing secondary studies; and second, to synthesize the empirical results derived from this literature mapping. By analyzing the state of the art through this systematic lens, this chapter identifies recurring challenges, current methodological gaps, and the limitations of existing RE frameworks when applied to AI and Machine Learning ecosystems. Ultimately, these synthesized findings provide the empirical baseline and theoretical justification that motivate the design requirements of our proposed solution.

5.1 Protocol

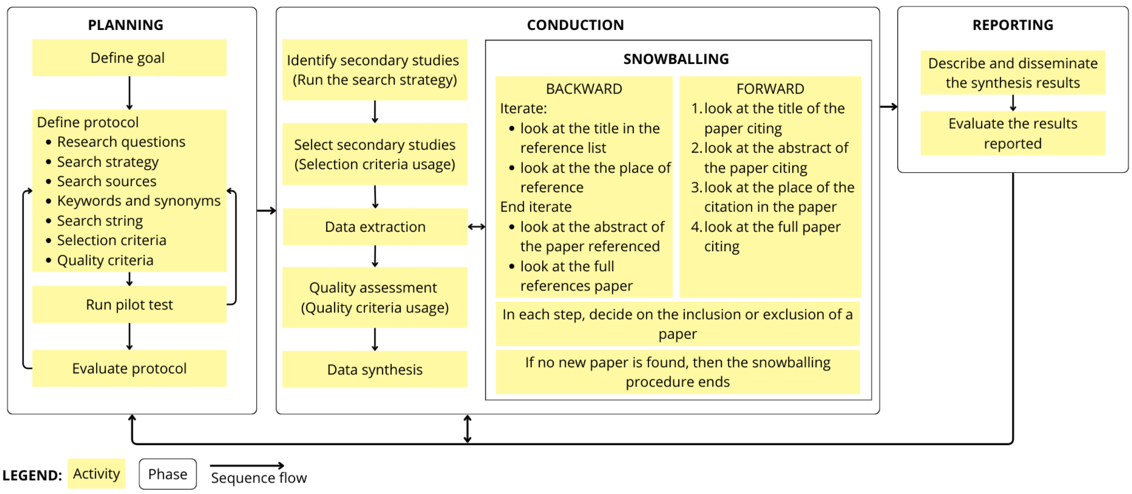

Figure 5.1: *A detailed view of the identification and selection processes of secondary studies.*

This tertiary study aims to consolidate a body of knowledge on RE for ML-based systems, drawing on secondary studies. Figure 5.2 presents the entire process, from protocol conception to data synthesis, collected from relevant secondary studies. This section describes the tertiary study protocol. Further details about this protocol can be seen elsewhere[1].

5.1.1 Objective

This tertiary study identifies the techniques, processes, tools, and metrics reported by secondary studies—such as Systematic Literature Reviews and Systematic Mappings—in the field of RE for AI.

5.1.2 Research questions, pilot research and research string

In accordance with the objective of this study, we formulated the following research questions (RQ):
RQ1: What is the state of the art of RE for ML, including the most researched topics on this topic?
Motivation: While the intersection of Requirements Engineering and Machine Learning (RE4ML) has seen an increase in publications, the literature remains fragmented across different domains, venues, and terminologies. Investigating RQ1 is essential to map the current baseline of the field, categorize the dominant research trends, and identify which specific lifecycle phases or application domains have received the most academic attention. This taxonomy establishes the foundation necessary to understand where the community has focused its efforts and where it has not.

RQ2: What are the challenges and gaps highlighted by the literature on RE for ML?
Motivation: The probabilistic and data-driven nature of ML introduces unique challenges—such as handling non-determinism, data drift, and complex non-functional requirements (e.g., explainability and fairness)—that traditional RE methods fail to address. Investigating RQ2 allows us to systematically consolidate the practical bottlenecks, methodological friction points, and open gaps reported in secondary literature. Pinpointing these explicit limitations is critical, as they serve as the empirical requirements and direct justification for designing new, lightweight frameworks tailored to the industry.

This study specifically identifies research and proposals regarding the application of requirements engineering for ML (RQ1), as well as the main RE approaches, strategies, and proposals for AI and the contexts in which they are being employed, as well as the tools and support functions involved in the system development process. We confirmed the difficulties and potential directions for future research on RE for AI through RQ2. To

[1] https://zenodo.org/records/14617471

address these study concerns, however, we employed a suitable search approach and as many pertinent secondary articles as we could.

5.1.3 Research String

To limit the search to Requirements Engineering within ML systems, the search string was created. Through pilot testing, we improved the search string to narrow it down to the field of RE for ML. Here's how we narrowed our search string:

The search string defined was: ***(("machine learning"OR "artificial intelligence"OR "data-driven") AND ("requirements engineering"OR "requirements elicitation"OR "requirements analysis"OR "requirements specification"OR "requirements validation"OR "requirements management") AND ("systematic review"OR "systematic literature review"OR "systematic mapping"OR "systematic literature mapping"OR "literature search")).***

The validation of the search string was performed by two researchers with experience in systematic studies in RE. The pilot testing process and other details are available at link.[2]

Six sources, including digital databases and search engines, were used to perform the automatic search. In addition, the *forward snowballing* method involved selecting and analyzing the citations of each secondary study to create the collection of studies relevant to this investigation.

We included ACM DL, IEEE Xplore, Scopus, Engineering Village, Wiley, and Web of Science among the sources chosen for automatic searching. On October 18, 2023, the automatic search was performed in these databases. We extracted the duplicate articles for each source from the automatic search.

5.1.4 Inclusion and Exclusion Criteria

We formulated selection criteria consisting of inclusion criteria (IC), exclusion criteria (EC), and the snowballing technique [45]. The IC and EC were applied to all 193 studies. Initially, we checked whether each non-duplicate study met at least one of the exclusion criteria. We considered the following ECs for our protocol:

- EC1: Full text not freely available on the Web or on the CAPES Journals Platform;
- EC2: Not written in English;
- EC3: Not a secondary study;
- EC4: Does not address requirements engineering for AI or ML;

[2] https://docs.google.com/document/d/1Vduulh-YHfHXXSRz7fMeHMyVZQERqiIZOlFFtDH0l_U/edit?usp=sharing

- EC5: Is a preliminary or summarized version of another article.

5.1.5 Extraction Form

In this section, we describe the information extracted during the complete reading of the selected secondary studies. We used a data extraction form based on the RQs. For each secondary study analyzed, relevant information related to each publication was extracted. The list of this information can be found below.

- Publication vehicle;
- Year of publication;
- Type of publication;
- Search source;
- Number of requests;
- How many search sources;
- Which search strategies;
- How many primary studies analyzed;
- What is the objective of the secondary study;
- What are the contributions of each secondary study;
- What are the contributions of Requirements Engineering for AI or ML, such as techniques, processes, tools and metrics;
- What is the maturity report of primary studies of Requirements Engineering for AI and ML?
- What are the challenges pointed out in each secondary study?
- Future work directed in each secondary study?

5.1.6 Quality Assessment

An important decision when performing a systematic tertiary study is to check the quality of secondary studies. The Centre for Reviews and Dissemination (CDR) maintains a database of systematic reviews in Medicine selected according to pre-established quality criteria [15]. These same quality criteria can also be used to assess systematic studies in the SE area [42, 19, 17].

In this tertiary study, we analyze each secondary study through eight questions — or quality criteria (QC) — based on the CDR criteria. The acceptable responses for each question are Y (Yes), P (Partially), and N (No). The following is the list of quality assessment-related questions.

QC1- Are the inclusion criteria (IC) appropriately described?

QC2- Are the exclusion criteria (EC) appropriately described?

QC3- Does the search cover all relevant studies?

QC4- Is the quality or validity of the included primary studies assessed?

QC5- Are primary studies adequately described?

QC6- Is the justification for the secondary study duly described?

QC7- Is the protocol validation properly described?

QC8- Is data extraction properly described and appropriate?

Considering the response to each QC, we assign the following score: two for ‘Yes’, one for ‘Partly’, and zero for ‘No’. The final quality score of each secondary study is the arithmetic mean of the scores for its quality questions. That quality score allows secondary studies to rank through the analysis of their protocol (QC), which is useful for guiding the synthesis activity of the secondary studies.

5.1.7 Data Collection

The data extraction process took place in two distinct stages. First, we extracted information from the selected articles through an automatic search. Then, we read the selected articles using the snowballing technique. To avoid any bias in the analysis of each publication, the data extraction stage was conducted by three researchers. Each of the researchers, during the extraction process, performed an independent analysis of each article. After reading the selected publications, the researchers met to compare and consolidate the extracted data.

Figure 5.2 depicts the whole process performed in our tertiary study, including protocol definition, studies identification through automatic search (**193**), duplicate studies removal (**-89**), studies selection via metadata reading and inclusion and exclusion criteria (**-95**), studies selection based on full reading of papers (**-5**), backward (**+5**) and forward (**0**) snowballing, quality assessment, and data synthesis (**9**). Therefore, nine secondary studies represent the state of the art on RE for ML-based systems.

After the full reading of the 12 papers identified (seven from automatic search and five from snowballing), three papers were eliminated, and nine remained for data extraction and synthesis. Still regarding studies selection, Table 5.1 summarizes the number of studies removed as duplicates and those excluded by a particular EC. We cut out a great number of duplicates (243), mainly due to the overlap between study sources and the recursive nature of snowballing. As primary studies represent most of the papers' references and citations during snowballing, hundreds of papers (789) were removed by EC3. Finally, the number of studies excluded by EC4 (155) means that lots of secondary studies exist in related themes but not specifically about RE and ML-based systems.

Table 5.1: *The number of studies removed as duplicates and applying an exclusion criterion.*

	Duplicate	EC1	EC2	EC3	EC4	EC5	Total
Automatic Search	89	2	0	31	62	0	184
1st Backward Snowballing	99	0	0	566	38	0	703
2nd Backward Snowballing	23	0	0	219	29	0	271
Forward Snowballing	31	2	1	73	15	0	122
Total	**242**	**4**	**1**	**889**	**144**	**0**	**1280**

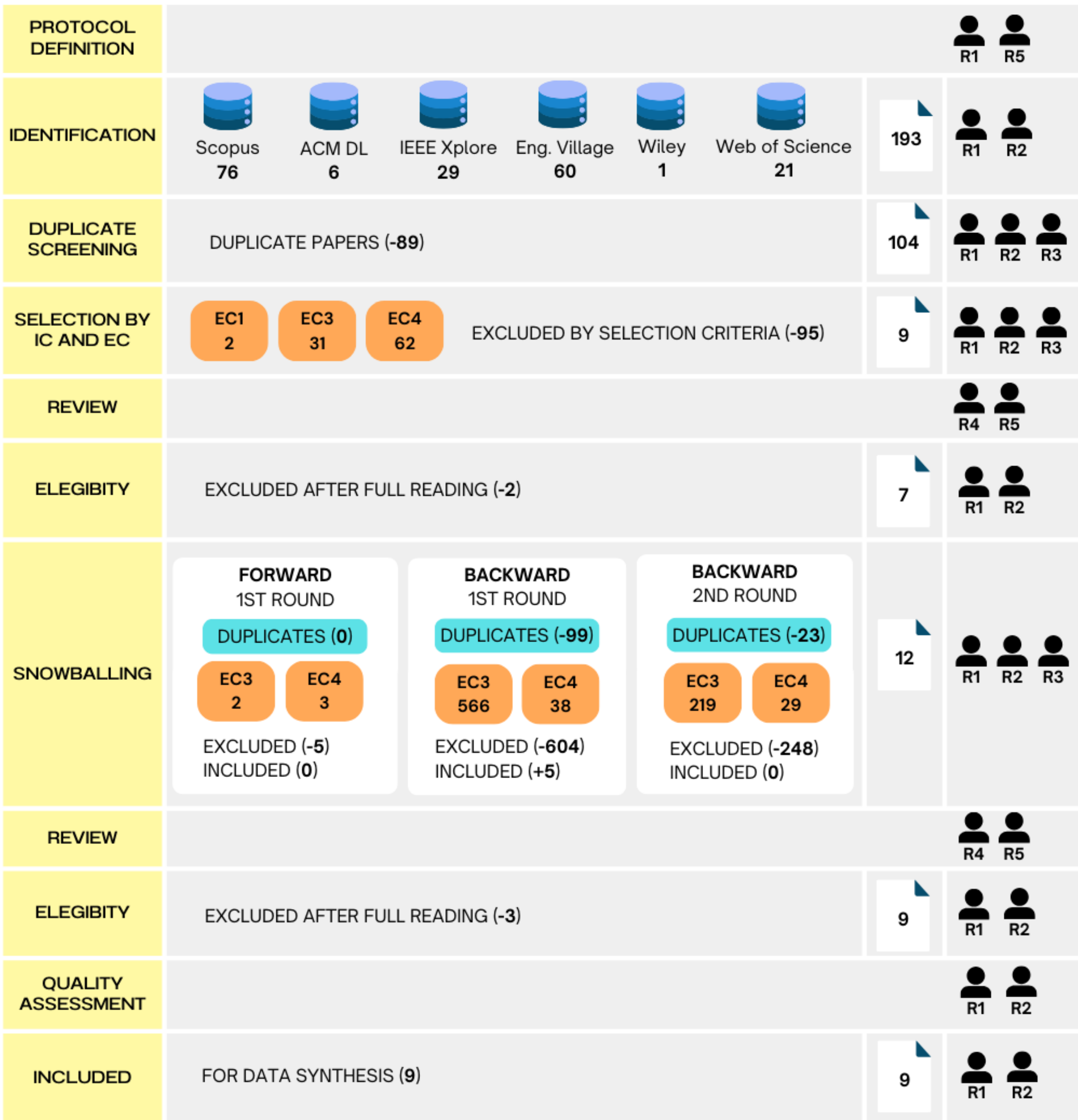

Figure 5.2: *A detailed view of the identification and selection processes of secondary studies.*

5.1.8 Results of Tertiary Study

This section presents a summary of the data extracted regarding the QPs of this tertiary study.

We extracted information from nine studies: three published in 2023, five published in 2021 and one published in 2019. The group of authors who contributed the most is formed by Klood Ahmad, Muneero Bano, John Grundy and Mohamed Abdelrazek. These authors are affiliated with Australia and, therefore, this was the country that contributed the most publications (two publications). Japan was another country that also had two publications and was the country that contributed the most to the topic. In addition to these, we also analyzed publications from Brazil, Germany, the United States, Turkey, and the Republic of Korea. Five of the nine studies are from journals. Only one publication had more than one, the Journal of Systems and Software with two publications, both from 2021. The articles with the highest quality indexes are from journals. These articles include two search strategies and use six automatic search sources.

Research Question 1

Regarding the research question: **”What is the state of the art of RE for ML and AI?”**

The state of the art of research involving RE applied to AI or AI is characterized in this tertiary study involving five systematic reviews, two mappings, one meta-review and one literature survey. Seven studies were applied to AI and two were more general and dealt with AT. Four studies dealt with RE, three with software engineering, one with security and one with explainability. The search strings were characterized by the conjunction of synonyms of RE and AI or ML. The study involving explainability included strings with keywords specific to the topic, such as: explainability, ethics, justice, morality, success, transparency, responsibility, etc. As well as the study involving security, which presented keywords such as: safe, robust, risk, reliability and security.

We identified that all studies carry out a discussion and synthesis of the findings, highlighting the results for the elicitation, analysis and specification phases.

For the **elicitation** phase, Study S3 shows that the most used technique is brainstorming and ad hoc methods are used for elicitation. Study S2 discusses that the elicitation discussed in the primary studies presents concerns with data requirements and NFR that were also found in studies S4, S5, S7 and S9. Among the studies, the NFRs of ethics, trust and explainability (E1, E2), accuracy (E4), reliability (E7), robustness and legal requirements (E5), security, interpretability, fairness and privacy (E8, S9) stand out. Study S2 also points out challenges in enabling communication and collaboration between stakeholders, which is corroborated by studies S3, E5, E7, E8 and E9. Study S9 highlights that a concern for AI-based AI systems is the importance of dealing with uncertainty and managing customer and user expectations, as also found in studies S3, S4 and S6. Study S5 also highlights the importance of obtaining new data sources, identifying data-sensitive characteristics, and obtaining quantitative markers for the quality of this type of system.

Studies S5 and S6, which deal with security and explainability, respectively, point out the importance of identifying situations that require explainability, as well as defining stakeholders and user characteristics. Study E5 contradicts E7 by stating that problem and domain definition have received less attention in the primary studies analyzed.

Regarding the **analysis** phase, studies S2, S5, and S9 identified that the language models most used by this type of system have been UML and GORE. Study S1 also points out that others were also cited in primary studies, such as SysML, ontoML, activity diagram, and semiformal UML model. Study S1 also presents the tools used for modeling: jUCMNav and two others developed by the researchers, Rius and Sirius. Study S1 also points out frameworks for analysis, holistic DevOps, and Ethics Aware. Study S5 presents the importance of discussing performance measures in connection with NFR. And Study S6 also presents the need to perform trade-off analyses, for example, fairness vs. accuracy or performance vs. explainability.

As for the **specification** phase, studies S5 and S7 highlight the importance of specifying data requirements, model requirements, and process requirements. Study S5 also presents the importance of documenting the dataset, versioning the model, and maintaining traceability. Study S6 outlines the importance of identifying the essential information that should be conveyed and to whom, establishing a correlation with the characteristics of the users identified during the elicitation phase. Study S9 demonstrates that, given the inherent characteristics of AI-based systems, the specification could be stated as a hypothesis to be tested as an experiment.

Study S3 shows that it is not clear in the primary studies analyzed which methods and tools are used in RE. Study S3 was the only one that presented four studies on **verification and validation**. And the **management** phase was not mentioned in the studies. The most investigated areas deal with autonomous driving and computer vision, according to studies S1 and S2. However, study E4 describes that the most investigated areas were: autonomous driving, aviation and health. Studies S1, S2 and S3 present the empirical evaluation methods that have been most commonly used in RE research for AI-based systems. Study S3 shows that there is a lack of RE techniques and tools for AI-based systems, and that there is a lack of evaluation to know if the existing methods and tools serve this type of system. However, study S7 discusses that although studies present the use of *ad hoc* methods in *startups*, there are claims that these practices improve performance, quality of deliveries and developer satisfaction. Finally, study S8 presents the need for adaptation in the RE process for AI-based systems.

Research Question 2

Regarding the research question: **"What are the challenges and gaps highlighted by the literature on RE for ML and AI?"**

The challenges highlighted in the secondary studies analyzed are: defining, addressing, and evaluating non-functional requirements such as ethics, explainability (S1, S2, S4, S5), security, fairness, interpretability (S8), privacy (S8, S9), and data requirements (S1, S2, S4). Studies S1, S3, and S5 highlight the challenge of integrating the development team and stakeholders and ensuring that everyone has adequate understanding. The challenges of understanding the domain and navigating uncertainty are highlighted in studies S3 and S4, allowing customers and users to understand and engage in the process with appropriate expectations. Studies S1 and S2 indicate the need to introduce new methodologies. Study S3 indicated a shortage of proposals. Study S3 also demonstrates how the lack of formal evaluation affects the certainty that improvised approaches are suitable for defining requirements in AI-based systems. Study S5 presents the challenge of addressing traceability and specifying metrics and acceptance criteria, which is corroborated by study S9. Study S7 and Study S9 present challenges related to the lack of a process and deadline pressure that may affect the adherence of RE with *ad hoc* methods for this type of system. This contribution is complemented by study S8, which argues that the requirements of AI-based systems change rapidly and therefore *ad hoc* techniques may not be appropriate. Study S9 shows that it is difficult to work together with industry and academia and evaluate different situations. This contribution is corroborated by studies S2 and S3, which present the challenge of evaluating different scenarios.

Future work of secondary studies can be divided into: bibliographical studies and research papers and solution proposals. Studies S3 and S8 do not present suggestions for future work. Studies S4, S6, S7 and S9 present initiatives for bibliographic studies as future work. Study S4 intends to expand its study by adopting gray literature and adding *feedback* from professionals to its analysis. Study S6 indicates future work to conduct a meta-review. S7 is a study that aggregates gray literature, case studies and interviews with professionals. And study S9 indicates a multifocal review as future work. Study S1 seeks to evaluate techniques in different areas. Studies S2 indicate future work to develop proposals for documentation and modeling language and a platform for stakeholder collaboration, action and communication. Study S5 presents future work proposing a methodology to guide the work of security teams.

5.2 Tertiary Study Considerations

According to studies, the most investigated areas deal with autonomous driving and computer vision. However, other study indicates that the most investigated areas were autonomous driving, aviation, and healthcare. Empirical evaluation methods are the most commonly used in RE research for ML-based AI systems. A lack of RE techniques and tools for ML-based AI systems, as well as a lack of evaluation to determine whether existing methods and tools are suitable for this type of system.

On the other hand, ad-hoc methods used in startups improve performance, delivery quality, and developer satisfaction. Unfortunately, those methods are not presented in the paper. Exists the need to adapt the RE process for ML-based AI systems. The most of contributions are to the elicitation, analysis and specification, highlighting structures, tools, and needs of these phases. Other point considered is treat trade-offs and explaining them in a hypothesis format.

Finally, studies identified that the most investigated activity was elicitation, with brainstorming being the most used technique. These studies highlight uncertainty about which existing approaches and tools are suitable for this type of system, suggesting that empirical assessments applied to different scenarios could provide clarity.

In Figure 5.3 we highlight the main contributions pointed out by the secondary studies according to RE activity [14].

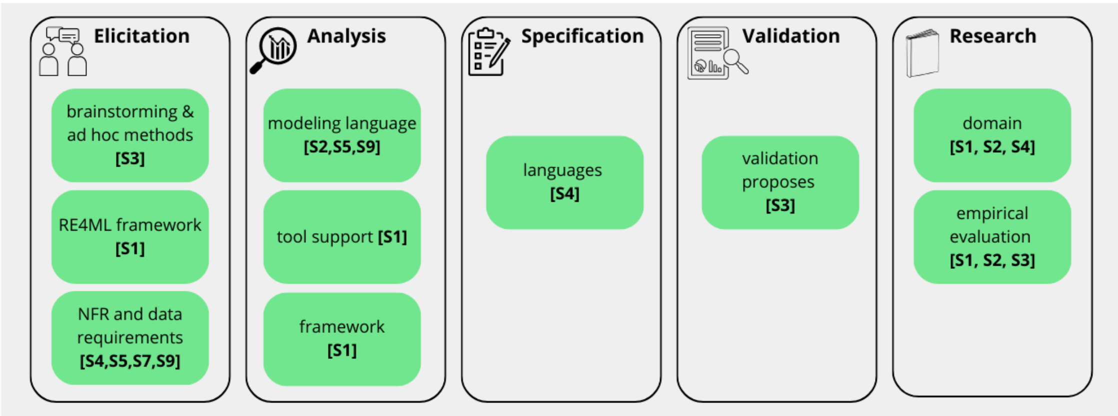

Figure 5.3: *Synthesis of main contributions found in secondary studies.*

Figure 5.4 depicts the main challenges identified in our analysis. Regarding requirements elicitation, we highlight collaborative communication between stakeholders (including data scientists) and stakeholders' expectations management due to the uncertainty of AI-based solutions, and data requirements and NFRs (e.g., security, explainability, ethics, trust, fairness, privacy, legal requirements, etc.).

For the requirements analysis activity, we give special importance to discussing performance measures with stakeholders and performing trade-off analysis, especially aligned with NFRs. Concerning the requirements specification, we point out the significance of specifying data, model, and process requirements in a versionable and traceable manner, and the alignment between specifications and the type of stakeholder.

The analysis of secondary studies calls attention to the low number of studies concerned with requirements validation and management. Challenges related are the definition and specification of performance measures and acceptance criteria for these requirements activities and the establishment of traceability between requirements artifacts. Finally, as challenges for future RE4ML research, studies refer to a tailored RE process for data-driven systems development, the proposal of new RE-oriented techniques and tools, and experimental use of new and existing RE approaches regarding their suitability for the ML domain.

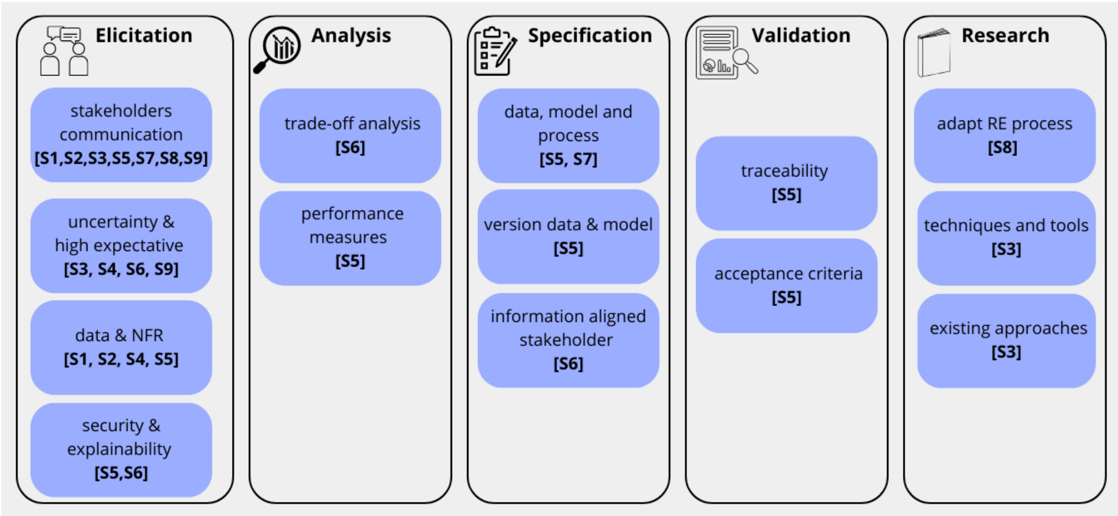

Figure 5.4: *Synthesis of challenges discussed in secondary studies.*

By synthesizing our findings, we identified key areas that require further investigation. Based on these insights, we propose a research agenda encompassing nine critical topics that address the challenges in ML-based systems. The Figure 5.5 presents the main points that will be discussed throughout this section.

A fundamental aspect of ML systems lies in their intrinsic characteristics, including **data properties**, **model features**, **ML processes**, and **infrastructure**. Defining data requirements is crucial to ensure completeness, representativeness, ethical considerations, privacy compliance, and fair distribution of samples. Equally important is the specification of model requirements, such as explainability constraints, performance limitations affecting algorithm selection, relevant performance metrics, execution and learning times, and concerns related to model complexity and degradation. Infrastructure considerations should also be clearly defined, covering aspects such as model provisioning, artifact storage, monitoring, integration needs, and cost management.

Another critical issue is **explainability**. Different stakeholders have varying levels of understanding about the ML-based system and, consequently, distinct needs regarding the interpretability of ML models. When documenting requirements, it is essential to specify what needs to be explained, to whom, and in what context. While explainability focuses on making model decisions comprehensible to different users, interpretability deals with understanding how the model arrives at its conclusions.

NFRs and **trade-offs** also play a significant role in the development of intelligent systems. Different applications may demand specific considerations related to ethics, security, uncertainty, and robustness. Ensuring that data is unbiased, that the model consistently produces reliable results, and that a certain tolerance for errors in predictions is established are all fundamental aspects that must be addressed. Security, in particular, requires careful attention, as it influences and is influenced by other requirements. This is especially relevant in high-stakes scenarios, such as healthcare, where ML models can impact automated decisions not supervised by humans.

Collaboration among stakeholders is another crucial factor for the success of ML-based solutions. Establishing effective communication and expectation management processes is essential for capturing requirements and analyzing problems from multiple perspectives. **Tools** such as structured frameworks and collaborative platforms can facilitate interaction, helping stakeholders align their understanding of the development process and the resulting ML-based solutions. Improving **traceability** mechanisms is also critical to maintaining quality throughout development. Artifacts should support iterative refinement and allow for continuous revisiting and modification of documented information from the goal and problem identification phase through solution delivery.

To ensure that ML models meet business objectives, it is crucial to define measurable **metrics** and **acceptance criteria**. Additionally, refining RE processes for ML systems is necessary to develop or adapt methodologies for requirement elicitation, analysis, specification, validation, and management. Aligning these processes with the ML-based system lifecycle enables better treatment of critical issues and trade-offs while also facilitating acceptance criteria specification and artifacts traceability.

Finally, **empirical studies** in real-world projects are essential for evidence of the effectiveness of RE approaches in ML-based systems development. Conducting controlled experiments and case studies can offer valuable insights into how RE practices influence the success of AI-driven solutions in industry settings.

By addressing these interconnected areas, we aim to advance research and practice in RE for ML-based systems, fostering the development of more transparent, secure, and efficient AI solutions.

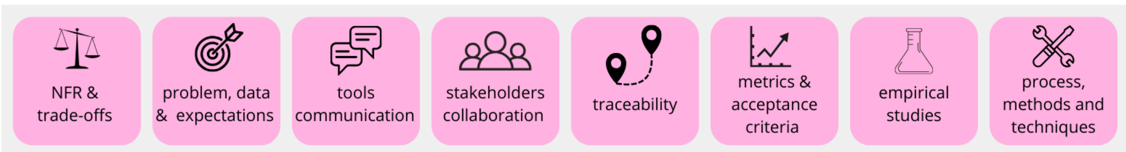

Figure 5.5: *The research agenda's main topics aligned with the challenges collected from secondary studies.*

5.3 Threats to validity

Reporting systematic literature studies presents intrinsic difficulties, e.g., findings are often relevant on a specific topic. We adopted the checklist presented by Ampatzglou et al. [11] to list the actions we took to mitigate threats to the validity of this study.

The tertiary study protocol was proposed by three researchers (R1 to R3) and reviewed by two more experienced ones (R4 and R5). Further information about the tertiary study protocol not available in this paper can be found in the supplementary artifacts.

To mitigate search string-related threats, we selected consensual knowledge about RE, as defined in SWEBOK v4.0, and terms related to AI/ML validated through multiple pilot searches. Inclusion and exclusion criteria were discussed among the research team of this tertiary study to obtain a common understanding of RE-related concepts in AI/ML-based systems. To identify relevant SLRs and mitigate search and selection biases, we searched papers through six sources and also did a recursive search on the references and citations (i.e., snowballing) of relevant papers. Besides, each paper was reviewed by three researchers (R1, R2, and R3). In case of disagreements, we resolved them through discussions and reconsiderations with the most experienced researchers (R4 and R5).

To ensure that the theoretical concepts were correctly represented during extraction and synthesis, we used the RE activities described in SWEBOK v.4.0. This decision aligns the observation of the extracted data with a reference literature. We also defined a data extraction form to ensure consistency in extracting relevant information, and we evaluated the data according to the research questions. In addition, we had at least three researchers who extracted the data independently. The more experienced researchers (R4 and R5) dealt with disagreements and divergences during the process.

Concerning the quality of the nine secondary studies found, we elaborated on quality criteria based on the CDR approach. To mitigate risks, two researchers carried out the quality evaluation process (R1 and R2), and two more experienced ones reviewed it. We also synthesized the results following the RE activities based on the SWEBOK v4.0 definitions.

Finally, despite all the effort spent to search for as many secondary studies on RE for AI/ML as possible, we are aware that some secondary research may not have been retrieved, which would restrict the generalization of our results.

5.4 Final Considerations

In this chapter, we present the tertiary study as part of the process based on the technology transfer model for identifying the gap identified in the industry (presented in the Chapter 2). We identified that many of the challenges reported in practice are confirmed from the perspective of the literature studies.

Our next step is to present the proposed candidate solution and how it was defined and refined to address one of the challenges reported in industry and academia.

Given the challenges identified in the study of interviews and the gaps identified in the tertiary study, we covered the initial stages of the MTT. Thus, we delimited the problem to be addressed in this research and previously presented, and thus delimited the candidate solution that will be presented in the following chapter.

The development of AI systems presents new challenges for SE and consequently for RE. It is a more complex activity when compared to the development of classical systems [40]. Several related issues still do not have an assertive answer, such as, for example, describing requirements in projects of an experimental nature, identifying the objective, aligning objective with data, managing stakeholder expectations, establishing communication and quality metrics.

There is a lack of proposals for RE processes for the development of AI systems [47]. There are some RE approaches for AI systems, however, these proposals deal with elicitation and specification. We propose a process based on ISO 12207 that deals with aspects from problem identification to system analysis. This DT-formatted process must still encompass all RE phases in an agile manner and appropriate to the experimental nature of RDI projects for AI systems. Thus, this proposal is a solution that covers all necessary ER activities for AI systems in innovation projects.

CHAPTER 6

Framework Conceptualization

This chapter presents the proposal developed throughout this research. We show how we used each of the concepts discussed in chapter 2 such as ISO 12207, Design Thinking, GUT matrix, 5W2H and ISO 31000.

The DIP-AI Canvas (Discovery and Innovation Process for AI) is the practical instantiation of the aforementioned mapping. It provides a visual, collaborative interface designed to mitigate the "experience gap"and communication silos identified in the empirical study (Chapter 4).

6.1 Standard-Based Process Mapping

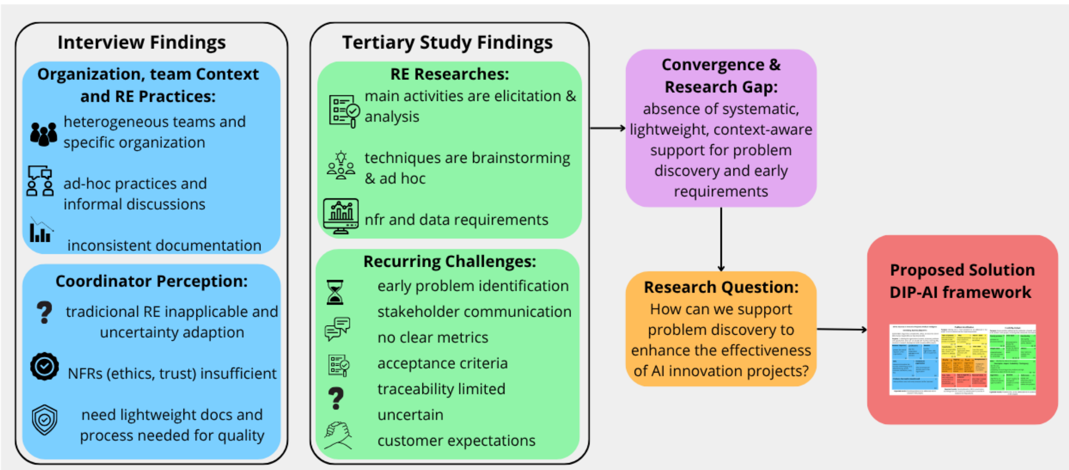

Figure 6.1: *Artifact Analysis Leading to the DIP-AI Framework*

The leftmost pane summarizes key Interview Findings, categorized into organizational contexts and coordinator perceptions. It highlights problematic current practices, such as "ad-hoc practices,informal discussions,"and "inconsistent documentation."It notes that coordinators perceive traditional Requirements Engineering (RE) as "inapplicable"and "uncertain,"emphasizing an acute need for "lightweight docs and process."

The central pane details Tertiary Study Findings, contrasting existing RE research with real-world recurring challenges. It identifies that current literature heavily focuses on elicitation and ad-hoc techniques like brainstorming. However, practical projects struggle with non-functional requirements (NFRs), stakeholder communication, lack of clear metrics, and uncertain customer expectations.

These findings are funneled into the Convergence and Research Gap pane, explicitly identifying the primary deficiency: the "absence of systematic, lightweight, context-aware support for problem discovery and early requirements."This gap is formalized into a fundamental Research Question, asking how to effectively support problem discovery in AI innovation projects. The logical end-point of this analytical process is the creation of the Proposed Solution: DIP-AI framework, represented by its multi-column canvas artifact, which seeks to directly address the synthesized industrial and theoretical requirements.

In this section we demonstrate the current status of the proposal. The description includes steps for building the catalogue and discussion of possible evaluation methods.

6.1.1 RE Process Development for AI

We studied the ISO 12207 technical processes related to ER. We listed the activities and tasks of each process and highlighted those that we considered essential for the execution of the ER phase so that they could not be omitted. The definition of the activities was carried out by myself and the supervisor of this research with experience in RE and with the support of a data scientist. The activities were discussed iteratively for four weeks.

In relation to the *business analysis* process, we include the following activities:

1. define the scope of the problem and analyze complaints,
2. characterize the solution space and identify alternative solutions,
3. evaluate alternative solutions and
4. maintain traceability.

For the process of *defining stakeholder needs*, we define the following activities:

1. identify stakeholders,
2. define needs, context of use and scenarios,
3. classify and prioritize needs,
4. identify constraints and relationships with non-functional requirements,
5. analyze the set of requirements and define performance measures,
6. discuss and give feedback,
7. obtain agreement and maintain traceability.

As for the *defining system requirements* process, we chose the following activities:

1. define the functional boundary,
2. identify modes of operation, restrictions and implementation risks and define requirements,
3. analyze requirements and
4. obtain agreement and maintain traceability.

Finally, we define the following activities for the *system analysis* process:

1. identify contexts and assumptions, analyze results, establish conclusions and recommendations, record results and
2. maintain traceability.

Once the essential activities have been defined, we seek to fit these activities into the double diamond approach of Design Thinking, as shown in Figure 6.2. Thus, in the *empathy* phase, we group the activities related to the *business analysis* process. In the *define* phase, we map activities related to the *stakeholder needs definition* process. In the *ideation* phase, we list the activities of the *system requirements definition* process. We model the *system analysis* process covering these three phases mentioned. In the *prototype* phase, we foresee the development of an ML-enabled system through prototyping, testing, evaluation, and refinement until it becomes a product that satisfies the stakeholders. Therefore, the requirements elicitation and specification activities are carried out iteratively; Requirements analysis is performed at all stages of the process, ensuring stakeholder involvement and understanding, followed by the requirements validation activity, where agreement and management are obtained to ensure traceability is maintained.

After designing the process, we identify the stakeholders involved in each stage. The process must be led by a requirements engineer, which is why this profile is not represented in the diagram. In the needs discovery phase, we list the stakeholders: business owner (BO), project manager (PM) and data scientist (DS). In the problem definition phase, we add the domain expert (DE). And in the ideation or requirements definition phase, we add stakeholders responsible for infrastructure (SE) and user experience (UX).

Since ISO 12207 provides the expected results for the application of each process, we also identify what is expected from each of the phases of the proposed framework. The empathy phase has as its expected result the discovery of the needs of customers and users. The expected result of the definition phase is the definition of challenges, problems, possible solutions and evaluation references. And finally, the ideation phase foresees as a result a greater knowledge of the proposed solution based on the multiple views of the stakeholders.

For each phase, we was defined artifacts and subprocesses suitable for the development of data-driven systems. Based on the process implemented in this work, we developed questions to identify a problem, analyze solutions, and raise other important

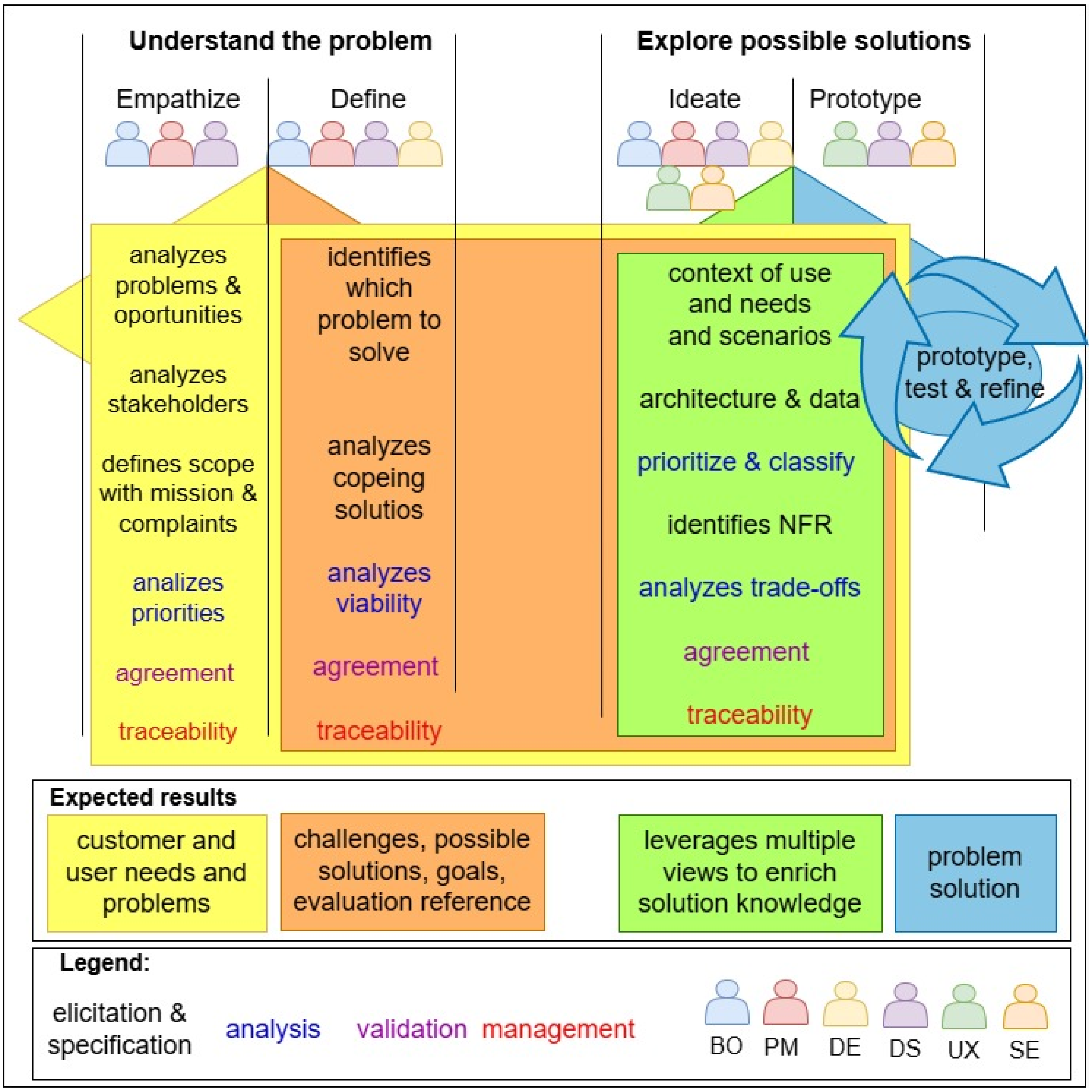

Figure 6.2: *RE Process to AI and Innovation Projects Design Thinking-BAsed*

questions based on the available data involving the different stakeholders important for the development of this type of system.

The **Empathy** phase presented in Figure 6.2 establishes that we must discover the needs and problems of customers and users. To do this, we seek an approach to explore and describe scenarios where a ML solution can be proposed. We adopted the 5W2H approach (*Who? What? Where? When? Why? How? How much?*, which is used to compose action plans. The requirements engineer initially consults the BO and DE to identify which scenarios exist in the organization and how these scenarios are operated. The idea is to identify workflows (WF) and decision-making (DT) that are done on an ad hoc basis and can be assisted or automated with an ML solution. The requirements engineer will then consult how these processes occur, which phases, who the actors involved are, why these processes occur in the organization, where these processes occur and when. This will identify the context of each scenario.

The requirements engineer will then consult the data scientist, to whom he will present all the identified scenarios and all the information previously established. The CD will then point out which of these scenarios have tasks that can be assisted by an AI solution. For example, optimization tasks, automation tasks or decision-making assistance. From the project's perspective, these tasks are problems to be solved with AI. Therefore, the requirements engineer will meet again with the BO and DE in order to determine the priority for the project. Then, the tasks that the CD identified as possible problems to be solved will be presented and the clients will evaluate the degree of criticality (GC) that will indicate the priority of each solution. The degree of criticality will be calculated based on the multiplication of the values of severity, urgency and tendency of that scenario to become worse. For each of these parameters, a value from 1 to 5 will be assigned, where 1 is not very critical and 5 is very critical. Thus, we will have a result that will indicate which problem has the highest score and therefore the highest priority to be solved in the project.

The **Define** phase presented in Figure 6.2 establishes that we must identify challenges, possible solutions, goals and obtain evaluation references. In this process we must evaluate the feasibility of the highest priority problem previously identified. Given that we have already identified the problems, priorities, and associated data tasks, we now seek to analyze the feasibility. To do this, we will analyze the expected input and output data, possible algorithms, the gold *dataset* for the solution, the objective to be achieved, possible evaluation metrics, data quality aspects, *baseline* established for the problem and references based on the state of the art. This information will be identified by the requirements engineer by consulting the CD, BO and DE. Finally, the data scientist will indicate whether the solution is viable based on the identified problem and the necessary and available data and the other attributes presented. If the highest priority problem is not viable, this step will be performed again considering the problem queue based on priority.

We defined the **Empathize** and **Define** steps by describing the processes, artifacts, stakeholders involved, and expected results. We was developed subprocesses for the **Ideate** and **Prototype** phases. We will subsequently evaluate the application of this process in an RDI unit involving AI projects.

6.1.2 The DIP-AI Canvas

We studied the RE-related ISOs technical processes. We listed the activities and tasks of each process and highlighted what we considered essential for executing the RE phase so that they would not be suppressed. We also defined activities with experienced requirements engineers and the support of a data scientist. The activities were discussed

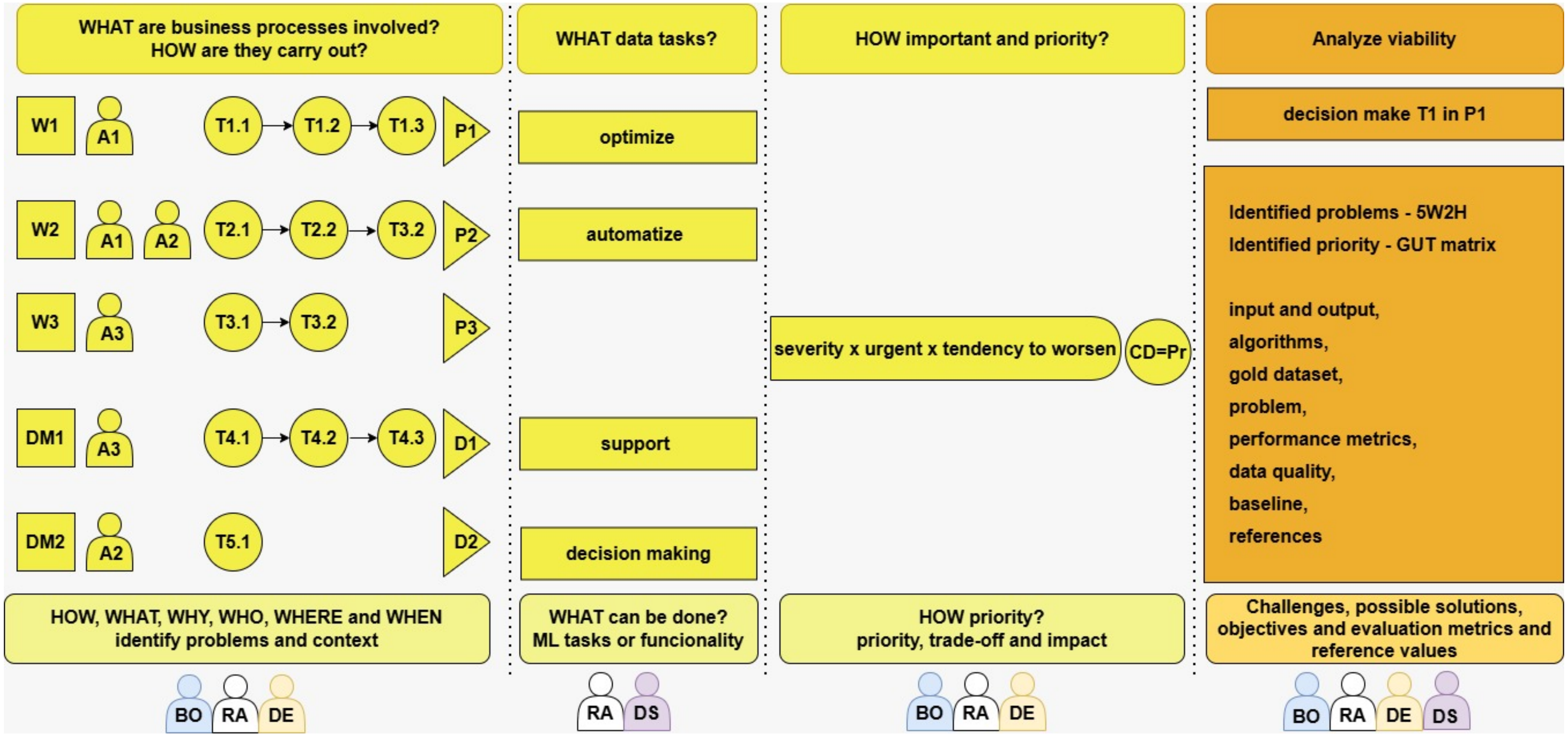

Figure 6.3: *A subprocess of RE for ML Systems based on DT - First Diamond phase.*

iteratively over four weeks.

Regarding the *business analysis* process, we included the following activities: (i) define the scope of the problem and analyze complaints, (ii) characterize the solution space and identify alternative solutions, (iii) evaluate alternative solutions, and (iv) maintain traceability. For the *stakeholders' needs definition* process, we defined the following activities: (i) identify stakeholders; (ii) define needs, context of use, and scenarios; (iii) classify and prioritize needs; (iv) identify restrictions and relationships with non-functional requirements; (v) analyze the set of requirements and define performance measures; (vi) discuss and give feedback; and (vii) obtain agreement and maintain traceability.

Concerning the *system requirements definition* process, we chose the activities as follows: (i) define the functional boundary; (ii) identify modes of operation, implementation restrictions, and risks, and define requirements; (iii) analyze requirements; and (iv) obtain agreement and maintain traceability. Finally, we defined the following activities for the *system analysis* process: (i) identify contexts and assumptions, analyze the results, establish conclusions and recommendations, record results, and (ii) maintain traceability. We believe that a simplified guide process can aid experimental and iterative AI and innovation development.

After defining the essential activities, we aligned them with the DT Double Diamond model. In the empathize phase, we positioned activities from the business analysis process. The define phase encompasses tasks related to the stakeholder needs definition process. The ideate phase integrates activities from the system requirements definition process. The system analysis process spans across all three phases, supporting iterative refinement and stakeholder alignment. Finally, the prototype phase involves the development, evaluation, and refinement of ML-enabled prototypes until a solution that meets stakeholder expectations is achieved [48].

[p]

Each DT phase also has expected outcomes aligned with ISO standards. The empathize phase aims to uncover user and business needs. The definition phase results in the articulation of challenges, problems, and evaluation criteria.

For each phase, we specify relevant information and subprocesses suitable for AI innovation projects. Based on this process, we formulated guiding questions to assist in problem identification, solution analysis, and data-driven stakeholder engagement critical to this development context. Table 6.1 presents the relationship between the main constructs of the canvas and ISO 12207 processes.

The 5W2H management tool has been used to make the planning and execution of projects, tasks, and actions clearer, ensuring that all relevant information is provided, assigned, and understood by all involved [28]. We identified that to understand the

Process/Fields Goups	**Question about context projects**	**5W2H**	**GUT Matrix**	**ISO 31000**	**Fesibility Analysis**
Business or mission analysis process	Define scope of the problem and characterize the solution space				Evaluate classes of alternative solutions
Stakeholder needs and requirements definition process		Define context of use	Prioritize and classify needs	Identify constraints	
System requirement definition process		Identify and define needs and justifications	Prioritize needs	Identify risk-related requirements	
System analysis process					Define strategy and perform analysis

Table 6.1: *Relation between ISO/IEC/IEEE 12207 modified by ISO/IEC/IEEE 5338 processes and DIP-AI Canvas*

problem we needed to answer some basic questions such as: **What** are the businesses involved? **How** are they carried out? **Who** are the stakeholders? **Where** do the processes take place in the organization? **How** do these processes happen and **why**? **What** are the costs and impacts of these activities? Aligned with the 5W2H questions. These questions help better characterize the organization's existing processes and problems. This allows for greater insight into the project management landscape and better alignment with stakeholders.

The GUT Matrix prioritizes problems based on Gravity, Urgency, and Trend, helping guide decision-making in dynamic project contexts [28]. This GUT matrix classifies problems to establish ordered priorities to deal with problems efficiently. Therefore, the processes, problems, and tasks are prioritized (using the GUT matrix). Then the feasibility analysis is carried out by collecting information such as input and output data, algorithms, metrics, references, etc.

We raise questions that support the development or definition of AI products. Furthermore, we define questions identified by [47] as challenges in the literature or that impact the development of AI innovation projects. These areas include: data quality, evaluation metrics, input and output data, algorithms, baselines, etc.

We also defined risks based on ISO 31000 [2]. The risk definition was made with description, impact, probability, and contingency plan. In order to identify risks, how they impact the project, the probability of occurrence defined as low, medium, and high, and what can be done to mitigate the risks.

We also identified the importance of understanding the task data associated with the problem as part of feasibility, as this determines whether an AI solution is applicable to the project. We refined DIP-AI by discussing it with project coordinators. The current version of the DIP-AI framework canvas is presented in Figure 6.5 and details about its structure follow.

The Figure 6.4 illustrates the systematic methodology employed to construct the DIP-AI framework. It details a rigorous engineering approach that integrates standardized processes from the software engineering domain with flexible, human-centered techniques from Design Thinking (DT). The resulting construction methodology is depicted as a multi-stage integration pipeline.

The process begins with the selection of standardized technical processes from ISO/IEC/IEEE 12207. Three foundational processes are selected: "Business Analysis, Stakeholder Needs and Requirements Definition,"and "System Requirements Definition."These are refined through iterative loops to incorporate context-aware adaptation and NFR priorities, ensuring relevance for the AI domain.

Simultaneously, the framework adopts the core philosophies of the Design Thinking Double Diamond model. This is represented by a visualization of the four primary phases: "Emphasize, Define, Ideate, and Prototype". These two paradigms are formally mapped during the "ISO-DT alignment"stage. For instance, the ISO Business Analysis process is aligned with the early DT phases to ensure deep problem understanding before definition.

In the next phase, "Integrate Toolset & AI Specific Extensions, the framework incorporates well-established industry tools to support specific activities identified in the combined processes. These tools are mapped directly to corresponding columns in the final Canvas Construction to fulfill specific objectives. The final output of this construction method is the standardized DIP-AI Framework Canvas, a unified visual workspace that enables the collaborative discovery of requirements for AI innovation projects, backed by the synthesized rigor of international standards and Design Thinking principles.

[p]

The proposed canvas is structured into three parts, following the first diamond of the Design Thinking model. The first part corresponds to the empathy and immersion phase, aimed at understanding the client's context and pain points. The second part focuses on defining and prioritizing problems and project objectives. The third part refers to ideation, where possible solutions for the selected priority problem are outlined, followed by a feasibility analysis.

Stakeholder involvement—an acknowledged challenge in AI projects [47]—is central to using the canvas properly. In the first phase, fields are completed primarily by the Business Owner (BO), who provides context, organizational background, existing challenges, and identifies other relevant stakeholders. In the second phase, the Domain Expert (DE) joins the BO to characterize and prioritize organizational processes, problems, gaps, and opportunities. At the end of this phase, the Data Scientist (DS) identifies which of the prioritized elements correspond to data-related tasks. In the third phase, the DS answers more advanced questions to evaluate the feasibility of developing AI-based solutions for the selected tasks.

By guiding this multi-stakeholder collaboration and addressing specific AI-related challenges, the canvas offers a practical approach tailored for AI innovation projects [47, 9]. It was designed to promote an iterative and adaptable approach suited for AI innovation projects, aimed at facilitating communication through its visual structure, enabling the alignment of business objectives and technical feasibility with data, and supporting the understanding and prioritization of problems and processes.

6.2 Final Considerations

In view of the above, our next step is to finalize the definition of the ER process for the development of ML innovation projects. Next, we seek to evaluate the process

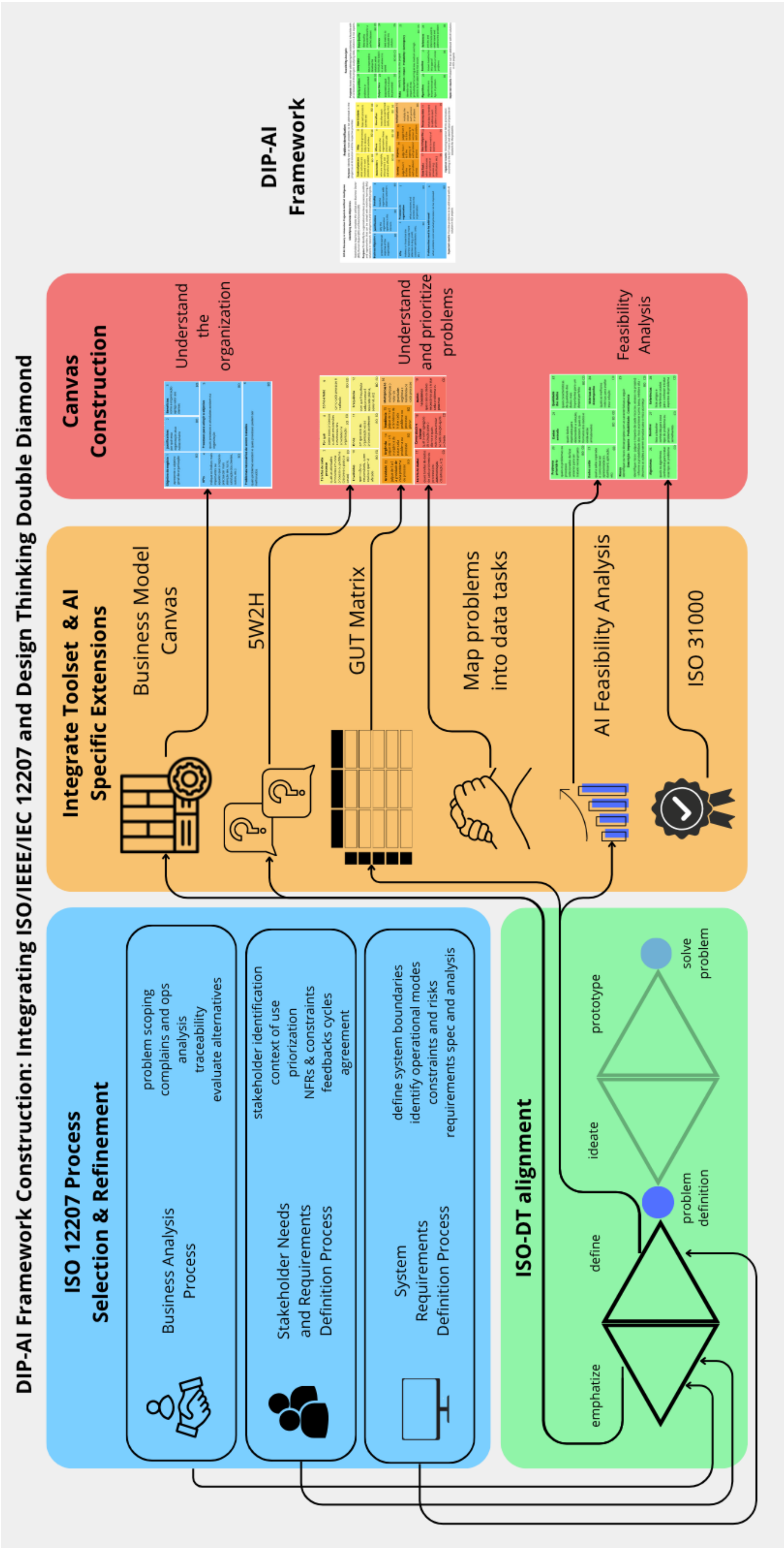

Figure 6.4: *DIP-AI Framework Construction Method.*

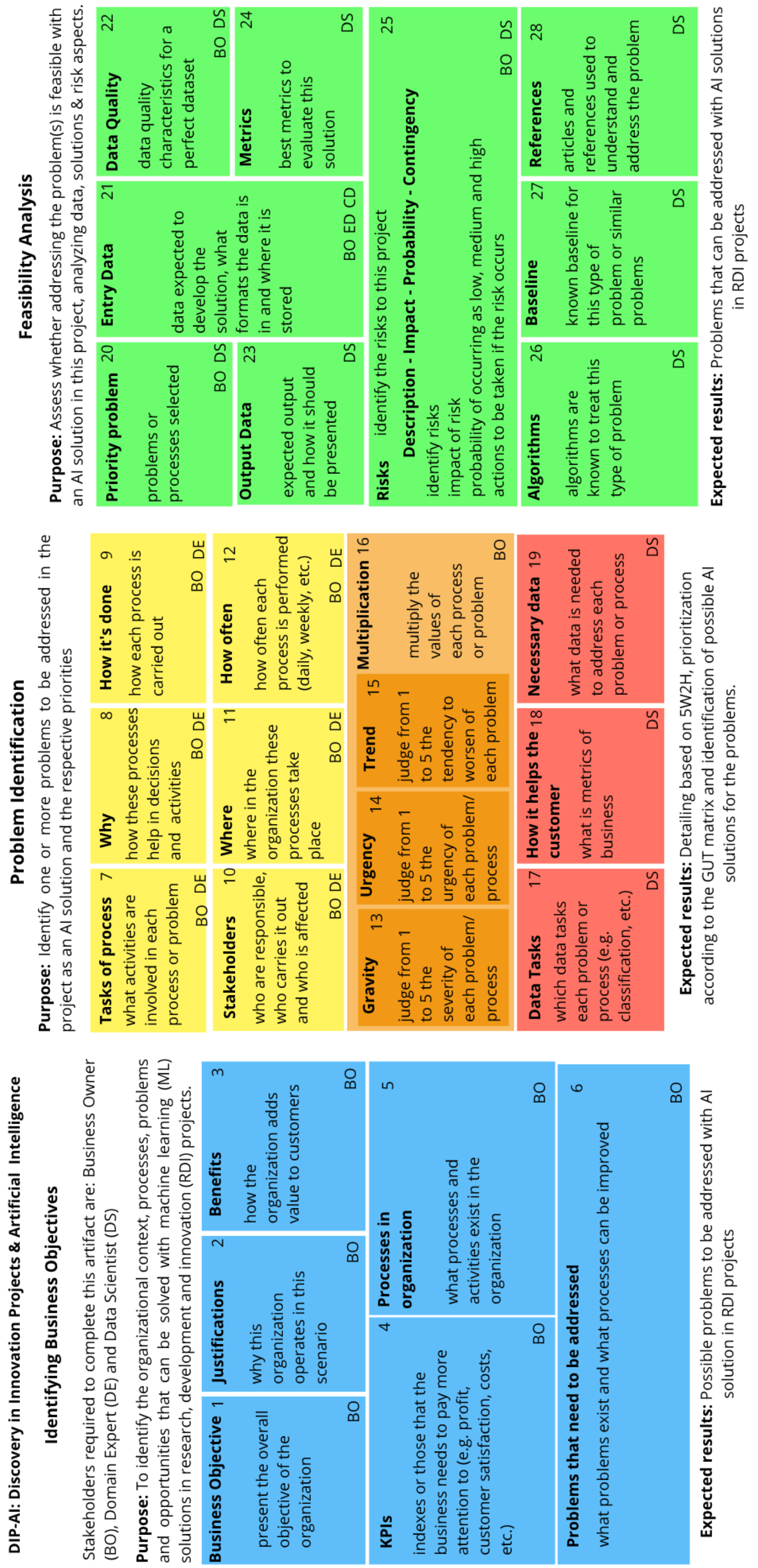

Figure 6.5: *DIP-AI Canvas.*

using the forms provided for in the MTT, analyzing aspects of efficiency, effectiveness, usability and ease of use.

The DIP-AI framework bridges the gap between formal ISO standards and the experimental nature of AI innovation. By providing a structured yet visual approach, it facilitates the elicitation of both functional and data-specific requirements. The next stage of this research involves a formal evaluation of the framework in a real-world RDI environment, assessing its effectiveness, usability, and impact on delivery quality.

CHAPTER 7

Validation

In this chapter we presented the validation sessions. We developed pilot test, validation with students, validation with coordinators in static scenario and validation in dynamic scenario.

7.1 Validation with Students

To evaluate the preliminary feasibility, learning curve, and perceived usefulness of the proposed framework, we conducted an empirical study with postgraduate students. This evaluation was integrated into the practical activities of an advanced graduate course focusing on the intersection of Software Engineering and Artificial Intelligence. Utilizing a postgraduate cohort serves a dual methodological purpose: it leverages participants who already possess foundational proficiency in both domains, and it provides a controlled environment to observe how the framework structures the discovery phase of complex, innovative projects. During the evaluation, students were exposed to the challenges inherent in discovery for non-deterministic systems, applying our approach to characterize real-world problem spaces. The following subsections detail the experimental protocol, including participant characteristics, task design, qualitative evaluation criteria, and the systematic analysis of the collected feedback.

7.1.1 Evaluation Design

The objective of these stages is to progressively refine the candidate solution to ensure its suitability for industry adoption. Following the Technology Transfer Model (TTM), **we carried out three evaluations in distinct contexts:** (i) an academic validation with graduate students in Software Engineering for AI, who used the approach to identify problems in their final course projects; (ii) a static validation in an industrial setting, where RDI professionals retrospectively applied the approach to real project artifacts—this type of validation is conducted offline, with real practitioners and data, but outside the context of active development; and (iii) a dynamic validation, embedded in a real-time industrial case study within an ongoing project. These iterative assessments were designed to promote early issue identification, enhance user satisfaction, foster continuous improvement, and build trust in the solution's applicability. An approved ethics review was conducted prior to performing validations involving human participants. Subsequently, a pilot study was carried out with five graduate students to assess the validation procedure and data analysis methods, as well as to identify potential issues [1].

We conducted a pilot test with 5 practitioners of R&D projects from different roles. The idea was to understand if the artifacts, the dynamics of data collection, analysis, and description were adequate, or if adjustments were needed before the start of the validations.

We followed the Goal-Question-Metric (GQM) goal definition template [66] to define our research goal as follows: ***Analyze*** **DIP-AI with the** ***purpose*** **of characterization** ***with respect to*** **the problem discovery capability, overall acceptance, and criticisms and suggestions for improvement, from the point of view of participants in the context of AI innovation projects.**

Based on this objective, we formulated the following Research Questions (RQs):

- **RQ1: How well does DIP-AI contribute to problem discovery in AI innovation projects?** To answer this question, we closely followed the experience of applying DIP-AI, analyzed the completed DIP-AI canvas artifact, specific follow-up questions, and considered the concrete result of the project (which, at the time of writing, has been completed and deployed at the industrial partner).
- **RQ2: How well is DIP-AI accepted in practice?** To answer this question, the follow-up questionnaire was designed based on the Technology Acceptance Model (TAM 3) [67].
- **RQ3: How can DIP-AI be further improved?** To answer this question, we qualitatively analyzed open-text responses to the follow-up questionnaire regarding criticisms and improvement suggestions.

In all the validation scenarios the data collected throughout the experience of the participants applying DIP-AI involved the completed DIP-AI canvas artifact, the follow-up questionnaire, and information on the concrete result of the project.

For RQ1 (contributions to discovery) we considered the overall problem discovery experience, questions about perceptions of contributions of the artifact (questions that we created and that are shown in group 1 in Heat Maps Figures), and the open question "In the projects in which you participated, what were the difficulties in identifying the problem or machine learning task to be addressed in that project? Could they be mitigated by using this artifact?".

[1]CAAE: 81475724.4.0000.5083

For RQ2 (acceptance), we considered the answers to the TAM-based [67] questions. Considering the research objectives and context, the study focused on Technology Acceptance Model (TAM) constructs related to perceived usefulness—specifically job relevance, output quality, and result demonstrability—and perceived ease of use, including computer self-efficacy, perception of external control, perceived enjoyment, and objective usability.

Finally, for RQ3, we considered the answers to the following open questions:

- What limitations, recommendations, or opportunities for improvement do you identify in this artifact?
- Do you consider that any questions or fields are missing from the artifact?
- Do you find the current structure and organization of the artifact to be clear and appropriate?

7.1.2 Data Collection

For academic validation, we invited graduate students enrolled in a Software Engineering for Artificial Intelligence course. Since the course required the development of an innovative AI product, we applied the framework to assist them in the process of discovering their solutions. We obtained initial feedback on the use of the tool.

The course had 36 participants, all of whom were invited to participate. Participants were able to use the framework after receiving a 1-hour lesson on the constructs and how to complete the framework. During this time, participants also completed the profile questionnaires and the Informed Consent Form.

Students had 4 hours to ask questions online about filling the framework and one week to submit the completed framework and answer the perception questionnaires. Of the participants, only 14 accepted and completed all questionnaires correctly. Therefore, we will report here the results regarding the perception of these participants.

7.1.3 Analysis Procedures

For the quantitative analysis of the TAM questions, we performed a reliability analysis using Cronbach's Alpha. Reliability indicates how reliably or accurately a questionnaire or test measures a true value. Cronbach's alpha is given by $\alpha = \frac{N\bar{c}}{\bar{v}+(N-1)\bar{c}}$, where N represents the number of items, $\bar{c}$ is the average inter-item covariance, and $\bar{v}$ is the average variance. The quantitative analysis was conducted by aggregating the number of responses for each questionnaire item, based on a seven-point Likert scale to assess the perceived contribution of the framework.

The qualitative analysis of open questions was conducted using the Atlas.TI tool [13], applying open and axial coding procedures from Grounded Theory [64].The method we used was to gather participants' responses to each open-ended question in the questionnaire. We then analyzed the contributions, coded each response, grouped the responses, and established possible relationships. Since we were unable to reach data saturation due to the small sample of participants, we did not perform selective coding.

In Figure 7.1, we present three heat maps that represents the responses of the participants to the 36 Likert scale items (the five created ones for artifact contribution, and the 31 from the 8 selected TAM 3 constructs) that form the perception questionnaire of this research. The questions are numbered in the form X.Y where X represents the group of questions and Y enumerates the questions in a group (questions of groups 2 to 9 in the heat maps, which are: usefulness, ease of use, ability to perform the task using the artifact, control of the artifact to perform the task, pleasure in using the artifact, perceived relevance of the artifact, output quality and demonstrability, and intention to use the artifact. The values in each field of the heat map represent the number of responses to a given question.

7.1.4 Profile of Participants

All participants are graduates, with 1 currently pursuing a master's degree and 5 holding a master's degree. The educational backgrounds included: 3 electrical engineers, 2 information systems engineers, 1 dentist, and 2 computer network engineers; the remaining backgrounds represented only one participant each, including computer scientists, designers, mathematicians, software engineers, etc.

Eight participants stated that they work with AI under supervision, 1 without supervision, and 5 supervise others. Participants who have worked on innovation projects detailed their roles as: AI Engineer, Project Manager, Scrum Master, Data Scientist, Developer, Product Owner, Requirements Analyst, Consultant, and Tech Lead. Only six participants expressed a positive perception regarding: being experienced in identifying AI problems in projects and being skilled in identifying AI problems in projects.

We performed Cronbach's alpha calculation and obtained the result 0.9610669991469256. More details about the questionnaire responses can be seen in Figure 7.1.

7.1.5 Results

RQ1: DIP-AI's contribution to problem discovery

In this scenario is important see that students have experience with software development and the functions like product owner, and management, developer, etc. And some have experience in start the enterprise of technology.

The students evaluated it is possible to observe that participants reported that use of the DIP-AI canvas positively influenced project execution by optimizing resources and supporting the achievement of objectives, particularly because problem identification was conducted accurately, efficiently, and effectively.

Additionally, participants acknowledged that the use of the DIP-AI artifact helped address common challenges encountered during the problem discovery phase in innovation and AI-related projects. In this group we have participants that never developed software, or AI solutions, or RDI projects then the results positive about the help framework in identifying problem activities in first time is very significative.

RQ2: DIP-AI overall acceptance

Participants expressed positive perceptions regarding the **usefulness** of the artifact (as reflected in the responses from Group 2). They considered the artifact to be beneficial in several ways: it enhances productivity, increases assertiveness in decision-making, and supports effective problem identification. And one participant identifying the collaboration of framework is neutral in your perception.

Regarding **ease of use** (Group 3), three of the participants reported that the artifact requires a lot of mental effort. And only one student reported that he did not find it easy to use the artifact for what he wanted, and the framework is not easy of use.

Regarding the group of questions (Group 4) related to the **ability to identify problems** using the artifact, two participants reported that they would have difficulty responding the activity without guidance. Additionally, we had three neutral perceptions regarding the ability to complete the work if someone showed me how to do it first, and we had positive perceptions regarding the ease of using artifacts already used by others similar to them.

Regarding the perception of **controlling the artifact** (Group 5), one person indicate negative response about the framework presents the necessary resources to identify the problem and two persons indicate the framework is not compatible with other systems and artifacts used in the organization of them.

Regarding the dimension of **pleasure** (Group 6) When performing the activity with the artifact, one participant reported "I find the artifact pleasant to use,"three indicated "The process of using the artifact is pleasant,"and four indicated "No, I have fun using the artifact."We consider that, since the artifact is related to a laborious and challenging activity, the participants will not really perceive pleasure; however, we can consider that it may be more pleasurable with this artifact than with others, but this was not considered in this questionnaire.

No negative perceptions were reported regarding the **relevance** of the artifact (Group 7) or the importance of its use in supporting tasks related to AI project development. We have negative comments concerning the **demonstrability and quality of the artifact's results** (Group 8), about the apparent results of use the framework, and the difficult of explain why use of artifact can be beneficial. Like some participants do not have experience in discovery activities we do not completely understand the necessary and the importance of the activity and the framework contribution. About the intend of use in the future (Group 9), two participants declared neutral, and others declared intend of use the framework.

In general analysis we can observed that usefulness, and ability to identify problems have the positive perceptions by the students. And the ease of use, pleasure do not have positive perceptions in general. Some the questions about demonstrability of results have the contradictory perceptions, sometimes positive and sometimes negative.

RQ3: DIP-AI criticisms and suggestions

Most participants reported the need for a shared repository with completed canvases to get an idea of what is expected in each field and possible adaptations in different scenarios. This was the contribution of P5, P13, and three other participants. As they are students, it is natural that such a contribution is relevant.

P7 and P13 indicated fields that are missing in their opinion. Meanwhile, P8 suggested that the canvas could be reduced and have fewer fields. P13 also indicated other suggestions, such as: completion with generative AI, a validation checklist, and integration with other tools commonly used in AI and innovation projects.

- P5: 'A repository with case studies or complete application examples.'
- P7: 'Missing fields such as "How long will the project take to develop?", "How many people will be impacted?", "What is the size of the development team?", "Project timeline"or "Delivery forecast".'
- P8: 'Perhaps reduce steps or stages'
- P13: 'Add fields such as "Data Quality", "Infrastructure Availability"and "Expected Technical Challenges"to validate if the problem can be solved with AI. Rephrase some questions in a more direct and structured way to facilitate completion by less experienced teams. Create a final validation checklist to ensure that all parts of the problem have been well defined before the model development. Automate the completion with generative AI. Integration with project management tools (Jira, Trello, Notion) to facilitate communication between teams.' Creating an interactive or automated decision flow could help teams select the best machine learning approach for their problem.

We realized that the canvas can be a bit lengthy and difficult to understand. We also identified that some fields are unclear or could be added. We also noted that some

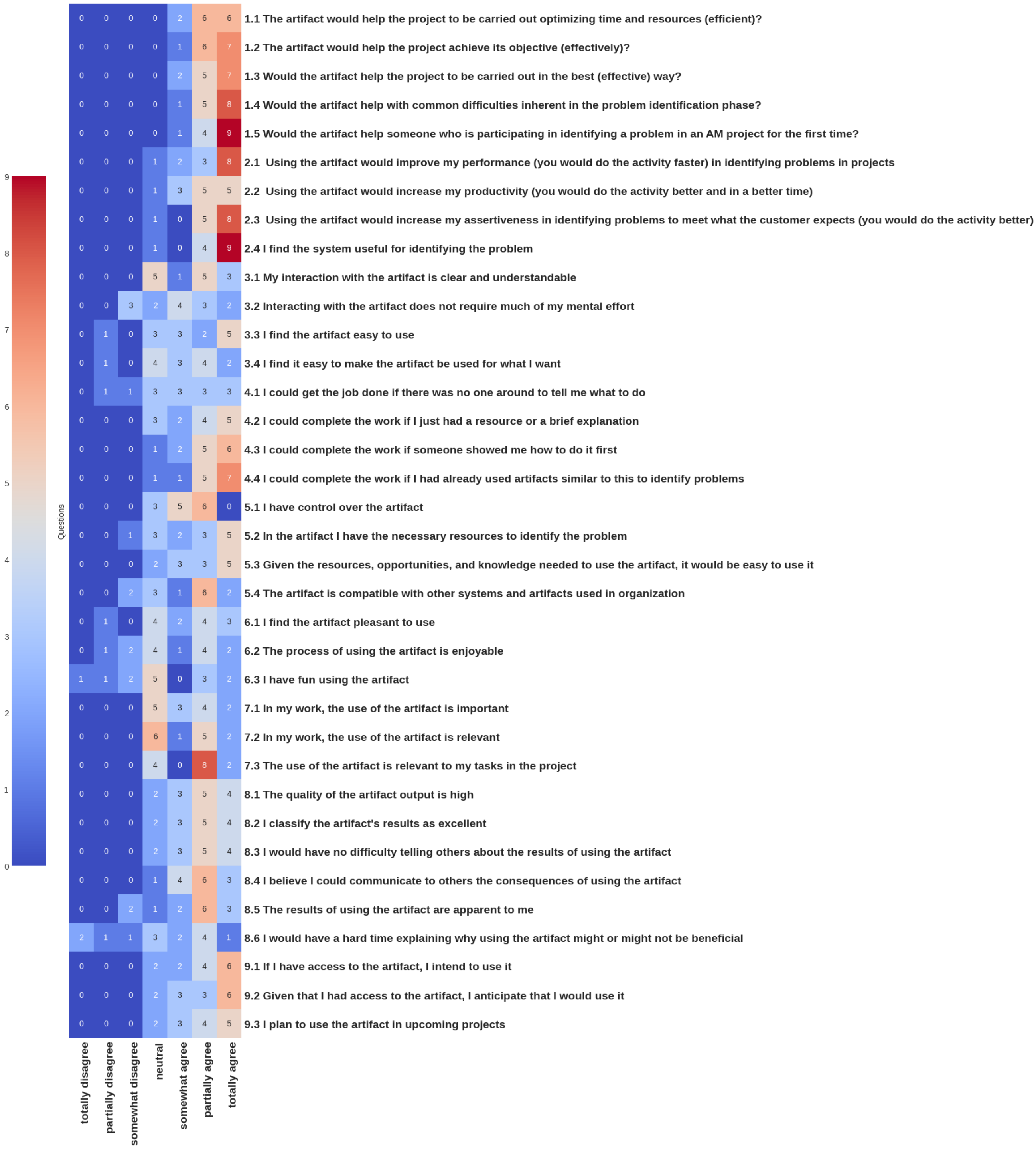

Figure 7.1: *Heat Map of Participant's Perception (TAM 3-based) in Academic Validation*

information that is normally elicited in other phases of the project was missing from the canvas, such as: timeline, number of people impacted, necessary infrastructure, etc. We emphasize that some of this information can be filled in the risks section, however, not all of it, as some is completed in phases subsequent to the discovery phase. We provided an interface in Miro and Canva for participants to integrate with their tools, and we did not receive feedback on whether this is sufficient or not. We also studied the possibility of integrating the canvas with generative AI to facilitate completion and provide suggestions for user responses.

7.1.6 Discussion

Although the evaluation conducted with students provides a more limited contribution to software engineering practice compared to industrial evaluations, the participant group consisted of postgraduate students, some of whom have relevant professional backgrounds. Several participants reported prior industrial experience, including cases with more than ten years of experience in software engineering. Therefore, their feedback still offers valuable insights regarding the artifact.

Overall, participants reported positive perceptions concerning the usefulness of the artifact and the sense of control when interacting with it. These results suggest that the artifact can support users in structuring and guiding discovery-related activities. However, the ease of use dimension received less favorable evaluations, indicating that participants experienced some difficulty when interacting with the artifact.

Additionally, negative perceptions were observed regarding enjoyment and mental effort. It is important to note that discovery activities, particularly those involving the identification of project needs and problem exploration, are inherently cognitively demanding.

Nevertheless, participants reported that the task became particularly tiring due to the large number of fields to be completed and the variety of aspects that must be considered during the process.

The open-ended responses provided further insights into potential improvements. Several participants suggested the inclusion of completed examples to help users better understand how to fill in the fields and interpret the required information. Other responses indicated the need for additional fields to capture relevant aspects of projects, such as impacted stakeholders, project duration, team size, data quality, infrastructure availability, and technical challenges or project constraints.

These observations highlight opportunities for refining the artifact, particularly by improving usability, providing guidance for completion, and expanding the representation of contextual project factors.

7.1.7 Threats to validity

Some threats to validity should be considered when interpreting the results of this evaluation. First, the evaluation was conducted with postgraduate students rather than industry practitioners. Although several participants reported prior experience in software engineering, many had limited or no experience specifically with AI-based projects. This may have influenced their perception of the artifact and their ability to fully assess its applicability in real industrial contexts.

Another potential threat relates to the academic setting in which the evaluation was conducted. Since the activity took place within a postgraduate course, factors inherent to educational environments—such as limited time, varying levels of engagement, and differences in students' backgrounds—may affect the quality and consistency of the results obtained.

Finally, the projects used as the basis for the evaluation were still in their early stages. These projects were intended to be further developed and investigated throughout the course and eventually used as the basis for the students' final projects. As a result, participants may not yet have had a complete understanding of the project context, which could have influenced their ability to provide more detailed or mature evaluations of the artifact.

7.1.8 Final Remarks

The qualitative feedback obtained through the open-ended responses proved particularly valuable, as it generated several insights for improving the proposed approach. These contributions highlighted opportunities for refinement in multiple aspects of the artifact. Suggested improvements include the addition of new fields to better capture relevant project characteristics, enhancements in usability to reduce user effort, and the potential development of a supporting tool to facilitate interaction with the canvas.

Participants also suggested the possibility of partially automating the completion process and creating a repository of completed canvases. Such a repository could serve both as a reference and as a source of inspiration for users, helping them better understand how to interpret and fill in the canvas fields. Collectively, these suggestions provide useful directions for future iterations of the approach, aiming to improve both its usability and practical applicability.

7.2 Static Validation

7.2.1 Evaluation Design

For static validation, we invited coordinators enrolled in AI and RDI projects. We invited coordinators who had ongoing projects and completed projects that had already gone through the discovery phase. The idea of static validation is to apply the framework to projects that have already completed the discovery phase without the framework, and we want to obtain the coordinators' perception of how it would be to use it in the context of their projects.

We invited ten coordinators but were only able to find an available time slot with eight. We recruited participants based on a list of projects completed within a one-year period at our university. We scheduled a 1.5-hour meeting with the participants and sent profile questionnaires and the Informed Consent Form. However, only 7 participants completed and correctly filled out all the forms and questionnaires to participate in the evaluation.

During the meeting, we attempted to complete the framework questionnaire fully regarding a completed project. After filling it, the coordinators filled out the perception questionnaire.

Profile of Participants

All participants hold doctoral degrees and are all professors at the Institute of Informatics at our university. They stated that their AI knowledge is as follows: 2 "know a little,"2 "know AI,"and 4 "know a lot."All stated that they work without supervision and collaborate on projects. All stated they have participated in more than 3 AI projects; three of the coordinators stated they have participated in more than 10 projects.

In addition to their collaborative roles, participants stated they have experience as software architects, requirements analysts, product managers, and developers. Five participants stated they are experts in identifying problems in AI projects.

We performed Cronbach's alpha calculation and obtained the result 0.897220. More details about the questionnaire responses can be seen in Figure 7.2.

7.2.2 Results

RQ1: DIP-AI's contribution to problem discovery

Although this evaluation considers projects that have already been completed, we want to obtain the perception of experienced professionals regarding how the framework

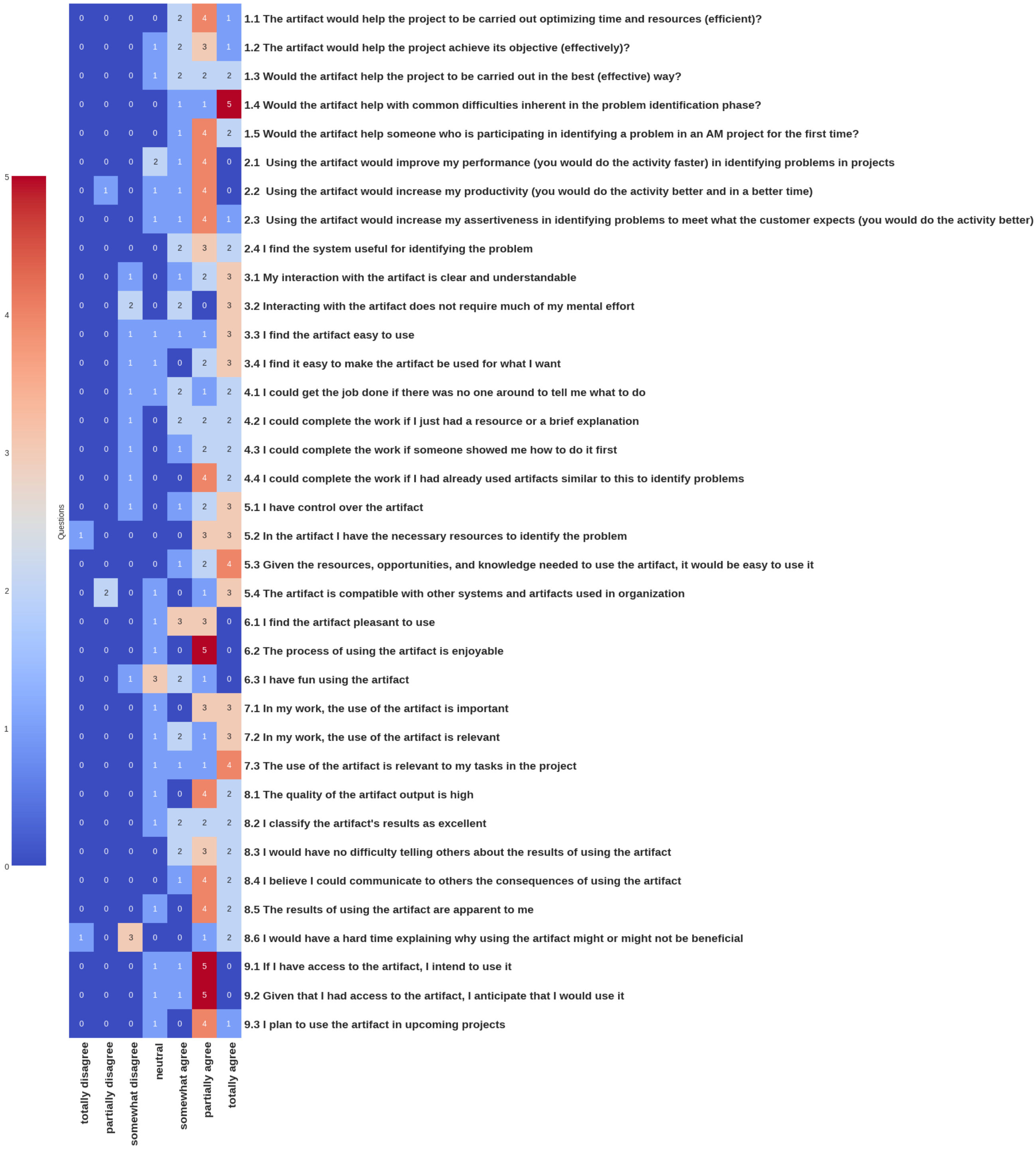

Figure 7.2: *Heat Map of Participant's Perception (TAM 3-based) in Static Validation*

would be used in a completed project, gain insights for improvement, and prepare ourselves for dynamic evaluation in real, ongoing projects.

The coordinators realized that the framework would optimize the project execution by working with time and resources and would help to achieve the project objective effectively. Most strongly agreed that the artifact would help overcome common project difficulties.

RQ2: DIP-AI overall acceptance

Participants expressed positive perceptions regarding the **usefulness** of the artifact (as reflected in the responses from Group 2). Most responses indicated partially positive perceptions regarding increased performance, improved productivity, and helpfulness when using the framework. A neutral perception was also noted for each of the previously mentioned topics. The coordinators also considered the framework useful for the task it is designed to support.

Regarding **ease of use** (Group 3), three participants reported that understanding the artifact is clear, requires no mental effort, and is easy to use and adapt. However, two participants gave negative responses regarding the absence of mental effort, and one participant gave a negative response to the other statements, slightly corroborating what was indicated in the evaluation with students.

Regarding the group of questions (Group 4) related to the **ability to identify problems using the artifact**, one participant reported difficulty completing the activity without guidance, and another was neutral in their response. Additionally, we had a slightly negative perception regarding the items: not being able to complete the work without a brief explanation, or completing it if someone showed me how to do it first, and being able to complete the work if I had already used similar artifacts.

Regarding the perception of **controlling the artifact** (Group 5), a slightly negative perception about controlling the artifact. A strongly negative perception regarding the artifact having the necessary resources to identify the problem and two slightly negative perceptions regarding the framework being compatible with other systems and artifacts used in the organization.

Regarding the dimension of **pleasure** (Group 6), we had five positive perceptions regarding the artifact being pleasant to use and three neutral perceptions about having fun using the artifact. Again, as the artifact is related to a challenging stage of the work developed in the project, it is natural that the responses do not include many positive perceptions regarding the pleasure of using the artifact. However, we noted an improvement in the participants' perception compared to the previous evaluation.

No negative perceptions were reported regarding the **relevance of the artifact** (Group 7) or the importance of its use in supporting tasks related to AI project development. We did receive negative comments concerning the **demonstrability and quality of the artifact's results** (Group 8), regarding the difficulty of explaining why the artifact's use can be beneficial. However, regarding other perceptions about the demonstrability of results using the artifact, we only had positive perceptions. Regarding the **intention to use it in the future** (Group 9), most responses showed interest in continuing to use the framework, and only one participant indicated a neutral position on its future use.

In general analysis, we can observe that the coordinators perceive that the framework can help with difficulties inherent in AI and innovation projects and indicated that the framework is useful and produces demonstrable results. Furthermore, they demonstrated an intention to use it in the future. However, we had negative perceptions regarding increasing productivity using the artifact, ease of use, and ability to perform the task with the artifact, therefore identifying room for improvement, as was identified in the previous evaluation.

RQ3: DIP-AI criticisms and suggestions

Two of the participants did not contribute to the open-ended questions. Only one participant indicated that they felt the **lack of a field** related to the technical feasibility of obtaining information, such as: need for personnel, algorithms, infrastructure, etc. Regarding the **organization of the artifact**, most responses indicated that it is adequate; however, we received **suggestions** for a tool to support data entry and examples to facilitate understanding of the fields.

- P3: 'Support for a web tool that implements canvases with user login support, basic project management and collaborative editing by multiple users, shared visualization of each collaborator's edit history, mechanism for recording each collaborator's contribution, etc.'
- P3: 'It would be important to visualize the relationships between the fields on the canvas, and how much one field could help fill in another.'
- P6: 'It could be a software tool to keep records and serve as a database for our RDI unit.'
- P7: 'An easier interaction, like a conversational agent, to ask me questions and explain them.'
- P7: 'The organization is clear. But it can always be improved by providing examples for those filling out the forms.'
- P7: 'I believe that perhaps what's lacking is the field of technical feasibility (e.g., people) and technological feasibility (e.g., algorithms).'

7.2.3 Discussion

For the evaluation conducted with coordinators, which focused on analyzing previously completed projects, several relevant contributions were identified. Similar to the feedback obtained in other evaluations, participants emphasized the importance of improving usability through the development of a dedicated web-based tool. Such a tool would facilitate access to the canvas, as well as support user and project management, enabling more structured interaction with the artifact.

Another important suggestion concerned the creation of a centralized repository or database containing the completed canvases of projects conducted within the organizational unit. This repository would allow coordinators and project members to access information from past projects, promoting knowledge sharing, improving traceability of decisions, and enabling comparisons across different initiatives.

A particularly relevant contribution was the suggestion of incorporating a conversational agent to assist users while completing the canvas. Participants indicated that such an agent could guide users through the process, clarify the meaning of specific fields, and help formulate appropriate responses. This type of assistance could reduce cognitive effort and improve the overall usability of the artifact, especially for users who are less familiar with discovery activities or AI-related projects.

Overall, the feedback from coordinators highlights opportunities to evolve the artifact from a static documentation instrument toward a more interactive and knowledge-supported platform, capable of facilitating project analysis, supporting decision-making, and preserving organizational knowledge generated across projects.

7.2.4 Threats to validity

A relevant threat to validity in this evaluation stems from the retrospective nature of the analyzed projects. Since the canvas was completed after the projects had already been concluded, the information provided relied largely on the coordinators' recollection and available project records. As a result, the accuracy and completeness of the data may have been influenced by memory limitations or by the availability and quality of documented information.

Consequently, some project details may have been partially reconstructed or interpreted based on the coordinators' current understanding of the projects rather than on contemporaneous records. This factor may have affected how certain fields of the canvas were completed and, therefore, should be considered when interpreting the results of the evaluation.

7.2.5 Final Remarks

Overall, the evaluation conducted with coordinators resulted in slightly more positive perceptions of the framework when compared to the evaluation carried out with students. This difference may be explained by the coordinators' greater familiarity with project management activities and their broader perspective on the needs of organizational projects.

Despite the generally positive feedback, some aspects—particularly ease of use and enjoyment—were identified as areas that require further improvement. These results suggest that the framework, while conceptually useful, would benefit from refinements aimed at reducing user effort and improving the overall interaction experience.

On the other hand, key dimensions such as usefulness, relevance, and intention to use received positive evaluations. These findings reinforce the practical value of the proposed framework and support its potential applicability in real project contexts. Consequently, the results provide encouraging evidence regarding the contribution of this work while also highlighting directions for future improvements.

7.3 Dynamic Validation

The objective of these stages is to progressively refine the candidate solution to ensure its suitability for industry adoption. Following the Technology Transfer Model (TTM), we carried out three evaluations in distinct contexts, in a dynamic validation, embedded in a real-time industrial case study within an ongoing project. These iterative assessments were designed to promote early issue identification, enhance user satisfaction, foster continuous improvement, and build trust in the solution's applicability. This section concerns detailing the DIP-AI framework and describing its application in the dynamic validation phase, conducted within an IAC project with a security company, where the framework was employed during the discovery to identify issues from the ground up in a real-world AI innovation project.

7.3.1 Case Study Design

This investigation is structured as an evaluative case study involving participants actively engaged in a discovery phase to identify and characterize a viable problem for the project. The methodological approach was guided by the case study research principles outlined by Runeson *et al.* [60].

7.4 Goal and Research Questions

We followed the Goal-Question-Metric (GQM) goal definition template [66] to define our research goal as follows: ***Analyze*** **DIP-AI with the** ***purpose*** **of characterization** ***with respect to*** **the problem discovery capability, overall acceptance, and criticisms and suggestions for improvement, from the point of view of participants in the context of IAC AI innovation projects.**

Based on this objective, we formulated the following Research Questions (RQs):

- **RQ1: How well does DIP-AI contribute to problem discovery in AI innovation projects?** To answer this question, we closely followed the experience of applying DIP-AI, analyzed the completed DIP-AI canvas artifact, specific follow-up questions, and considered the concrete result of the project (which, at the time of writing, has been completed and deployed at the industrial partner).
- **RQ2: How well is DIP-AI accepted in practice?** To answer this question, the follow-up questionnaire was designed based on the Technology Acceptance Model (TAM 3) [67].
- **RQ3: How can DIP-AI be further improved?** To answer this question, we qualitatively analyzed open-text responses to the follow-up questionnaire regarding criticisms and improvement suggestions.

7.4.1 Validation Context

The context of this study is an industry-academia collaboration innovation project initiative designed to enhance the education of undergraduate and postgraduate students. In this initiative, participants play a central role in the design and development of innovative AI-based technology solutions aimed at addressing concrete business challenges of real industry partners. The industrial partner involved in the particular case study is a technology company that develops digital solutions for multiple clients, with a product portfolio focused on technological security across various industry sectors.

Problem discovery is an important part of an AI innovation project, enabling the identification of problems rooted in the genuine needs of the industrial partner. The project in question was suitable for our case study as it did not yet have a clearly defined problem. We spent a month on-site accompanying the participants while carrying out the Discovery phase of the project with the support of DIP-AI. We participated in the onboarding meeting with the customer, daily meetings, and applied the questionnaire.

Regarding the profile of the participants, the project team had eight undergraduate students and two doctoral students. All participants had at least one year of prior experience in AI innovation projects, having been involved in between two and six such projects. Six of the participants declared themselves experienced and skilled in identifying problems in AI projects. More details about the participants' profiles are shared with other materials from this evaluation [2]. Figure 7.3 illustrates the participants in the context of the case study and the DIP-AI framework used in the background.

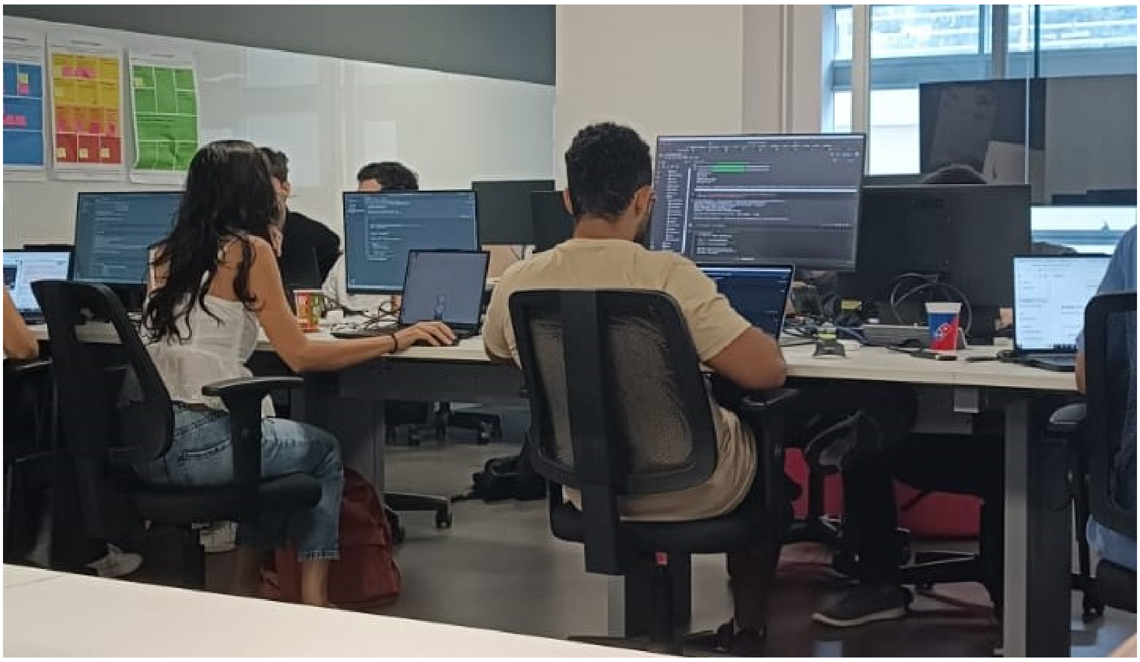

Figure 7.3: *Participants in the Case Study Context and the Use of DIP-AI in the background.*

Figure 7.4 illustrates the DIP-AI framework canvas completed during the case study. The canvas was first introduced to the project coordinators on January 15, 2025, when the constructs employed in the canvas, the potential benefits of its evaluation, and the components of the artifact were presented. On January 22, 2025, the first author conducted a presentation to the participants, providing an overview of the canvas and explaining each field in detail. During this session, participants posed questions about its use and engaged in discussions about the relevance of the canvas and how it might help to mitigate challenges in their project. Each presentation lasted approximately one hour.

In the subsequent days, the first author documented difficulties reported by participants in their respective project that could potentially be addressed through the proposed approach. Additional contextual information about the industrial partner and the project challenge was also gathered. On January 29, 2025, the project objective was revised in collaboration with the partner company and the innovation unit due to feasibility concerns. As a result, the canvas was completed for a second time on January 30, 2025, between 3:30 p.m. and 6:00 p.m. During this session, participants identified and prioritized multiple project-related tasks using the GUT Matrix.

[2] https://zenodo.org/records/15732851

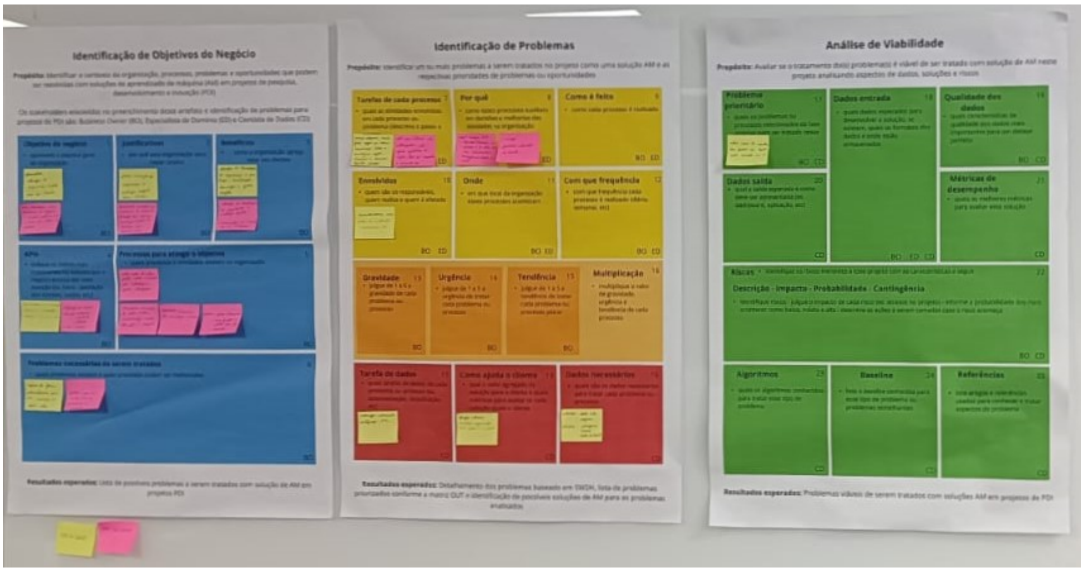

Figure 7.4: *DIP-AI completed by participants in the case study.*

In the days that followed, participants continued to raise questions about specific canvas fields, although they did not resume completion immediately. The process was resumed on February 6, 2025, from 3:20 p.m. to 4:45 p.m., at which point participants were better prepared to engage with the artifact. Having met with company representatives to clarify uncertainties and deepened their understanding of the problem, they completed all remaining fields.

7.4.2 Data Collection

Data collected throughout the experience of the participants applying DIP-AI involved the completed DIP-AI canvas artifact, the follow-up questionnaire, and information on the concrete result of the project.

For RQ1 (contributions to discovery) we considered the overall problem discovery experience, questions about perceptions of contributions of the artifact (questions that we created and that are shown in group 1 in Figure 7.5), and the open question "In the projects in which you participated, what were the difficulties in identifying the problem or machine learning task to be addressed in that project? Could they be mitigated by using this artifact?".

For RQ2 (acceptance), we considered the answers to the TAM-based [67] questions. We selected eight of the TAM 3 constructs (questions of groups 2 to 9 in Figure 7.5), which are: usefulness, ease of use, ability to perform the task using the artifact, control of the artifact to perform the task, pleasure in using the artifact, perceived relevance of the artifact, output quality and demonstrability, and intention to use the artifact.

Finally, for RQ3, we considered the answers to the following open questions:

- What limitations, suggestions or opportunities for improvement do you suggest for this artifact?
- Do you think that questions or fields are missing?
- Do you believe that the current organization of the artifact is clear and adequate?

7.4.3 Analysis Procedures

For the quantitative analysis of the TAM questions, we performed a reliability analysis using Cronbach's Alpha. Reliability indicates how reliably or accurately a questionnaire or test measures a true value. Cronbach's alpha is given by $\alpha = \frac{N\bar{c}}{\bar{v}+(N-1)\bar{c}}$, where N represents the number of items, $\bar{c}$ is the average inter-item covariance, and $\bar{v}$ is the average variance. The quantitative analysis was conducted by aggregating the number of responses for each questionnaire item, based on a seven-point Likert scale to assess the perceived contribution of the framework.

The qualitative analysis of open questions was conducted using the Atlas.TI tool [13], applying open and axial coding procedures from Grounded Theory [66].The method we used was to gather participants' responses to each open-ended question in the questionnaire. We then analyzed the contributions, coded each response, grouped the responses, and established possible relationships. Since we were unable to reach data saturation due to the small sample of participants, we did not perform selective coding.

7.4.4 Results

In Figure 7.5, we present a heat map that represents the responses of the 10 participants to the 36 Likert scale items (the five created ones for artifact contribution, and the 31 from the 8 selected TAM 3 constructs) that form the perception questionnaire of this research. The questions are numbered in the form X.Y where X represents the group of questions and Y enumerates the questions in a group. The values in each field of the heat map represent the number of responses to a given question. We obtained a Cronbach's Alpha of 0.950, which means that the internal consistency of the questionnaire can be considered excellent. Hereafter, we answer each of the research questions.

RQ1: DIP-AI's contribution to problem discovery

Regarding RQ1, considering the overall discovery experience, the team filled the DIP-AI canvas and used it to guide the identification of the problem. In fact, it supported a strategic redefinition of the problem by analyzing risks and opportunities early. The team started working on understanding a particular problem, but then identified risks and a more promising alternative, changing direction. It is noteworthy that at the time of writing, an MVP solving the identified problem was successfully developed and deployed at the industry partner.

Furthermore, based on the first group of questions, it is possible to observe that participants reported that use of the DIP-AI canvas positively influenced project execution by optimizing resources and supporting the achievement of objectives, particularly because problem identification was conducted accurately, efficiently, and effectively.

Additionally, participants acknowledged that the use of the DIP-AI artifact helped address common challenges encountered during the problem discovery phase in innovation and AI-related projects. However, two participants disagreed about the artifact's usefulness for individuals identifying a problem in a project for the first time.

Participants' answers to the question "In the projects in which you participated, what were the difficulties in identifying the problem or machine learning task to be addressed in that project? Could they be mitigated by using this artifact?".

- P3: "I believe that using this artifact right from the beginning could help us look at data and request possible pending issues that may exist."
- P6: "I believe that using this artifact the difficulties could be mitigated"
- P8: "Lack of clarity of processes related to the problem and there is a need to structure the necessary data in a clear way."
- P7: "Lack of clarity in communication with the customer and in the company's processes are difficulties that could be mitigated, since the artifact would help the team prioritize elements of an AI task."

RQ1 Answer: The use of DIP-AI was suitable, and it was considered to contribute positively to the discovery phase of the AI innovation project.

RQ2: DIP-AI overall acceptance

Participants expressed positive perceptions regarding the **usefulness** of the artifact (as reflected in the responses from Group 2). They considered the artifact to be beneficial in several ways: it enhances productivity, increases assertiveness in decision-making, and supports effective problem identification.

Regarding **ease of use** (Group 3), four of the participants reported that the artifact requires a lot of mental effort. And only one student reported that he did not find it easy to use the artifact for what he wanted.

Regarding the group of questions (Group 4) related to the **ability to identify problems** using the artifact, two participants reported that they would have difficulty completing the activity without guidance. Additionally, one participant indicated that relying solely on a support resource or a brief explanation would not be sufficient to carry out the task effectively. These perceptions align with comments from other participants who noted that the artifact is not entirely intuitive or easy to use.

Regarding the perception of **controlling the artifact** (Group 5) to perform the activity, three participants disagreed with having control over the artifact. Perhaps because they consider it not a very customizable tool.

Regarding the dimension of **pleasure** (Group 6) in performing the activity with the artifact, one participant reported not experiencing any sense of enjoyment, and three others stated that they did not find the task engaging or fun. These responses are consistent with the broader perception that the artifact is not particularly easy to use, as it is associated with a cognitively demanding activity that requires significant mental effort.

No negative perceptions were reported regarding the **relevance** of the artifact (Group 7) or the importance of its use in supporting tasks related to AI project development. Similarly, there were no negative comments concerning the demonstrability and quality of the artifact's results (Group 8). On the contrary, participants emphasized that the artifact contributes positively to stakeholder communication, reinforcing its practical value in project contexts.

Finally, nine responses were positive regarding the **intention to use** (Group 9) the artifact in the future.

RQ2 Answer: The use of DIP-AI was well accepted by agile team involved in innovation project.

RQ3: DIP-AI criticisms and suggestions

Participants generally reported that the artifact was well-structured and adequately organized, with no missing fields. They also mentioned that the use of the framework could help mitigate challenges previously encountered in other projects.

Participants indicated that DIP-AI is particularly valuable during the discovery phase of AI innovation projects. Three responses highlighted that the framework contributed to understanding the problem and facilitated communication with stakeholders. Additionally, four participants noted that the DIP-AI canvas played a key role in stakeholder discussions concerning the availability of data, an essential aspect for the success of such projects. Participants also identified specific advantages of using the artifacts: one response emphasized the artifact's role in task prioritization, and three others mentioned that the canvas helped reduce common difficulties in innovation and AI initiatives.

Despite the positive feedback, participants also offered several suggestions for improvement. One participant noted that the artifact is extensive and proposed completing it in iterative rounds. Another suggested that some fields could be merged to reduce the overall complexity. Additional suggestions included the inclusion of a glossary to clarify

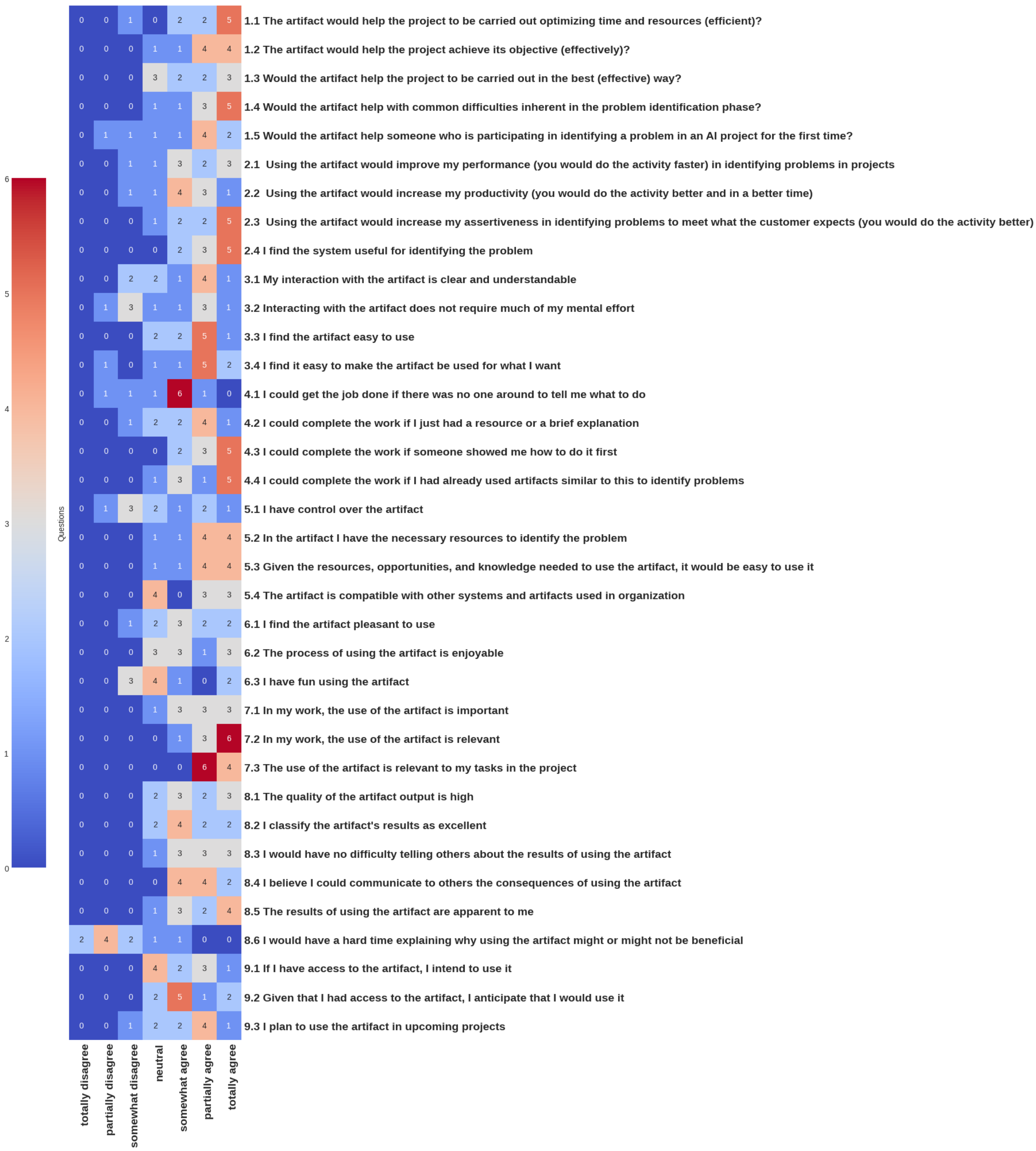

Figure 7.5: *Heat Map of Participant's Perception (TAM 3-based)*

terms and acronyms, as well as providing guidance to assist stakeholders in completing the artifact. One participant recommended enriching the "priority problem"field by linking it to problem characterization — specifically, incorporating information from the GUT matrix and task data into the analysis of findings. Another participant proposed integrating the canvas with agile methodologies, such as Scrum, to periodically reflect on the project's objectives during each sprint. Finally, two respondents suggested that the canvas would benefit from tool support to enhance user interaction and usability.

- P3: "I believe that using this artifact right from the beginning could help us look at data and request possible pending issues that may exist."
- P6: "I would add, along with the algorithms, the characterization of the priority problem. For example: it is a classification problem, regression, supervised learning, unsupervised, it will involve Deep Learning, LLMs."
- P8: "I would like some concepts of the artifact to be explained through a glossary, because if we did not have help, it would be confusing to understand some terms."
- P7: "I would use the artifact as an aid to SCRUM, just like KANBAN. Filling out the artifact in "rounds"is one of its greatest values: giving a temporal notion of the team's perception of the project as a whole. It may even have too many fields, which could lead to exhaustion for those using the artifact and cut short the creative innovation process a bit by keeping the user too tied to a (possibly) long and closed step-by-step process. In my opinion, the artifact could benefit a lot from tool support with an inviting and clean UX."
- P10: "During the dynamic, I felt that some fields were confused with each other, and perhaps, without the help of the researcher, I would not have properly understood what differentiated them. For example, the GUT matrix was a little confusing to me at first, although it was very useful. I found it more practical to use it virtually."

RQ3 Answer: DIP-AI contributed to problem discovery, communicating with stakeholders, and prioritization of problems. We also received suggestions to make the canvas more user-friendly, such as adding a glossary and tool support.

7.4.5 Discussion

DIP-AI is a proposal aligned with the RE phase applied to AI projects. In AI project development processes, there is concern with the Discovery phase, where the problem is identified and characterized; the data is discovered, and the project is prepared for the other phases. Project coordinators pointed out the difficulty in identifying the problem and in building and maintaining this documentation for their projects [47]. This difficulty was also pointed in studies on the state of the art and practice [71, 9]. The approach was developed to help discovering and selecting problems to be addressed in a project.

In the evaluative case study, the proposed framework was perceived as both useful and relevant. The quantitative analysis supports this perception, indicating that the artifact effectively aids in problem identification. These findings are corroborated by the qualitative analysis, which reinforces the positive evaluations observed in the quantitative data. The results suggest that the framework facilitates communication and assists in collecting project-related information that may signal potential future risks, such as the viability and availability of data and models. However, perceptions regarding the ease of use were not uniformly positive. The qualitative feedback provided valuable insights on how to enhance the proposal's usability. Suggestions include simplifying the completion process through the use of a glossary, guided assistance, or iterative completion rounds.

7.4.6 Threats to validity

In this section, we discuss the threats to the validity of our study, according to the categories presented by Runeson *et al.* [60].

In terms of **construct validity**, potential challenges include the influence of assessment apprehension, which may mobilize participants to avoid negative feedback. To mitigate this threat, the questionnaire was administered anonymously. In addition, hypothesis guessing emerges as a potential threat, where participants may have provided overly positive responses, anticipating the research being conducted. To minimize this problem, we clarified that the questionnaire would be used to investigate ways to understand and improve the experience of identifying problems in projects. Participants were encouraged to be honest and critical in order to improve the approach in the future.

For **internal validity**, issues such as diffusion or imitation of treatments come to the fore. Some participants may have sought information about the motivation for some responses from their peers, compromising the integrity of our study. To minimize this threat, we included open-ended questions at the end of the questionnaire to complement the quantitative analysis of the results and better understand participants' perceptions when answering each question. Furthermore, we applied the questionnaire individually and synchronously to avoid the influence of comments among participants.

Regarding **external validity**, we faced the selection and treatment interaction due to the non-representative sample achieved by our study. The participants chosen for the research were currently active in AI innovation projects and had already performed problem identification in previous projects. Despite this, we recognize the importance of carrying out more case study evaluations with different participant profiles and projects for generalization.

Finally, regarding **reliability**, we conducted a single case study. However, as expected for case studies, we carefully described the context, and our case study fills a gap in the literature by identifying problems of innovation and AI projects in an IAC context with a large Brazilian security company. To improve the reliability of our analyses, we peer-reviewed all of our analysis procedures involving two independent researchers and made all data available online for audit in our open science repository.

7.4.7 Final Remarks

This section presented DIP-AI, an approach designed to support the identification of key problems in AI innovation projects. The approach is align through the DIP-AI canvas, which contains 28 fields aimed at capturing business objectives, characterizing problems and opportunities, and analyzing feasibility.

To evaluate the approach, we conducted a case study during the discovery phase of an ongoing AI project, involving ten team members with prior experience in AI innovation projects. Workshops were organized to guide participants in completing the canvas and to clarify questions related to information gathering and customer interaction.

The results indicate positive perceptions regarding the usefulness and relevance of the canvas. However, participants also reported challenges related to ease of use and enjoyment, suggesting improvements such as the inclusion of a glossary and tool support for completing the canvas.

This work contributes to academia by advancing discussions on problem discovery methodologies aligned with requirements engineering standards, such as ISO 12207 and ISO 5338. From an industrial perspective, it demonstrates the practical applicability of the approach in a real AI innovation project. As future work, additional case studies and improvements to the DIP-AI approach will be pursued based on the feedback obtained.

7.5 Discussion of Empirical Findings and Framework Evolution

This section correlates the theoretical challenges identified in the tertiary literature mapping with the practical bottlenecks raised during the interviews with innovation project coordinators. Furthermore, it details how the proposed *DIP-AI* framework addressed these challenges across its three evaluation cycles (postgraduate students, retrospective evaluation with coordinators, and dynamic real-world validation).

The systematic traceability between the identified challenges, their empirical manifestations during validation, and the structural mitigations provided by the framework is mapped in Table 7.1.

Table 7.1: *Traceability Matrix: Challenges vs. DIP-AI Resolutions*

ID	Identified Challenge (Source)	Empirical Evidence in Validation	DIP-AI Treatment and Evolution
C1	**Lack of Stakeholder Alignment** *(Tertiary: S1, S3, S5; Coordinators)*	**Positive:** Participants reported optimization of resources, higher assertiveness, and robust expectation management. In the dynamic validation, it enabled a successful **strategic pivot** prior to MVP development.	**Human-Centered Design:** Integrates *Design Thinking* principles, forcing multidisciplinary teams (business, software engineering, and data science) to co-create early in the cycle.
C2	**Neglect of Data and Model Requirements** *(Tertiary: S1, S2, S4; Coordinators)*	**Positive:** Participant P3 (Dynamic) noted that the framework explicitly forced the team to analyze data availability and identify pending risks at the very beginning.	**Dedicated Domains:** Incorporates tailored sections for data characterization and model limitations, operationalizing technical standards (ISO 5338 and 12207) into practical discovery blocks.
C3	**Uncertainty and NFR Management** *(Tertiary: S1, S2, S4, S5; Coordinators)*	**Positive:** Participant P7 (Dynamic) affirmed that the artifact successfully supports the prioritization of AI task elements under high uncertainty.	**Integrated Risk Matrices:** Embeds structured tools like *5W2H* and *GUT* matrices to map statistical and socio-technical uncertainties (e.g., explainability, model drift) as business hypotheses.
C4	**Cognitive Overload and Usability Friction** *(Emergent finding across all evaluations)*	**Negative:** Significant mental effort was reported by students and coordinators alike. Participants expressed difficulty completing fields without active guidance or pre-existing examples.	**Technological Evolution (LLM):** To mitigate this friction, an assistive tool powered by **Large Language Models (LLMs)** was designed to operate as a conversational agent guiding users through data entry.

7.5.1 Consolidated Synthesis of Evaluation Results

Multidisciplinary Collaboration and Problem Discovery (C1 & C3)

The tertiary literature study indicated a severe shortage of formal Requirements Engineering (RE) approaches designed for upstream discovery, often forcing teams to rely on *ad-hoc* methods under extreme deadline pressures (Studies S7, S8, S9). The interviews with coordinators corroborated this gap, emphasizing that the absence of structured early-stage documentation inevitably triggers technical debt and client misalignment.

Across the evaluation cycles, the *DIP-AI* framework proved its utility in addressing these limitations. Postgraduate students, retrospective coordinators, and dynamic validation team members uniformly agreed that the canvas **optimizes project execution**, preserves valuable organizational resources, and enhances decision-making assertiveness. The most compelling evidence emerged during the **dynamic validation**: while populating the canvas, the development team identified critical technical risks early enough to execute a timely strategic pivot toward a more viable alternative, resulting in the successful deployment of an MVP within the partner industry.

Handling Data and AI-Specific Constraints (C2)

Existing literature (S1, S2, S4) and domain experts stress that engineering requirements for AI-driven systems requires expanding beyond traditional software paradigms to cover data quality, model behavior, and infrastructure constraints.

DIP-AI directly addresses this necessity by guiding engineers and data scientists through structured, question-driven prompts regarding data distribution and algorithmic complexity before a single line of code is written. As highlighted by participant P3 in the real-world deployment, the canvas successfully surfaces data-dependent bottlenecks that typically remain hidden until late-stage integration or production phases.

The Usability Bottleneck and the LLM-Driven Response (C4)

Despite the framework's conceptual success and high perceived usefulness—validated by strong positive scores in the *Usefulness* and *Relevance* dimensions of the Technology Acceptance Model (TAM 3)—all three evaluation loops revealed a significant friction point regarding *Ease of Use* and *Perceived Enjoyment*:

- Participants consistently reported that populating the artifact requires a **high level of cognitive effort** and can be intimidating for less experienced teams.
- Open-ended feedback heavily emphasized the need for centralized repositories, pre-completed examples, and interactive tool support.

Following the evolutionary philosophy of Gorschek's Technology Transfer Model (TTM), this negative feedback was treated not as a failure, but as empirical data motivating the next engineering iteration. To lower the entry barrier and mitigate the identified cognitive overload, this research advanced to design an **assistive LLM-powered tool**. Aligned with the suggestion of participant P7 (who requested "an easier interaction, like a conversational agent, to ask questions and explain them"), the integrated generative AI component acts as an automated facilitator, providing real-time examples, refining entries, and reducing the learning curve of the framework.

7.5.2 Concluding Remarks on Methodological Adherence

The cross-analysis of our empirical data demonstrates that the *DIP-AI* framework successfully fulfills its intended methodological role as an upstream discovery instrument capable of structuring the inherent uncertainty of AI Research, Development, and Innovation (RDI) projects. While baseline approaches like the *ML Canvas* and *PerSpecML* assume that the core problem has already been bounded, *DIP-AI* actively targets the ambiguous phase that precedes formal specification. By evolving from a static canvas into an AI-assisted ecosystem, this approach balances the analytical rigor demanded by software engineering standards with the pragmatic flexibility required for industrial adoption.

7.6 Final Remarks of Chapter

The proposed approach was validated through three complementary stages: an evaluation with postgraduate students, a retrospective evaluation with project coordinators, and a dynamic evaluation conducted in a real AI innovation project. Together, these evaluations provided insights from different perspectives, ranging from initial usability perceptions to practical applicability in real project contexts.

Across the three stages, the results consistently indicated positive perceptions regarding the usefulness and relevance of the framework for supporting problem discovery and structuring discussions about AI innovation projects. At the same time, recurring challenges were identified, particularly related to ease of use, user effort, and overall interaction experience when completing the canvas.

The qualitative feedback collected during the evaluations also generated important suggestions for improvement. These include enhancing usability, providing clearer guidance and examples for completing the fields, and developing tool support to facilitate the interaction with the framework. Participants also highlighted the potential value of features such as repositories of completed canvases and guided assistance during the completion process.

Based on these findings, the next step of this research involves the development of a software tool to support the use of the framework, incorporating Large Language Models (LLMs) to assist users in completing the canvas, clarifying fields, and suggesting relevant information based on project context. This evolution aims to reduce user effort, improve usability, and transform the framework into a more interactive and intelligent support system for discovery activities in AI innovation projects.

CHAPTER 8

DIP-AI Tool

This chapter introduces the computational tool developed to support the DIP-AI framework presented in Chapter 6. Initially designed to facilitate the visualization, sharing, and collaborative completion of the canvas, the tool underwent significant iterations based on initial validation feedback. Participants indicated that the original canvas was overly extensive, potentially hindering ease of use. Consequently, we implemented a streamlined interface and modular workflow to enhance usability and reduce cognitive load.

A key evolution of the tool was the integration of **Large Language Models (LLMs)** to assist users during the elicitation process. This feature addresses a critical bottleneck identified in the literature: the communication gap between technical and business stakeholders [47, 9]. By leveraging LLMs, the tool provides contextual explanations for the canvas fields and suggests domain-specific responses, thereby facilitating a more intuitive and efficient requirements discovery process.

8.1 Tool overview

This section presents the design, implementation, and main features of the DIP-AI Tool, a web-based environment that assists the validated DIP-AI Framework to support structured discovery in AI and Research, Development, and Innovation (RDI) projects. The tool digitizes and extends the framework, enabling collaborative and guided project definition aligned with Requirements Engineering (RE) principles and ISO/IEC standards.

The DIP-AI Tool provides an interactive canvas where multiple stakeholders can collaboratively create and refine discovery artifacts. It supports asynchronous collaboration, ensuring that different users can access and edit shared projects while maintaining traceability and consistency. The tool preserves the conceptual structure of the original framework—covering business understanding, problem identification, prioritization, and feasibility analysis—while incorporating automation and intelligent support through context-adapted questions and AI-generated suggestions that facilitate comprehension and completion. Illustrative images of the DIP-AI Tool are available in our open science repository[1]

8.2 Requirements

The **functional requirements** were derived from the framework's processes and the needs identified in its evaluation. The tool must support: *User management* – registration, authentication, and access control; *Project management* – creation, sharing, and version control of discovery projects; *Canvas-based interaction* – visual building blocks that structure discovery according to framework phases; *Collaborative completion* – asynchronous, multi-user editing of shared artifacts with traceability; and *Export and documentation* – generation of outputs in PDF and JSON formats for reuse and reporting.

Key **non-functional requirements** emphasize: *Usability* - an intuitive, responsive interface following UX best practices to reduce cognitive load; *Security* - authenticated access and permission control ensuring data confidentiality and integrity; *Reliability and persistence* - robust data management ensuring ACID compliance in the PostgreSQL database; *Compliance* adherence to ISO/IEC/IEEE 12207 and ISO/IEC 5338 process guidelines for AI and software system development.

As shown in Figure 8.1, after logging into the tool, a project management screen is presented, where it is possible to view, edit, and delete the logged-in user's existing projects and projects shared with that user. It is also possible to create a new project. When accessing a project, it is possible to view the entire canvas or select parts of the canvas to be presented in blocks to facilitate completion and discussion with stakeholders. The user can also disable the suggestions generated by the LLM. Each field has a number indicating the order in which it should be completed, the field title, the original question, and an icon identifying the type of stakeholder who could help to provide that answer.

Figure 8.2 highlights a field selected for completion with the LLM suggestion turned on. This allows the user to see the original question from the canvas, the contextualized question (in blue) that helps with understanding the completion and shows the field for the user to answer, and the suggested answer generated by the LLM (in green) that can be used and edited by the user.

Once the question is saved with LLM suggestion enabled, the canvas updates, and the message "generating suggestions"appears. At the end of processing, the canvas and suggestions are updated, and the message "All suggestions updated"appears. The tool also displays the completed fields, detailed and marked with "filled."

[1] https://zenodo.org/records/17379596

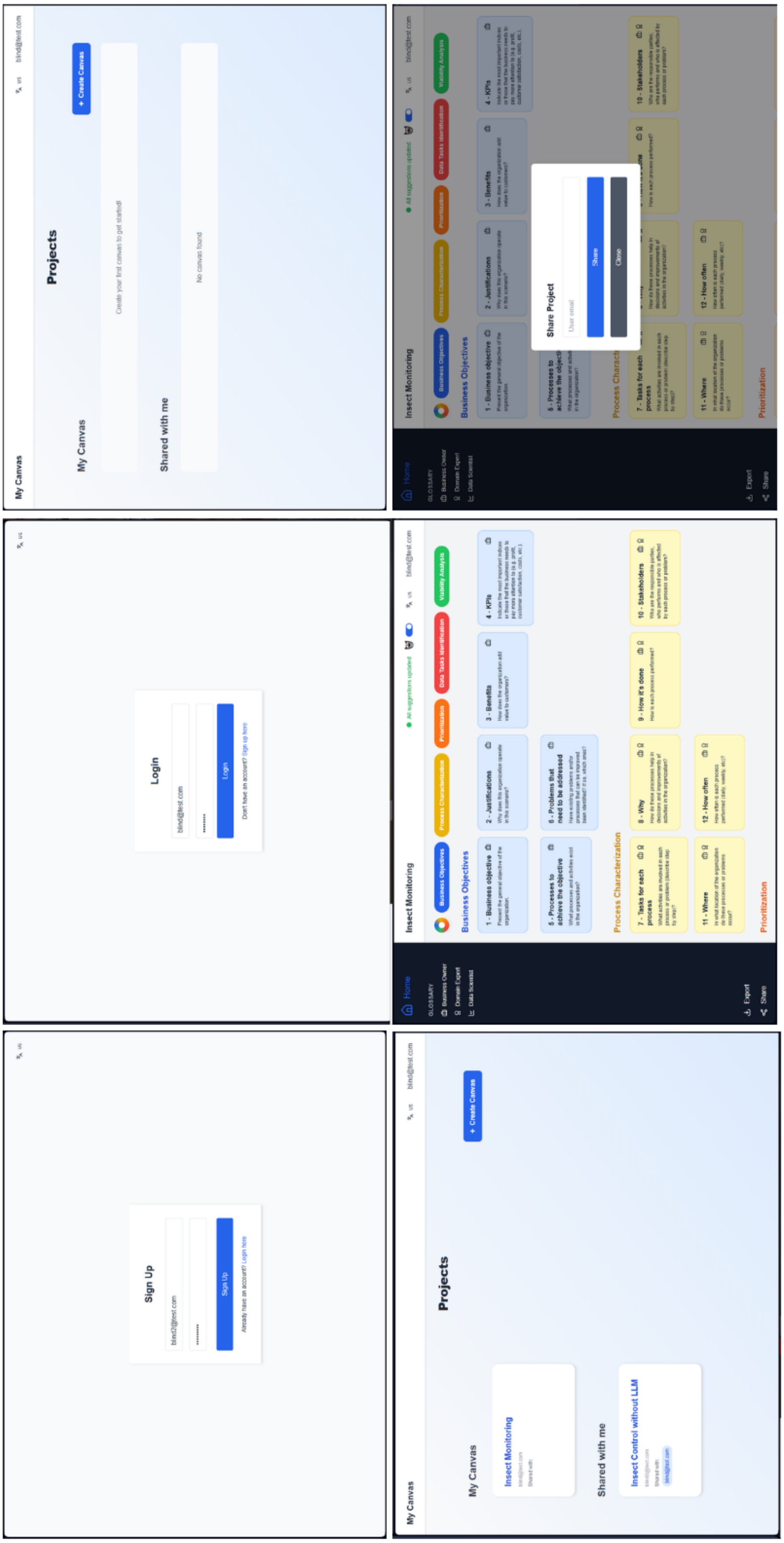

Figure 8.1: *Interface showing project management (personal and shared projects) and canvas view with LLM-generated suggestions.*

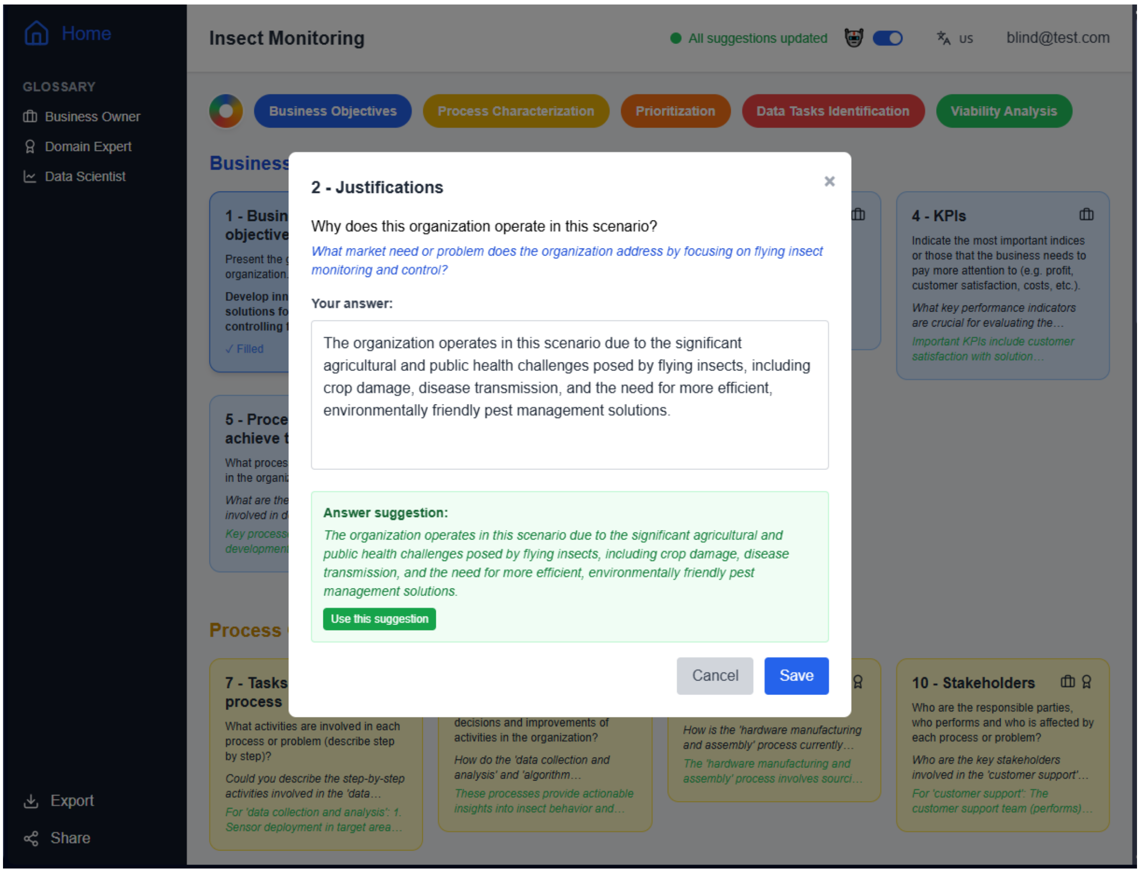

Figure 8.2: *Interface showing the rewriting of a canvas question (in blue) and an automatically generated answer suggestion (in green) produced by the LLM module.*

8.3 Development

The tool's development followed an iterative approach grounded in the feedback obtained from the DIP-AI Framework evaluation, which highlighted three key limitations of the manual version: (i) cognitive effort during canvas completion; (ii) lack of automated guidance; and (iii) difficulty maintaining consistent documentation.

To address these gaps, the development focused on three main components: Account and project management, enabling controlled access and shared workspaces; Interactive canvas visualization, presenting a structured representation of the framework's discovery process; and Intelligent assistance via LLMs, introduced to reduce cognitive effort and support users through contextualized guidance.

The tool was developed using React on the frontend for a scalable and responsive interface and FastAPI (Python) on the backend for efficient asynchronous processing. These technologies were selected for their compatibility with modular, service-oriented architectures and their ability to handle concurrent user interactions smoothly.

In addition to the previously elicited requirements, we adopted the LLM module for **Question Rewriting**: dynamically reformulates each canvas question in simpler, domain-oriented language to clarify its intent; and **Answer Suggestion**: proposes initial responses based on prior project information and user-provided context. This intelligent guidance mitigates ambiguity, reduces manual effort, and improves user engagement in the discovery process.

8.3.1 Architecture

The DIP-AI Tool adopts a modular, service-oriented architecture composed of three layers (Figure 8.3):

- The frontend layer, developed in React, offers a responsive and modular interface that supports user authentication, project management, and interactive canvas completion.
- The backend layer, implemented using FastAPI (Python), manages data persistence, user requests, and communication with external AI services.
- The intelligent assistant module connects to an LLM (currently Gemini 2.5 via API), generating contextualized question reformulations and answer suggestions. The design allows seamless integration with alternative LLM providers.
- The PostgreSQL database stores project data, user information, and canvas content, ensuring ACID-compliant operations and persistent traceability.

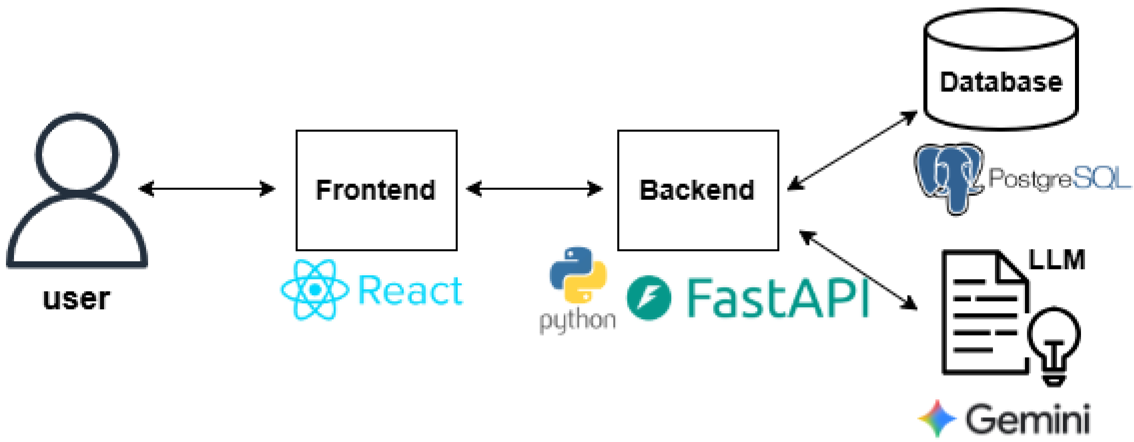

Figure 8.3: *DIP-AI Tool architecture.*

8.4 Asynchronous Processing Flow

To address one of the main challenges in integrating LLMs into interactive environments—the latency and variability of response time—the DIP-AI Tool employs an asynchronous request management architecture (Figure 8.4).

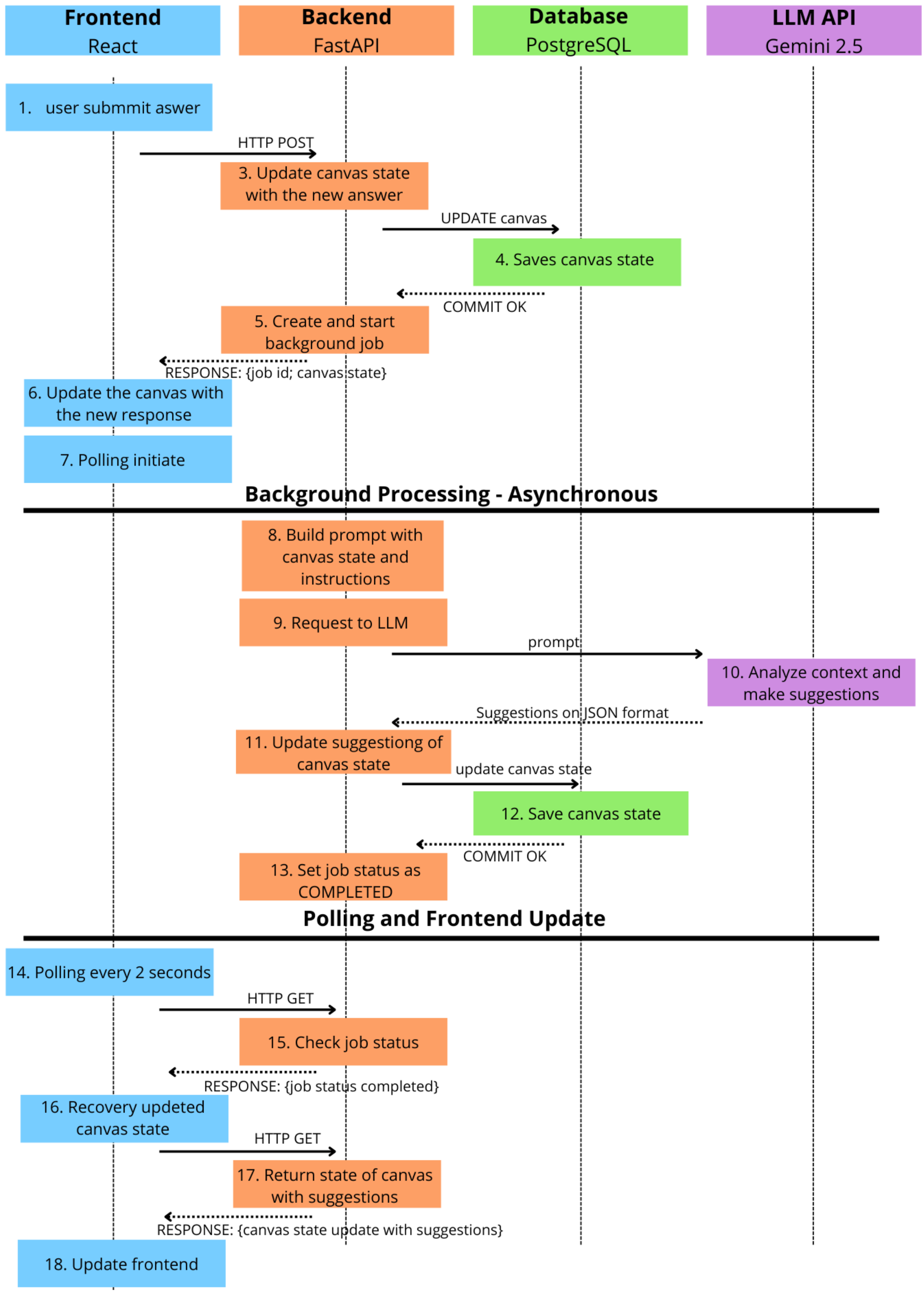

Figure 8.4: *Asynchronous processing of the tool. The sequence diagram illustrates the interaction between the frontend, backend, database, and LLM API.*

When a user submits an entry on the canvas, the system immediately records it in the database and updates the interface (Steps 1–6), allowing the user to continue editing other fields. In parallel, a background process (Steps 8–13) asynchronously sends the contextualized prompt to the LLM, retrieves suggestions, and stores them for later retrieval. The frontend periodically polls the backend for updates (Steps 15–19), displaying new LLM-generated content without interrupting user interaction.

This design ensures non-blocking interaction, smooth user experience, and efficient utilization of computational resources. Additionally, users may choose to disable the LLM module entirely, in which case the system executes only the direct data update flow (Steps 1–6), maintaining full functionality even without AI assistance. This hybrid execution model—combining synchronous data management and asynchronous AI augmentation—offers flexibility for environments with limited connectivity or privacy constraints while preserving the benefits of intelligent assistance where available.

8.5 Discussion

The tool improves the practical applicability of the DIP-AI Framework, transforming it into a collaborative and intelligent environment for project discovery. By equipping the framework's structured guidance, the tool provides requirements engineers and stakeholders with a systematic approach to understanding business objectives, identifying AI opportunities, and assessing technical feasibility.

LLM adoption suggests that intelligent assistance can reduce cognitive effort, improve communication among stakeholders, and enhance completeness of discovery artifacts. The LLM assistant clarify ambiguous terminology, provide context-sensitive examples, and reduce effort during completion. Too help gain agility in filling out, improved understanding of fields, and discussion of problems and opportunities were noticed. The asynchronous architecture also proved effective in mitigating response latency, maintaining usability during interactive sessions.

The tool described in this research represents an innovative solution to advance the integration of RE and AI engineering by embedding intelligent assistance in early development stages to promote traceability, collaboration, and methodological consistency.

8.6 Final Remarks

This chapter presented the DIP-AI Tool, a web-based environment that assists the framework to support structured discovery in AI and RDI projects. The tool assists multidisciplinary teams in identifying, prioritizing, and assessing the feasibility of AI opportunities through a standards-based process aligned with ISO/IEC 12207 and 5338.

In summary, the tool represents a concrete step toward intelligent, useful, and collaborative discovery of AI innovation projects, combining methodological rigor with the generative capabilities of modern language models.

In summary, the tool transforms the conceptual framework into a practical, interactive, and intelligent environment for project discovery. Its modular design, asynchronous LLM integration, and standards-based structure enable effective collaboration and reduce cognitive load throughout the early phases of AI and innovation project definition.

CHAPTER 9

Validation with Experts

This chapter presents the final evaluation of the DIP-AI framework and its supporting tool. The primary objective of this stage was to gather expert feedback to validate the proposal's efficacy, usability, and practical relevance in real-world AI innovation contexts.

To achieve this, we adopted the **Focus Group** as our primary data collection method. This approach allows for an extension of the individual interview format into a dynamic group discussion, fostering a collaborative environment where experts can deliberate on the proposed solution. As defined by Morgan [50], the focus group is a "research technique that collects data through group interaction on a specific topic chosen by the researcher."

In the context of this research, the focus group method was particularly valuable for simulating the multi-stakeholder interaction required by the DIP-AI Canvas. By bringing together project coordinators, data scientists, and domain experts, we were able to observe the synergy between different roles and identify how the framework facilitates communication and mitigates the challenges previously identified in the empirical stages of this study.

9.1 Focus Group Design

To perform a comprehensive evaluation of the DIP-AI framework and assess its practical value, we designed a focus group study targeting professionals in the AI and innovation fields. The recruitment strategy followed the "over-invitation"principle to mitigate the common risk of no-shows in professional settings. We initially invited 15 experts with diverse backgrounds; of these, 12 confirmed their availability.

We conducted two synchronous remote focus groups following the methodology proposed by Kontio [44]. The sessions were moderated by the lead researcher, supported by a co-author responsible for logistics, recording, and technical monitoring (audio, video, and connectivity). The first session included 3 professionals (after two last-minute cancellations), while the second session comprised 5 participants. Despite the cancellations, the group sizes remained within the functional range for qualitative depth in software engineering research.

The sessions were structured into a 90-minute protocol, organized as follows:

1. **Opening and Contextualization:** Participant introductions and a formal presentation of the DIP-AI framework.
2. **Framework Deliberation:** Participants were encouraged to provide critical feedback and suggestions for improvement based on their professional experience with AI discovery activities. Contributions were organized in rounds to ensure balanced participation and avoid "groupthink."
3. **Tool Demonstration and Feedback:** The moderator presented the supporting tool, highlighting functionalities such as canvas exportation and the **Large Language Model (LLM)** integration for data entry support. A second round of discussion followed, focusing on the tool's usability and technical feasibility.

To ensure engagement, reminders were sent 24 hours and 30 minutes prior to each session, including the meeting link, the Informed Consent Form (ICF), and a profile questionnaire.

9.2 Data Collection

Data collection was governed by audiovisual technology guidelines for online qualitative research [20]. The sessions were hosted on a secure video-conferencing platform with explicit participant authorization for recording.

Two primary data streams were utilized to ensure triangulation:

- **Audiovisual Records:** Synchronous interactions were recorded and subsequently transcribed. This allowed for a precise capture of participants' insights, nuances in discourse, and instances of verbal agreement or dissent.
- **Observation and Field Notes:** While the moderator led the discussion, a note-taker documented non-verbal cues and technical incidents, as recommended by Marques et al. [23]. These notes provided context for the transcription analysis.

To safeguard data quality, a "pilot-test"of the audiovisual infrastructure was conducted prior to the sessions. This prevented connectivity issues from interrupting the flow of the qualitative inquiry, ensuring a stable environment for expert deliberation [74].

9.3 Analysis Procedures

The qualitative analysis followed a systematic approach consistent with empirical software engineering research. The procedures were structured into three main phases: transcription, thematic coding, and data triangulation.

9.3.1 Transcription and Data Preparation

The transcription and cleaning of the data were fundamental to ensuring the integrity of the subsequent analysis phases. Audio records were converted into text using automated tools followed by manual verification to ensure technical accuracy. During this stage, we strictly enforced anonymity protocols to protect participants' identities, a crucial step for generating unbiased artifacts for the coding phase. Finally, a discourse refinement was performed to ensure the quality of the text while preserving the participants' original intent and context.

9.3.2 Coding and Categorization

For the coding stage, the transcribed material was processed using the **Atlas.TI** software. We followed a systematic coding protocol:

- **Open Coding:** Initially, the text was fragmented into key excerpts to identify primary contributions and emerging concepts.
- **Axial Coding:** These codes were then grouped into thematic categories. Specifically, we identified clusters related to the *DIP-AI Tool*, the *Framework*, and *Future Research Directions*.

Regarding the **Tool**, the analysis revealed specific feedback on usability improvements, the integration of **LLM** to support user interaction, and the perceived benefits of automation. Concerning the **Framework**, participants suggested the inclusion of new fields and highlighted positive aspects regarding its structural logic.

9.3.3 Triangulation and Synthesis

To enhance the validity of the findings, we performed a multi-source triangulation. We conducted a content analysis of the two focus groups to identify the most frequent and significant contributions. These results were then cross-referenced with data from questionnaires and the discourse analysis. Finally, the focus group outcomes were compared with findings from the earlier interviews (Chapter 4) and the tertiary study (Chapter 5). This holistic approach allowed us to refine the framework and the tool based on a solid foundation of both literature and practice.

9.4 Profile of Participants

Regarding the profile of the Focus Group participants: We had 3 participants with ongoing doctoral studies, 3 with completed doctoral degrees, and 2 with completed master's degrees. One participant has between 1 and 3 years of experience as a data scientist; another participant has more than 10 years of experience in this position, and the others stated having between 3 and 10 years of experience. Three participants stated having experience as a project manager. Six participants stated having between 1 and 10 years of experience as an AI engineer.

Four participants stated experience in the public sector. Four participants stated experience in startups. Five participants stated working in consulting. Six participants worked in the private sector, and five of them worked in organizations with more than 100 employees.

Regarding the types of innovation projects, 7 participants stated working on R&D projects, and 7 participants stated that they had done innovation projects for digital products. All participants worked on AI and MV projects; two also worked in open innovation, and three worked on digital transformation projects.

Participants also reported on their roles in these projects. Six participants stated they worked as leaders or managers, seven worked in technical AI development and research. Six participants worked in software development, and five stated they worked in consulting.

Seven participants stated they have extensive experience with ML. Five of them declared having significant experience in Deep Learning, and four declared the same expertise in recommendation systems. Five participants stated they have experience with model languages, and four confirmed this expertise in AI agent development.

Regarding requirements in AI projects, five participants stated they have extensive experience in discovery, solution ideation, requirements specification, and validation. Two participants stated little experience in project discovery, and one participant stated they have experience in discovery. We observed no differences in profile between the participants of Focus Group 1 and the participants of Focus Group 2.

9.5 Results

The participants evaluated the framework based on the TAM identified with a questionnaire. Table 9.1 presents the contributions obtained in the Focus Group for the framework and for the tool regarding ease of use, usefulness, and intention to use in the future. We obtained positive perceptions regarding ease of use, usefulness, and intention to use in the future for both the framework and the tool.

We observed that the average contribution for ease of use for the framework was 5.857, while for the tool it was 5.714. The average usefulness score for the framework was 6.000, while for the tool it was 5.428. Finally, the intention to use the framework was evaluated with an average of 6.428. The intention to use the tool was evaluated with an average value of 5.857. Therefore, it is noted that professionals see more value in the framework and its contribution, ease of use, and demonstrate interest in using it than in the tool. Despite this, we see that the evaluation was still positive for the tool regarding the same aspects.

We conducted transcripts of the focus group and performed thematic analysis. We performed open and axial coding and analyzed the results. We were able to group the codes according to negative and positive perceptions regarding the framework, perceptions about the use of the tool, and about the use of LLM to assist in data entry.

In Figures 9.1, 9.2 and 9.3 we observe the contributions obtained in the Focus Group regarding the proposed framework, the supporting tool. The coded data were organized into three analytical dimensions: framework-related insights, tool-related perceptions an the use of LLM, and research and future-use perspectives.

9.5.1 About the framework

In first dimensional (in purple) presented in Figure 9.1 The participants indicate a need of emphasis on compliance in the framework. Among all framework-related codes, compliance emerged as the most frequently mentioned concept (n = 8), highlighting participants' concern with regulatory alignment, accountability, and risk management in AI-driven projects.

Participants referred to the need to explicitly address project constraints, including infrastructure (n = 1), cost (n = 1), security (n = 1), people (n = 1), and environmental limitations (n = 1). These constraints were conceptually grouped under the operational card (n = 4), which was perceived as a central artifact responsible for consolidating contextual and organizational factors that influence project feasibility.

Decision-support artifacts such as the project baseline, go/no-go decisions, ROI card, and project impact assessment were discussed as complementary elements that enable structured evaluation throughout the project lifecycle. Together, these mechanisms support informed decision-making, particularly in contexts involving AI adoption.

Regarding methodological aspects, participants emphasized the importance of formalizing the discovery process and explicitly defining its scope (n = 4). The identification of KPIs was perceived as a relevant. The filling of framework was perceived as non-exhaustive activity. The framework's flexibility and adaptability to different project contexts. Highlighting of importance of framework and contribution for practical and art state (n = 9).

Validation checklist (n = 3) and the validation meta layer (n = 5) were highlighted as important to ensure consistency and rigor. These elements were perceived as safeguards against incomplete or incoherent documentation, particularly when multiple stakeholders are involved. As a form to obtain the validation between stakeholders.

9.5.2 About tool and LLM use

The second analytical dimension (in orange) presented in Figure 9.2 focuses on participants' perceptions of the tool and its integration with Large Language Models (LLMs). Overall, the results reveal a predominantly positive perception of LLM usage, with participants recognizing its advantages for efficiency and support in documentation tasks. Despite this positive outlook, hallucination risks (n = 2) were explicitly raised, indicating awareness of potential limitations when relying on generative models. To mitigate these risks, participants proposed several improvement mechanisms, most notably the use of guardrails (n = 3), minimum response thresholds before generation (n = 2), and the integration of knowledge bases (n = 2).

Transparency and traceability also emerged as relevant concerns. Suggestions such as identifying sources of information, providing textual justification for generated answers, and incorporating user feedback mechanisms reflect an effort to align LLM-generated outputs with established quality and accountability standards.

From a usability perspective, the tool was perceived as an enabler of operational efficiency. Features such as filters, better organization, and structured generation of artifacts were associated with reduced time and effort, as well as with facilitating the deployment of the framework in real-world projects.

9.5.3 About future works

The third dimension (in green) presented in Figure 9.3 captures participants' perceptions regarding the research contribution and future applicability of the framework and tool. The importance of the contribution was the most frequently mentioned code across this category (n = 9), indicating a strong perceived relevance of the proposed approach.

Participants expressed interest in future adoption of the framework and emphasized its flexibility, particularly its ability to support reasoning, decision-making, and memory of what is most relevant during discovery and planning phases. These characteristics were associated with the framework's potential sustainability beyond the study context.

		Framework			Tool		
Focus Group	Participant	Ease of Use	Usefulness	Intention of use	Ease of Use	Usefulness	Intention of use
1	P1	6	7	6	7	5	5
1	P2	7	5	6	7	5	6
1	P3	5	7	6	5	7	6
2	P4	5	6	6	5	6	6
2	P5	7	7	7	6	6	6
2	P6	5	6	7	5	5	5
2	P7	6	4	7	6	4	7

Table 9.1: *Perceptions of TAM about Ease of Use, Usefulness and Intention of use in the future for the Framework and the Tool*

Concerns about knowledge transfer and adoption strategies were also discussed, including the use of workshops and explicit explanations of the framework to teams. Additionally, participants highlighted the generation of project documentation using LLMs after the discovery phase as a valuable outcome.

Finally, evaluation-related codes—such as applying the framework with clients and assessing usability and demonstrability—reinforce the empirical grounding of the proposal and point to clear directions for future validation in applied settings.

Overall, the qualitative results suggest that the proposed framework is perceived as a structured yet flexible approach that addresses critical governance and decision-making needs in AI projects. The supporting tool, particularly through the controlled use of LLMs, was viewed as a valuable enabler, provided that appropriate safeguards are in place. Finally, the strong emphasis on perceived contribution and future use highlights the framework's potential impact on both research and practice.

9.6 Threats to validity

To ensure the rigor and credibility of our qualitative findings, we identified and addressed potential threats to validity following established guidelines for online focus groups and empirical software engineering.

- **Internal Validity:** A primary concern in synchronous online research is technical instability, which can lead to fragmented discussions and loss of data. Following the recommendations of Willemsen et al. [74] and Daniels et al. [20], we mitigated this by providing participants with detailed audiovisual instructions and technical support prior to the sessions. Additionally, a dedicated technical assistant was present during all focus group sessions to resolve connectivity or platform issues in real-time.
- **External Validity (Generalization):** While qualitative studies do not aim for statistical generalization, we sought to enhance the representativeness of our findings

by inviting participants from diverse organizational contexts and project domains. Following Kontio's guidelines [44], the group included experts with varying levels of experience in AI development across different industries. However, we acknowledge that the results reflect specific RDI (Research, Development, and Innovation) ecosystems and may not fully represent all industrial AI settings.

- **Construct Validity (Interaction Bias):** To mitigate the risk of dominant participants or researcher bias, we strictly followed a moderation protocol inspired by Marques et al. [23]. The moderator ensured balanced participation, specifically prompting those who were less vocal to contribute. This strategy was crucial to minimize "groupthink"and ensure that the evaluation of the DIP-AI framework and its LLM integration was based on a consensus of diverse perspectives.
- **Reliability (Dependability):** To ensure the reliability of our qualitative analysis, we implemented a peer-review process where two independent researchers coded the data. Any discrepancies were discussed and resolved through consensus. Furthermore, to support transparency and accountability, we have made all anonymized artifacts, tools, and analysis procedures available in an open-access scientific repository, allowing for future replication or auditing of our results.

9.7 Final Remarks

This chapter presents the final validation of the research, consolidating expert perceptions on the framework and the support tool. The evaluation used a mixed-methods approach: quantitative, through the Technology Acceptance Model (TAM), and qualitative, through Thematic Analysis of the focus group transcripts.

9.8 Quantitative Evaluation

The results demonstrate a positive acceptance of both the framework and the tool in three main dimensions: Ease of Use, Usefulness, and Intention to Use.

Professionals consistently assigned higher scores to the framework compared to the tool. The intention to use the framework reached the highest average (6.428), while the perceived usefulness of the tool was the lowest (5.428), although still within the positive range.

Experts perceived superior strategic value in the framework (process and structure), while the tool is perceived as an operational facilitator that still has room for interface refinements.

9.8.1 Qualitative Analysis

The thematic analysis, carried out through open and axial coding, allowed the experts' contributions to be organized into three critical dimensions:

Perceptions of Framework

The most discussed dimension was Compliance and Governance. The code "Compliance"was the most frequent (n=8), revealing that, in the context of AI, regulatory traceability and risk management are top priorities.

Operational Constraints: Participants emphasized that the framework should explicitly state limitations regarding infrastructure, cost, and security.

Flexibility: The framework's adaptability to different project contexts was praised, being seen as a non-exhaustive guide that supports decision-making (Go/No-Go).

Perceptions of LLM use

The integration of Large Language Models (LLMs) was viewed predominantly positively for increasing documentation efficiency. However, technical concerns arose:

Risk Mitigation: Guardrails (n=3) and integrated knowledge bases were suggested to prevent model hallucinations.

Transparency: Experts requested mechanisms to identify the origin of information generated by AI and provide textual justifications for suggestions given by Canvas.

Contributions to Research

The relevance of the research was widely validated (n=9 mentions of the importance of the contribution).

Knowledge Sustainability: The framework was highlighted for its ability to act as "organizational memory"during the discovery phases.

Suggestions include applying the framework directly with end clients and automating the generation of post-discovery phase project documents using LLM.

9.9 Synthesis

The results suggest that the framework fills a critical gap in AI project governance, offering a structured yet flexible approach. The tool, enhanced by the controlled use of LLMs, acts as a productivity accelerator, provided that ethical and technical safeguards are rigorously implemented.

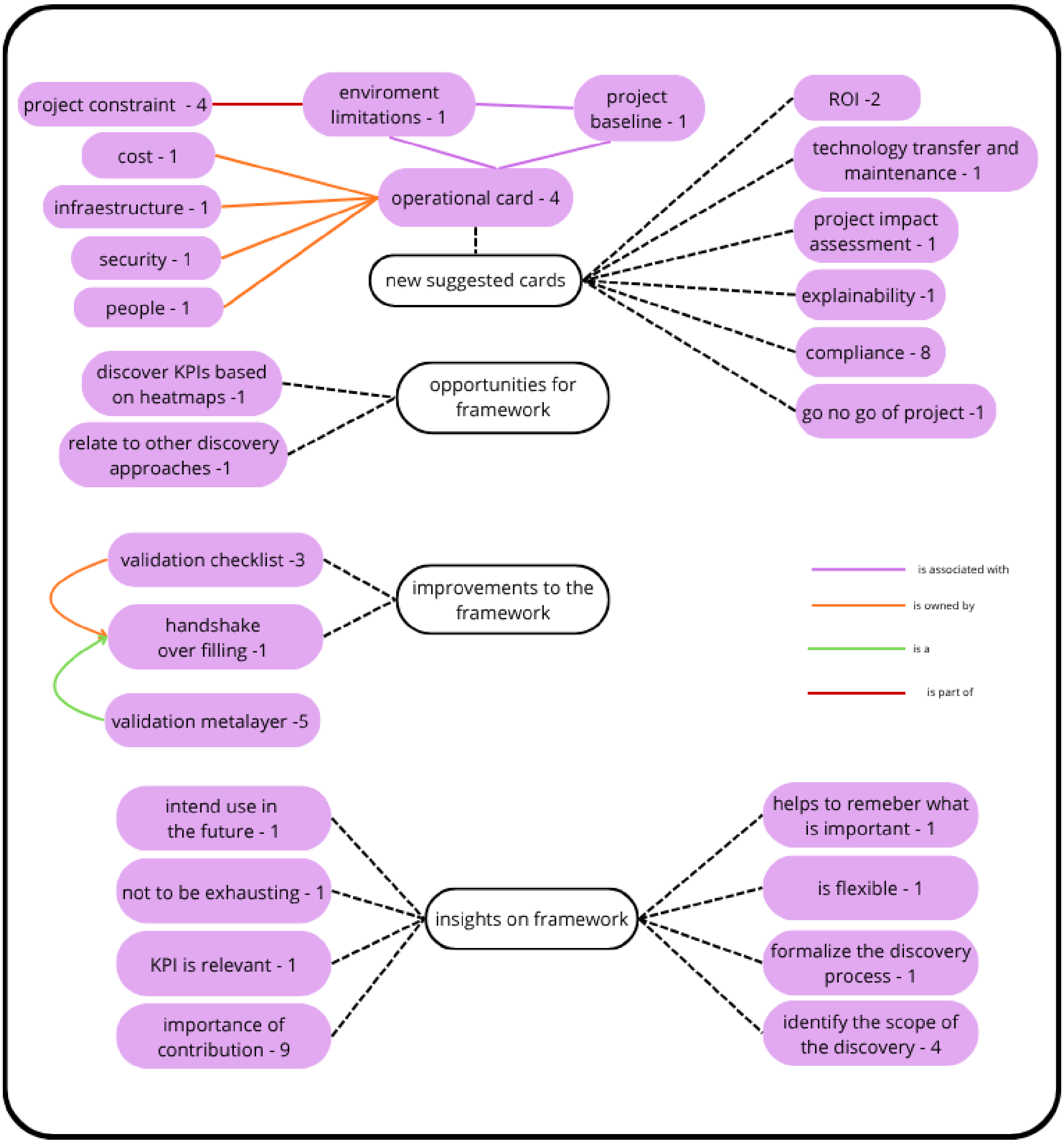

Figure 9.1: *Contributions of Focus Group -*

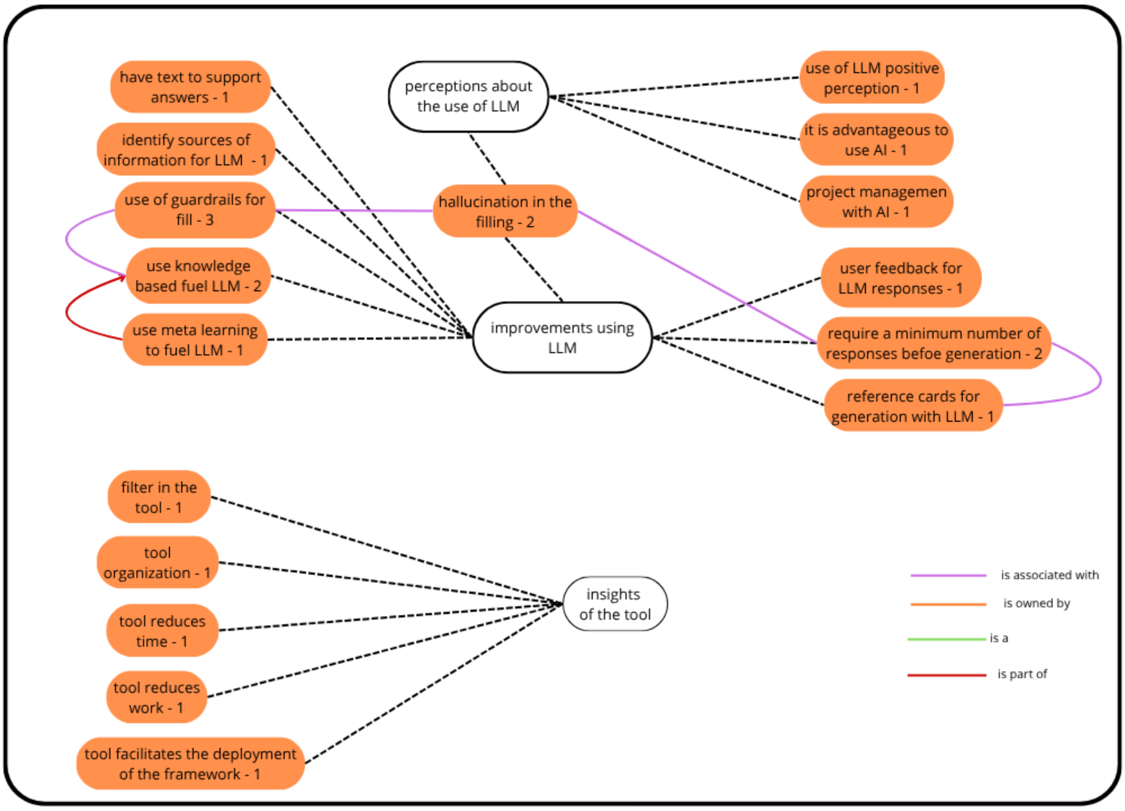

Figure 9.2: *Contributions of Focus Group.*

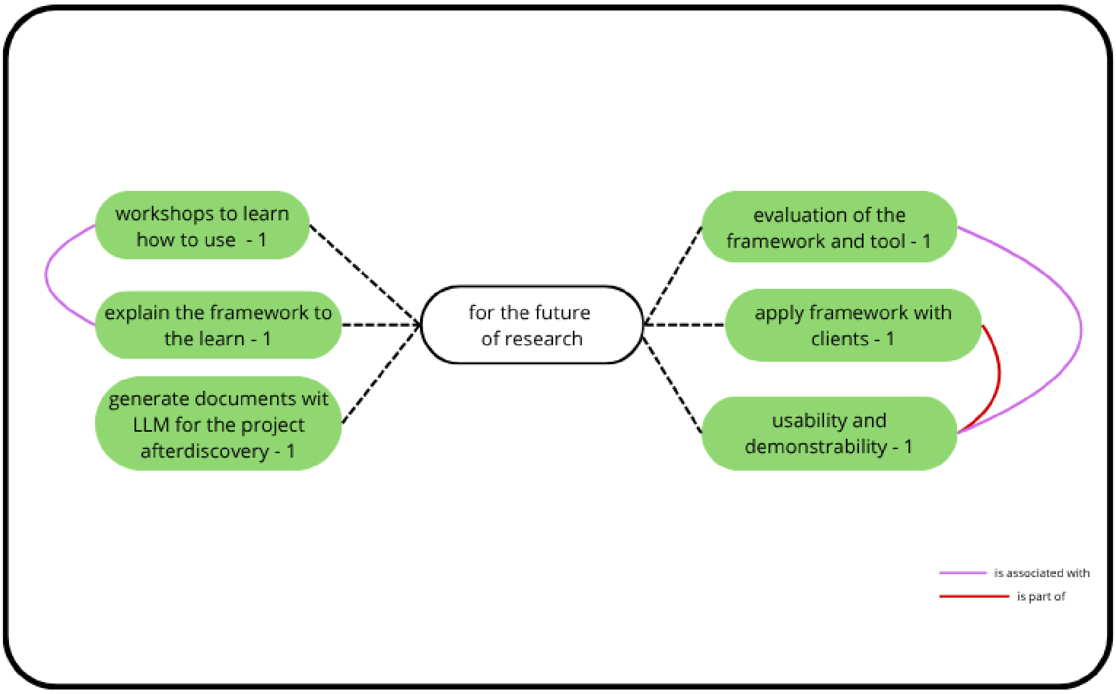

Figure 9.3: *Contributions of Focus Group.*

CHAPTER 10

Conclusions

This chapter synthesizes the research journey, reflecting on the objectives achieved and the contributions made to the field of Requirements Engineering (RE) for Artificial Intelligence (AI). We present a summary of contributions, methodological reflections, a final assessment of the research questions, and a roadmap for future investigations.

10.1 Summary of Contributions

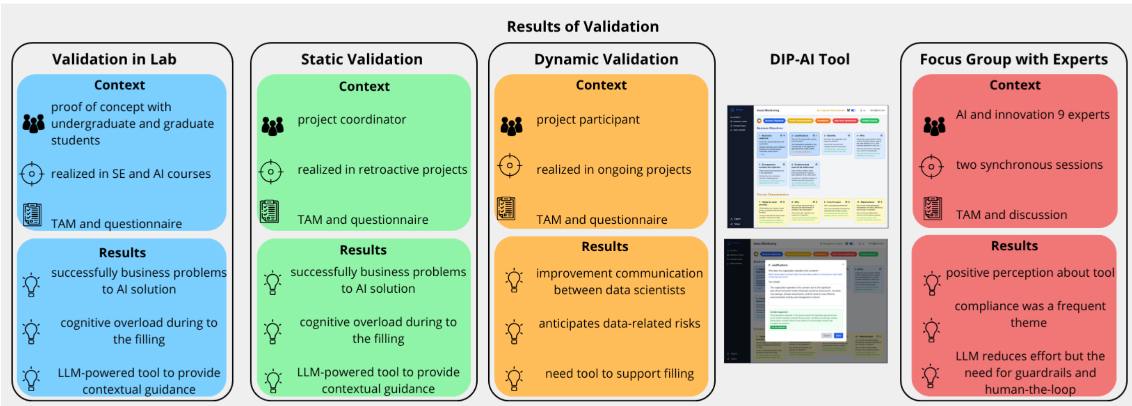

Figure 10.1: *Comprehensive Summary of Framework Validation Results*

The process applied in this thesis presented in Figure 6.1 advanced the state of the art by providing an in-depth investigation into the state of practice of RE within AI-driven innovation projects. Our empirical discovery phase identified critical bottlenecks, such as the difficulty of specifying requirements for experimental software, aligning business objectives with data feasibility, and managing the high expectations of stakeholders in non-deterministic systems.

Other contribution of this work is the tertiary study, results contrasting existing RE research with real-world recurring challenges. It identifies that current literature heavily focuses on elicitation and ad-hoc techniques like brainstorming. However, practical projects struggle with non-functional requirements (NFRs), stakeholder communication, lack of clear metrics, and uncertain customer expectations.

To address these challenges, we introduced the **DIP-AI Framework**, a hybrid approach that harmonizes the structural rigor of **ISO/IEC/IEEE 12207** with the creative agility of **Design Thinking** and **5W2H**. The framework was iteratively refined through the **Technology Transfer Model (TTM)**, ensuring that each version incorporated feedback from real-world practitioners as presented in Figure 10.1.

Furthermore, to overcome the usability limitations identified during initial trials, we developed a dedicated supporting tool. A significant contribution of this work is the integration of **Large Language Models (LLMs)** into the tool, which facilitates the elicitation process by providing context-aware guidance and automated suggestions. This integration proved essential in bridging the communication gap between technical teams and business owners, ultimately enhancing the quality and completeness of the requirements discovery phase.

10.2 Answering the Research Question

The central research question of this study was: *"How does the framework structure and support the requirements discovery phase in AI-based innovation projects?"*

Our findings demonstrate that the **DIP-AI Framework** effectively supports the discovery phase by providing a structured yet flexible workspace (the Canvas). The final evaluation through Focus Groups confirmed the framework's **effectiveness** in identifying viable AI problems and its **utility** in providing a clear roadmap for project feasibility. While early iterations showed limited usability, the subsequent development of the LLM-powered tool significantly improved the user experience, making the process more intuitive for non-experts.

10.3 Methodological Reflections

The adoption of the **Technology Transfer Model (TTM)** was pivotal in bridging the gap between academia and industry. The partnership with prominent innovation hubs (CEIA and ExACTa) allowed the research to be grounded in actual industrial problems from inception to final validation.

The use of Mixed Methods enabled a robust triangulation of data. The quantitative rigor of the **Technology Acceptance Model (TAM3) [67]** provided objective measures of perceived usefulness and ease of use, while **Grounded Theory** (following the Strauss and Corbin approach [16]) offered deep qualitative insights into the underlying processes of AI discovery through open, axial, and selective coding. Finally, the **Focus**

Group method [44] proved to be a cost-effective and highly interactive environment for validating the framework with seasoned professionals, yielding insights that individual interviews could not capture.

10.4 Limitations

Despite the positive results, some limitations must be acknowledged:

- **Contextual Scope:** While the framework was validated within the EMBRAPII/CEIA ecosystem, the findings may reflect specific regional or organizational cultures. Further validation in diverse global RDI settings is required to enhance generalization.
- **Technological Frontier:** The current version of DIP-AI was designed prior to the explosion of autonomous AI Agents. Its efficacy in projects exclusively centered on Agent Workflows has not yet been fully explored.
- **Non-Functional Requirements (NFRs):** Although the framework addresses feasibility and performance, it does not yet provide specialized sub processes for deep ethical audits, security protocols, or explainability metrics (XAI).
- **Reliance on Perceptual Evidence:** This method relies heavily on qualitative, perceptual evidence, which introduces subjective bias. What users say they do or think often deviates from their actual behavior. In software or system evaluation, relying solely on how participants perceive system performance lacks empirical validation, as subjective satisfaction does not always correlate with actual system efficiency or accuracy.
- **Absence of a Baseline Comparison:** A critical limitation of this evaluation is the lack of a baseline comparison. Without establishing a clear benchmark—such as a previous version of the system, a standard industry alternative, or a control group—it is impossible to isolate variables or definitively prove that the proposed solution represents a quantifiable advancement. The feedback remains isolated, making it difficult to gauge true progress.
- **Focus Restricted to ML Projects and Specific Evaluation Contexts:** The evaluation scope is heavily restricted to Machine Learning projects and tightly bound to a specific evaluation context. Because ML systems possess unique traits—such as probabilistic outputs, data dependency, and evolving performance—the feedback collected here is highly specialized. Consequently, these findings cannot be easily generalized or applied to traditional software engineering projects or differing organizational contexts.

10.5 Future Work

The following avenues for future research are proposed:

- **Adaptation for AI Agents:** Investigate the necessary extensions for the DIP-AI Canvas to support the specificity of discovery of projects in LLM-based agents and multi-agent systems.
- **NFR Meta-Layers:** Develop specialized "plug-ins"or meta-layers for the framework focusing on **Ethics, Security, and Explainability**, ensuring these are treated as first-class citizens during discovery.
- **Longitudinal Industrial Studies:** Conduct long-term case studies to evaluate the impact of using DIP-AI on the long-term maintenance and technical debt of AI systems in production.

10.6 Final Remarks

The DIP-AI framework represents a significant step toward a more mature Engineering of AI systems. By empowering Requirements Engineers to act as orchestrators between Business Owners and Data Scientists, this work helps mitigate the high failure rates of AI projects. In summary, DIP-AI provides a robust foundation for reducing resource waste and fostering a culture of "Quality by Design"in the rapidly evolving landscape of Artificial Intelligence.

References

[1] **Iso/iec/ieee international standard - systems and software engineering – software life cycle processes**. *ISO/IEC/IEEE 12207:2017(E) First edition 2017-11*, p. 1–157, 2017.

[2] **Abnt/nbr/iso brazilian standard - risk management – guidelines**. *ABNT/NBR/ISO 31000:2018 Risk Management - Guidelines. 2nd. Geneva: ISO,2018*, p. 1–16, 2018.

[3] **Information technology — artificial intelligence — ai system life cycle processes**. Standard ISO/IEC 5338:2023, International Organization for Standardization, 2023.

[4] **Information technology — artificial intelligence — guidance on risk management**. Standard 23894:2023, International Organization for Standardization, 2023.

[5] AHMAD, K.; ABDELRAZEK, M.; ARORA, C.; BANO, M.; GRUNDY, J. **Requirements engineering for artificial intelligence systems: A systematic mapping study**, 2022.

[6] AHMAD, K.; BANO, M.; ABDELRAZEK, M.; ARORA, C.; GRUNDY, J. **What's up with requirements engineering for artificial intelligence systems?** In: *2021 IEEE 29th International Requirements Engineering Conference (RE)*, p. 1–12, 2021.

[7] AHMED, B.; DANNHAUSER, T.; PHILIP, N. **A lean design thinking methodology (ldtm) for machine learning and modern data projects**. In: *2018 10th Computer Science and Electronic Engineering (CEEC)*, p. 11–14, 2018.

[8] ALVES, A. P.; ARAÚJO, J.; ARAÚJO, M.; ALBERTINI, W.; KALINOWSKI, M. **Uma análise da necessidade de um modelo de referência mps para inteligência artificial**. In: *Anais do XX Workshop Anual do MPS*, p. 16–20, Porto Alegre, RS, Brasil, 2024. SBC.

[9] ALVES, A. P. S.; KALINOWSKI, M.; GIRAY, G.; MENDEZ, D.; LAVESSON, N.; AZEVEDO, K.; VILLAMIZAR, H.; ESCOVEDO, T.; LOPES, H.; BIFFL, S.; MUSIL, J.; FELDERER, M.; WAGNER, S.; BALDASSARRE, T.; GORSCHEK, T. **Status quo and pro-**

blems of requirements engineering for machine learning: Results from an international survey. In: Kadgien, R.; Jedlitschka, A.; Janes, A.; Lenarduzzi, V.; Li, X., editors, *Product-Focused Software Process Improvement*, p. 159–174, Cham, 2024. Springer Nature Switzerland.

[10] AMERSHI, S.; BEGEL, A.; BIRD, C.; DELINE, R.; GALL, H.; KAMAR, E.; NAGAPPAN, N.; NUSHI, B.; ZIMMERMANN, T. **Software engineering for machine learning: A case study**. In: *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, p. 291–300, 2019.

[11] AMPATZOGLOU, A.; BIBI, S.; AVGERIOU, P.; VERBEEK, M.; CHATZIGEORGIOU, A. **Identifying, categorizing and mitigating threats to validity in software engineering secondary studies**. *Information and Software Technology*, 106:201–230, 2019.

[12] ARTASANCHEZ, A. **Why your machine learning project will fail**. http://thedatascience.ninja/2018/07/12/why-your-machine-learning-project-will-fail/.

[13] 2024. https://atlasti.com.

[14] BOURQUE, P.; FAIRLEY, R. E.; SOCIETY, I. C. **Guide to the Software Engineering Body of Knowledge (SWEBOK(R)): Version 3.0**. IEEE Computer Society Press, Washington, DC, USA, 3rd edition, 2014.

[15] CENTRE FOR REVIEWS AND DISSEMINATION. **The database of abstracts of reviews of effects (dare)**. *Effectiveness Matters*, 6(2):1–4, Dec. 2002.

[16] CORBIN, J.; STRAUSS, A. **Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory**. SAGE Publications, Inc., 2008.

[17] COSTAL, D.; FARRÉ, C.; FRANCH, X.; QUER, C. **How tertiary studies perform quality assessment of secondary studies in software engineering**. In: *Proceedings of the XXIV Iberoamerican Conference on Software Engineering (ESELAW@CIbSE)*, p. 1–14, 2021.

[18] COUNCIL, D. **Framework for innovation**, 2024.

[19] CRUZES, D. S.; DYBÅ, T. **Research synthesis in software engineering: A tertiary study**. *Information and Software Technology*, 53(5):440–455, 2011.

[20] DANIELS, N.; GILLEN, P.; CASSON, K.; WILSON, I. **Steer: Factors to consider when designing online focus groups using audiovisual technology in health research**. *International Journal of Qualitative Methods*, 18, Jan. 2019.

[21] DAVIS, F. **A Technology Acceptance Model for Empirically Testing New End-User Information Systems**. 01 1985.

[22] DORARD, L. **Machine learning canvas**, 2015.

[23] DOS SANTOS MARQUES, I. C.; THEISS, L. M.; JOHNSON, C. Y.; MCLIN, E.; RUF, B. A.; VICKERS, S. M.; FOUAD, M. N.; SCARINCI, I. C.; CHU, D. I. **Implementation of virtual focus groups for qualitative data collection in a global pandemic**. *The American Journal of Surgery*, 221(5):918–922, May 2021.

[24] DUNBAR, B. **National aeronautics and space administration (nasa). technology readiness level**. https://esto.nasa.gov/trl/, 2017.

[25] EASTERBROOK, S.; SINGER, J.; STOREY, M.-A.; DAMIAN, D. **Selecting Empirical Methods for Software Engineering Research**, p. 285–311. Springer London, 2008.

[26] FAZAL, S.; DANZI, M. C.; XU, I.; KOBREN, S. N.; SUNYAEV, S.; REUTER, C.; MARWAHA, S.; WHEELER, M.; DOLZHENKO, E.; LUCAS, F.; WUCHTY, S.; TEKIN, M.; ZÜCHNER, S.; AGUIAR-PULIDO, V. **Rexprt: a machine learning tool to predict pathogenicity of tandem repeat loci**. *Genome Biology*, 25(1), Jan. 2024.

[27] FEBIYANTI, H.; YUNIAR, D.; YIHARODIYAH, L.; ANWAR, S. **Implementation of risk management with the iso 31000 approach in an effort to improve the performance of educational services**. *JMKSP (Jurnal Manajemen, Kepemimpinan, dan Supervisi Pendidikan)*, 9:123–135, 01 2024.

[28] GALLEGOS, R. A. P. **Ferramentas de Gestão Voltadas para Melhoria da Qualidade nas Empresas**. Freitas Bastos.

[29] GIRAY, G. **A software engineering perspective on engineering machine learning systems: State of the art and challenges**, 2021.

[30] GLASER, B.; STRAUSS, A. **Discovery of grounded theory: Strategies for qualitative research**. Routledge, 2017.

[31] GLASER, B. G. **Basics of grounded theory analysis: Emergence vs forcing**. Sociology Pr, 1992.

[32] GORSCHEK, T.; GARRE, P.; LARSSON, S. B. M.; WOHLIN, C. **A model for technology transfer in practice**. *IEEE Software*, 23, 2006.

[33] GUSTAFSSON, D. **Analysing the double diamond design process through research & implementation**. 2019. Accessed in: 12 Jul. 2026.

[34] HABIBULLAH, K. M.; GAY, G.; HORKOFF, J. **Non-functional requirements for machine learning: understanding current use and challenges among practitioners**. *Requirements Engineering*, 28(2):283–316, Jan. 2023.

[35] HEHN, J.; UEBERNICKEL, F. **The Use of Design Thinking for Requirements Engineering: An Ongoing Case Study in the Field of Innovative Software-Intensive Systems**. In: *2018 IEEE 26th International Requirements Engineering Conference (RE)*, p. 400–405, 2018.

[36] HEHN, J.; UEBERNICKEL, F.; MENDEZ FERNANDEZ, D. **DT4RE: Design Thinking for Requirements Engineering: A Tutorial on Human-Centered and Structured Requirements Elicitation**. In: *2018 IEEE 26th International Requirements Engineering Conference (RE)*, p. 504–505, 2018.

[37] HEHN, J.; OTHERS. **On Integrating Design Thinking for Human-Centered Requirements Engineering**. *IEEE Software*, 37(2):25–31, 2020.

[38] HERZBERG, G.; PANIKKAR, R.; WHITEMAN, R.; SAHU, A. **The imperatives for automation success**. https://www.mckinsey.com/capabilities/operations/our-insights/the-imperatives-for-automation-success, 2020. Accessed on March 14, 2024.

[39] HEYN, H.-M.; KNAUSS, E.; MUHAMMAD, A. P.; ERIKSSON, O.; LINDER, J.; SUBBIAH, P.; PRADHAN, S. K.; TUNGAL, S. **Requirement engineering challenges for ai-intense systems development**, 2021.

[40] ISHIKAWA, F.; YOSHIOKA, N. **How do engineers perceive difficulties in engineering of machine-learning systems? - questionnaire survey**. In: *2019 IEEE/ACM Joint 7th International Workshop on Conducting Empirical Studies in Industry (CESI) and 6th International Workshop on Software Engineering Research and Industrial Practice (SERIP)*, p. 2–9, 2019.

[41] JORDAN, M. I.; MITCHELL, T. M. **Machine learning: Trends, perspectives, and prospects**. *Science*, 349(6245):255–260, July 2015.

[42] KITCHENHAM, B.; PRETORIUS, R.; BUDGEN, D.; BRERETON, O. P.; TURNER, M.; NIAZI, M.; LINKMAN, S. **Systematic literature reviews in software engineering – a tertiary study**. *Information and Software Technology*, 52(8):792–805, 2010.

[43] KOCIELNIK, R.; AMERSHI, S.; BENNETT, P. N. **Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems**. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–14, New York, NY, USA, 2019. Association for Computing Machinery.

[44] KONTIO, J.; BRAGGE, J.; LEHTOLA, L. **The Focus Group Method as an Empirical Tool in Software Engineering**, p. 93–116. Springer London, 2008.

[45] KUDO, T. N.; BULCÃO-NETO, R. F.; VINCENZI, A. M. **Requirement patterns: a tertiary study and a research agenda**. *IET Software*, 14(1):18–26, 2020.

[46] MARBN, S.; MARISCAL, G.; SEGOVI, J. **A Data Mining amp;amp; Knowledge Discovery Process Model**. I-Tech Education and Publishing, Jan. 2009.

[47] MARTINS, M.; KUDO, T.; BULCÃO-NETO, R. **A qualitative study on requirements engineering practices in an artificial intelligence unit of the brazilian industrial research and innovation company**. In: *Anais do XXVII Congresso Ibero-Americano em Engenharia de Software*, p. 46–60, Porto Alegre, RS, Brasil, 2024. SBC.

[48] MARTINS, M.; KUDO, T.; BULCÃO-NETO, R. **A requirements engineering process for machine learning innovation projects**. In: *Anais do XXVII Workshop de Engenharia de Requisitos*, Buenos Aires, Argentina, 2024. Even3, Brazil.

[49] MITCHELL, T. **Machine Learning**. McGraw-Hill International Editions. McGraw-Hill, 1997.

[50] MORGAN, D. **Focus Groups as Qualitative Research**. SAGE Publications, Inc., 1997.

[51] NAKAGAWA, M. **5w2h–plano de ação para empreendedores**. *São Paulo: Globo, Movimento Empreenda*, 2014.

[52] PALOMARES, C.; QUER, C.; FRANCH, X. **Requirements reuse and requirement patterns: a state of the practice survey**. *Empirical Software Engineering*, 22(6):2719–2762, Dec. 2016.

[53] PEREIRA, L.; OTHERS. **Towards an understanding of benefits and challenges in the use of design thinking in requirements engineering**. p. 1338–1345, 03 2021.

[54] PFLEEGER, S. **Understanding and improving technology transfer in software engineering**. *Journal of Systems and Software*, 47(2):111–124, 1999.

[55] PIATETSKY, G. **Crisp-dm, still the top methodology for analytics, data mining, or data science projects**. https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html, 2014.

[56] PRESSMAN, R. **Engenharia de Software**. AMGH, 2011.

[57] QUAN, J.; SUN, X. **Credit risk assessment using the factorization machine model with feature interactions**. *Humanities and Social Sciences Communications*, 11(1), Feb. 2024.

[58] 2024. https://app-v2.reshape.com.br.

[59] ROOK, P. **Controlling software projects**. *Software Engineering Journal*, 1(1):7, 1986.

[60] RUNESON, P.; HÖST, M. **Guidelines for conducting and reporting case study research in software engineering**. *Empirical Software Engineering*, 14(2):131–164, Dec. 2008.

[61] RUSSELL, S.; NORVIG, P. **Inteligência Artificial: Uma Abordagem Moderna**. 2022.

[62] SALTZ, J. S. **The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness**. In: *2015 IEEE International Conference on Big Data (Big Data)*, p. 2066–2071, 2015.

[63] SCULLEY, D.; HOLT, G.; GOLOVIN, D.; DAVYDOV, E.; PHILLIPS, T.; EBNER, D.; CHAUDHARY, V.; YOUNG, M.; CRESPO, J.-F.; DENNISON, D. **Hidden technical debt in machine learning systems**. In: Cortes, C.; Lawrence, N.; Lee, D.; Sugiyama, M.; Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[64] STOL, K.-J.; RALPH, P.; FITZGERALD, B. **Grounded theory in software engineering research: A critical review and guidelines**. In: *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, p. 120–131, 2016.

[65] TOCKEY, S. **Insanity, hiring, and the software industry**. *Computer*, 48(11):96–101, 2015.

[66] VAN SOLINGEN (REVISION), R.; BASILI (ORIGINAL ARTICLE, 1994 ED.), V.; CALDIERA (ORIGINAL ARTICLE, 1994 ED.), G.; ROMBACH (ORIGINAL ARTICLE, 1994 ED.), H. D. **Goal Question Metric (GQM) Approach**. John Wiley Sons, Ltd, 2002.

[67] VENKATESH, V.; BALA, H. **Technology acceptance model 3 and a research agenda on interventions**. *Decision Sciences*, 39(2):273–315, May 2008.

[68] VILLAMIZAR, H.; KALINOWSKI, M.; LOPES, H. **A catalogue of concerns for specifying machine learning-enabled systems**, 2022.

[69] VILLAMIZAR, H.; KALINOWSKI, M.; LOPES, H. **Towards perspective-based specification of machine learning-enabled systems**, 2022.

[70] VILLAMIZAR, H.; KALINOWSKI, M.; LOPES, H.; MENDEZ, D. **Identifying concerns when specifying machine learning-enabled systems: A perspective-based approach**, 2023.

[71] VILLAMIZAR, H.; KALINOWSKI, M.; VIANA, M.; FERNÁNDEZ, D. M. **A systematic mapping study on security in agile requirements engineering**. In: *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, p. 454–461, 2018.

[72] VOGELSANG, A.; BORG, M. **Requirements engineering for machine learning: Perspectives from data scientists**, 2019.

[73] WAN, Z.; XIA, X.; LO, D.; MURPHY, G. C. **How does machine learning change software development practices?** *IEEE Transactions on Software Engineering*, p. 1–1, 2020.

[74] WILLEMSEN, R. F.; AARDOOM, J. J.; CHAVANNES, N. H.; VERSLUIS, A. **Online synchronous focus group interviews: Practical considerations**. *Qualitative Research*, 23(6):1810–1820, June 2022.

[75] WIRTH, R.; HIPP, J. **Crisp-dm: Towards a standard process model for data mining**. 2000.

[76] WOHLIN, C. **Guidelines for snowballing in systematic literature studies and a replication in software engineering**. In: *18th International Conference on Evaluation and Assessment in Software Engineering, EASE '14, London, England, United Kingdom, May 13-14, 2014*, p. 38:1–38:10, 2014.

[77] YANG, Q.; ZHU, X.; LIU, Y.; HE, Z.; XU, H.; ZHENG, H.; HUANG, Z.; WANG, D.; LIN, X.; GUO, P.; CHEN, H. **Reduced representative methylome profiling of cell-free dna for breast cancer detection**. *Clinical Epigenetics*, 16(1), Feb. 2024.

APPENDIX A

Interview

A.1 Protocol

The protocol presented in this appendix refers to obtaining information from interviews conducted with RDI project coordinators to ascertain the state of practice in requirements engineering for AI and innovation projects.

A.1.1 Objective

General: To obtain the state of the Requirements Engineering practice in RDI projects of the EMBRAPII/CEIA/UFG Unit

Assessment Objectives (GQM):

- Object of study [What will be analyzed?]: the requirements engineering practice in innovation projects at CEIA
- Purpose [Why will the object be analyzed?]: to characterize the state of requirements practice in machine learning innovation projects at CEIA
- Quality focus [Property of the object to be analyzed]: characteristics related to aspects of the process, methodologies, and artifacts related to the requirements engineering phase
- Point of view [who will use the collected data]: coordinators of innovation projects at CEIA (faculty)
- Environment [environment in which the analysis is done]: Center of Excellence in Artificial Intelligence of the Institute of Informatics of the Federal University of Goiás (location of innovation project development) (with the coordination of professors) Development of the Evaluation Plan
- Semi-structured interview, with a deductive approach (funnel), in person or with recording via web conferencing platform for transcription (with an average duration of 1 hour2). The interviews have three dimensions3, defined below for this research:
  1. Temporal dimension: synchronous and interactive

2. Spatial dimension: remote or in person
3. Structural dimension: scripted, classified as a semi-structured interview
4.

A.1.2 Hypothesys

1. Does not implement good traditional or agile elicitation practices in their projects
2. Does not implement good traditional or agile analysis practices in their projects
3. Does not implement good traditional or agile specification practices in their projects
4. Does not implement good traditional or agile validation practices in their projects
5. Does not implement good traditional or agile management practices in their projects
6. Non-functional requirements are neglected in projects
7. Coordinator characteristics affect the adoption of good requirements engineering practices in projects
8. Team characteristics affect the adoption of good requirements engineering practices in projects

1. Implements good traditional or agile elicitation practices in their projects
2. Implements good traditional or agile analysis practices in their projects
3. Implements good traditional or agile specification practices in their projects
4. Implements good practices Traditional or Agile Validation Methods in their Projects
5. Implements good traditional or agile management practices in their projects
6. Non-functional requirements are not neglected in projects
7. Coordinator characteristics do not affect the adoption of good requirements engineering practices in projects
8. Team characteristics do not affect the adoption of good requirements engineering practices in projects

A.1.3 Research Question

1. How do projects implement requirements elicitation?
2. How do projects implement requirements analysis?
3. How do projects implement requirements specification?
4. How do projects implement requirements validation?
5. How do projects implement requirements management?
6. How are non-functional requirements handled in these projects?
7. Do coordinators with more experience or greater knowledge of requirements engineering adopt best requirements engineering practices?
8. Do teams with greater knowledge of requirements engineering adopt best requirements engineering practices?

9. What requirements engineering challenges are faced in these projects?
10. What is generally neglected in the requirements phase in these projects?

A.1.4 Metrics

Q1 - How do projects implement requirements elicitation?

To answer this question, the elicitation questions from the requirements engineering questionnaire will be analyzed.

techniques sources tools

Q2 - How do projects implement requirements analysis?

To answer this question, the analysis questions from the requirements engineering questionnaire will be analyzed.

techniques priorities types of requirements requirements detailing

Q3 - How do projects implement requirements specification?

To answer this question, the specification questions from the requirements engineering questionnaire will be analyzed.

techniques requirement characteristics structure

Q4 - How do projects implement requirements validation?

To answer this question, the validation questions from the ER questionnaire will be analyzed.

checking aspects analyzed documentation validation Q5 - How do projects implement requirements management?

To answer this question, the requirements management questions from the ER questionnaire will be analyzed.

record of Change Management

Q6 - How are non-functional requirements handled in these projects?

To answer this question, the concerns and needs questions from the ER questionnaire will be analyzed.

Functional requirements vs. non-functional requirements Types of non-functional requirements Documentation of non-functional requirements

Q7 - Do coordinators with more experience or greater knowledge of requirements engineering adopt best practices in requirements engineering?

To answer this question, the profile questionnaire questions and the concerns and needs questions from the ER questionnaire will be analyzed.

Q8 - Do teams with greater knowledge in requirements engineering adopt best practices in requirements engineering?

To answer this question, the project questionnaire questions and the concerns and needs questions from the ER questionnaire will be analyzed.

Q9 - What requirements engineering challenges are faced in these projects?

To answer this question, the concerns and needs questions from the RE questionnaire will be analyzed.

Q10 - What has generally been neglected in the requirements phase of these projects?

To answer this question, the concerns and needs questions from the RE questionnaire will be analyzed.

A.2 Methodology

A.2.1 Strategy of Research

Quantitative: The data count regarding the items to be evaluated will be used to obtain evidence for research hypotheses.

Qualitative: The analysis will be carried out using the answers to open-ended questions regarding the practice of requirements engineering.

Desktop realization: Groups of coordinators representing the population of interest will provide retrospective information regarding the projects carried out.

Exploratory research: To gain greater familiarity with the research problem, aiming to build hypotheses.

Strategy to group select: The participants in the case study will be coordinators of RDI projects at the Center of Excellence in Artificial Intelligence.

A semi-structured interview will be conducted to obtain the state of the practice of requirements engineering in RDI projects at the Center of Excellence in Artificial Intelligence.

Documents:

- TCLE
- Coordinator Profile
- Project Profile
- Project Profile
- RE Activities

Pre-Conditions:

- Project approved by the Ethics Committee
- Consent from the CEIA-UFG board regarding the participation of the coordinators in the research
- Participants must have a coordinating role in at least one ongoing or completed project at CEIA.

A.2.2 Planning

Context Selection: The research will be conducted during 2023 with project coordinators of RDI at CEIA. The face-to-face interviews will take place at the Institute of Informatics.

Participant Selection: Potential participants will be identified through collaboration and dissemination of information by the EMBRAPII unit itself. These participants will be recruited by invitation via public email made available on the website of the Institute of Informatics of the Federal University of Goiás.

Participants will be selected non-randomly; all coordinators will be invited to participate in the interview, and all will be subjected to the same script.

Treatment to be performed:

1. The coordinators will be approached by the researchers who will explain the importance of this study as well as its possible contributions.
2. Each coordinator may schedule a time to answer questions regarding each project and their professional trajectory.
3. On the scheduled day and time, the researcher will interview them and record the responses and register acceptance of the Informed Consent Form.
4. Statistical and qualitative analysis of the responses to the profile, project, and requirements questions will be performed.
5. Synthesis and dissemination of the results obtained in the form of reports and/or scientific articles, safeguarding the identities of the participants.

Guidelines:

Interviews, when in person, will be conducted in the city of Goiânia-GO.

As well as filling out the consent forms, profile questions, project and requirements.

Once the CEIA board approves, the coordinators will be contacted to agree to the Informed Consent Form and schedule the interview.

Research location:

Goiânia - Goiás. It will be conducted at the Institute of Informatics - INF of the Federal University of Goiás - UFG.

Criteria for closing the research:

The criterion for closing the research as expected is the collection and analysis of the participants' responses.

The criteria for unexpectedly ending the research are: non-voluntary participation of the coordinators non-release of information from CEIA non-consent from the CEIA board accident involving the principal investigator that prevents the progress of the research

Measurement Instruments:

The measurement instrument will consist of the questions that make up the interview regarding the practice of requirements engineering in RDI projects at CEIA.

A.3 Threats to Validity

Conclusion Validity: Low participation of coordinators, which may reduce the statistical significance of the analyzed data;

Internal Validity: The participant's familiarity with the measurement instrument when coordinating multiple projects may lead to responses at different levels;

The interview script may affect the measurement of the data;

Participants may choose not to respond if they have coordinated several projects;

Construct Validity: Social threat, the researcher may bias the participants' responses with their assumptions and expectations;

External Validity: Time-space conflict of the interview application with the participant's other activities;

A.4 Coordinator Profile

1. Name:
2. Years of practical experience in machine learning projects?
3. Years of experience as a coordinator of machine learning projects?
4. Role at CEIA
5. Experience with software engineering?
6. What is your level of knowledge in software engineering?
7. Experience with requirements engineering?
8. What is your level of knowledge in requirements engineering?
9. Years of experience coordinating projects
10. How many projects have you participated in at CEIA?
11. How many projects have you managed/coordinated at CEIA?
12. How many projects have you maintained at CEIA?
13. Number of projects previously coordinated:

A.5 Project Profile

1. Project Name:
2. Coordinator(s):

3. What is the size of the machine learning component of the product as a whole?
4. What is the TRL level of this project?
5. What type of product was delivered?
6. Start Date
7. Status
8. Team size
9. How was this team organized?
10. What is the application domain of this project?
11. What type of machine learning was used in this project?
12. What learning tasks were addressed in this project?
13. Did your team have someone who was a software engineering specialist?
14. Did your team have someone who was a requirements engineering specialist?

A.6 RE Practice

A.6.1 General

1. •

A.6.2 General

1. Do you apply the activities related to requirements engineering in this project?
    - Elicitation
    - Analysis
    - Specification
    - Validation
    - Management
    - Not applicable
2. What is the justification for not performing requirements engineering in this project?
    - High financial cost
    - Time-consuming
    - Coordinator lacks knowledge
    - No one on the team has the knowledge
    - Other?
3. Was any agile methodology used in the project?
    - Scrum

- Lean
- Kanban
- SMART
- Other
- We did not use it

4. What are the criteria for choosing the agile methodology?
    - Knowledge
    - Experience
    - Ease of use
    - Cost
    - Other
    - We did not use it
5. How do you initially capture the expectations or requirements for a machine learning project?
6. How do you ensure that all team members have the same understanding? Do you use any specification or detail document?
7. How do you communicate changes in expectations?
8. How do you assess the quality of a machine learning project? Quality of the machine learning algorithm, data, software, etc.
9. What are the relevant quality attributes/measures considered in this project?
10. How do you assess the progress of a machine learning project against client expectations?
11. At the end of the project, how do you finally assess whether a machine learning project meets expectations?

A.6.3 Elicitation

1. Which elicitation techniques do you use?
    - interview
    - JAD - Job Analysis and Development
    - brainstorm
    - documentation analysis
    - mind mapping
    - prototyping
    - reverse engineering
    - reuse
    - scenarios
    - questionnaires

- observation
- meeting
- others

2. What are the sources for elicitation?
    - stakeholders
    - users
    - clients
    - documents
    - policies
    - scientific articles
    - manuals
    - books
    - minutes, contracts...
    - legislation
    - similar products
    - prior knowledge of the team

3. Which elicitation and documentation tools did you use?
    - bizagi
    - draw.io
    - figma
    - google forms
    - google docs
    - google sheets
    - microsoft word
    - ML canva
    - miro
    - LucidChart
    - Whimsical
    - Xmind
    - Meet
    - Zoom
    - Teams
    - others

A.6.4 Analysis

1. What analysis techniques are performed?

  - clustering
  - priorization
  - requirements classification
  - modeling
  - others?

2.
3. Which stakeholders are consulted in the analysis phase?
   - data scientist
   - project coordinator
   - client
   - user
   - software engineer
   - other
4. How are requirements priorities defined?
5. Do you distinguish between the types of requirements? If so, which ones?
   - functional requirements
   - non-functional requirements
   - product
   - process
   - data requirements
   - deliverable: time, schdule
   - budget
   - others
6. What types of diagrams are used/generated for requirements modeling?
   - use case
   - flowchart
   - activities/swimsuits
   - entity-relationship diagram
   - state machine
   - classess
   - business process
   - others

A.6.5 Specification

1. Are the requirements documented in the project? Who documents them? Does it follow any documentation standard, for example IEEE 29148/2018?

2. What specification techniques are used?
    - conversation flow
    - user story
    - natural language processing
    - modeling
    - user screens
    - formulas, metrics and mathematical measures
    - use case specification
    - scenarios
    - table
3. What are the attributes of the requirement?
    - uique identifier
    - author record
    - source record for that need
    - registration need
    - classification
4. Is there a concern with data and non-functional requirements? How are they documented?
    - efficiency
    - compatibility
    - usability
    - confidentially
    - security
    - maintainability
    - portability
5. What tool do you use for documenting requirements?
    - jira
    - helix RM
    - trello
    - gitScrum
    - google docs
    - google sheets
    - other

A.6.6 Validation

1. What techniques are used for requirements validation and verification?

- acceptance test scenarios
- inspection
- peer review
- prototyping
- simulation
- step-by-step
- other

2. With which stakeholders are the requirements validated?
    - data scientist
    - project cooordinator
    - client
    - user
    - other
3. What tools are used in the validation?
    - jira
    - helix RM
    - trello
    - gitScrum
    - google docs
    - google sheets
    - other

A.6.7 Management

1. Is there a record of changes (versions) to a requirement? Who is responsible? Where and how is this done?
2. Is it documented (change impact analysis - traceability matrix) what impacts or requirements are affected by the change when a requirement changes?
3. Which tool do you use for requirements management?
    - google sheets
    - jira
    - trello
    - helix RM
    - gitScrum
    - other
4. Are changes to requirement attributes recorded?
    - version

- data of change
- who request the change
- responsible for the record

A.6.8 Concerns and needs

1. Do you think the requirements elicitation for machine learning could have been/be improved in this project? How?
2. Do you think the requirements specification for machine learning could have been/be improved in this project? How?
3. Do you think non-functional requirements for machine learning could have been/be improved in this project? How?
4. Do you think data requirements for machine learning could have been/be improved in this project? How?
5. How important is it to you to have a systematized requirements engineering process for the development of machine learning?
6. What difficulties did you face/are you facing in this project?
    - quality of deliverables
    - delivery deadline
    - team: size, format, organization
    - financial resources
    - infrastructure
    - other
7. What do you think can be done to reduce the difficulties faced in this project?
    - higher quality datasets
    - clearer definition of project objectives
    - better detailing of tasks
    - requirements elicitation and specification
    - requirements validation
    - control of requirements changes
    - improvement in CEIA infrastructure
    - improvement in communication with stakeholders
8. Are or have lessons learned been documented and communicated?

A.7 Results

A.7.1 Profile of Coordinators

Table A.1: *Appendix: Professional Profile and Project Context of the Interviewed Coordinators*

ID	Role at CEIA	SE Knowledge	RE Knowledge & Exp.	Active Projects	Team Size (Avg.)	% AI Staff
P1	Coordinator, Tech Lead	Experienced (2005)	Experienced (2007)	1	17	25%
P2	Coordinator	Highly Exp. (2005)	Highly Exp. (2005)	2	13 / 22	45 / 40%
P3	Coordinator	Knowledgeable	Some Experience	3	10 / 6	20 / 16%
P4	Coordinator, Supervisor	Some Knowledge	Some Experience	5	20	80%
P5	Coordinator, Supervisor	Highly Exp. (2018)	Some Experience (2018)	2	11	45%
P6	Coord., Vice, Supervisor	Experienced (2009)	Experienced (2009)	1	18	80%
P7	Coordinator	Experienced (2013)	Experienced (2013)	1	10	30%
P8	Coordinator, Supervisor	Low Experience	Low Exp. (Graduation)	2	13	60%
P9	Coordinator	No Experience	Low Experience	2	12	50%
P10	Coordinator	Some Knowledge (2006)	Low Experience (2003)	1	14	35%

A.7.2 Techniques and Practices of RE in AI and RDI Projects

Table A.2: *Appendix: Technical AI Characteristics and Requirements Engineering Phases*

ID	Domain	TRL	Learning Type & AI Task	RE Applied Phases	Agile Method
P1	Mental Health	4	Supervised | Prediction	Elicitation, Analysis, Spec., Validation	Scrum (GitScrum)
P2	Health, Chatbot, Process Imp.	6	Unsupervised, Semi | Diag., Pred.	Elicitation, Spec., Val., Management	Scrum, Kanban
P3	Commerce, HR Management	6	Supervised | Prediction	Elicitation, Spec., Validation	Scrum
P4	Health, Law, Energy, Entert.	2-8	Supervised, Unsup., Semi | Diag., Pred.	Elicitation, Spec., Validation	Scrum, Kanban
P5	Public Health	7	Supervised | Image Recognition	Elicitation, Spec., Validation	Scrum
P6	Robotics, Auto. Navigation	4	Supervised | Pred., Image Recog., Nav.	Elicitation, Analysis, Spec., Val.	Kanban
P7	B2B	5	Supervised, Unsupervised | Diag., Pred.	Elicit., Analysis, Spec., Val., Mgmt.	Scrum
P8	Robotics for Health	3	Supervised, Unsupervised | Image, Nav.	Elicitation, Analysis, Spec., Val.	Scrum (Sprint)
P9	Agribusiness, Honest Market	7 & 5	Supervised, Unsupervised, Reinforc. | Pred.	Elicitation, Analysis, Spec., Val.	Scrum
P10	Industrial	7	Supervised | Prediction	Elicitation, Spec., Validation	Scrum

A.7.3 Quality Perceptions, Challenges and Lessons Learned

Table A.3: *Appendix: Quality Management, Challenges, and Lessons Learned in AI Projects*

ID	RE Importance	Progress Evaluation	Expectation Shift	Project Difficulties & Lessons Learned
P1	Highly Important	GitScrum updates	Weekly client meetings	Deliverable quality, needs systematic process.
P2	Highly Important	Planning checklists	Weekly/Monthly meetings	Data gathering, lack of experience. Needs spec guidelines.
P3	Reasonably Imp.	Weekly / Macro-deliveries	Team joins negotiations	Team and infrastructure constraints. Needs mockup tangibility.
P4	Highly Important	Weekly / Macro-deliveries	Weekly meetings	Team and infrastructure. Framework needs agile adaptations.
P5	Important	Not yet formal	Weekly meetings	Difficult initially, relies on methods from other projects.
P6	Highly Important	Weekly reviews	Weekly reviews	Team communication, funding, and lack of data protocols.
P7	Highly Important	Follow-up meetings	Weekly meetings	Financial resources. Higher data quality and clear goals needed.
P8	Important	Macro-delivery steps	Weekly meetings	Time management, team engagement. Formalism helps but needs exploration.
P9	Highly Important	Macro-deliveries	Weekly meetings	Data collection, client expectations, and model refinement.
P10	Reasonably Imp.	Macro-delivery fulfillment	Weekly meetings	Team maturity, industry data access, over-expectations (client expects "magic").

APPENDIX B

Tertiary Study

B.1 Protocol

B.1.1 Research question

1. What is the state of the art of requirements engineering for machine learning systems and what are the most researched topics in requirements engineering for machine learning? - Identify research and proposals regarding the use of requirements engineering for machine learning and artificial intelligence. Identify the main methods, techniques, proposals, uses and when and by whom they occurred, in addition to the supporting tools and roles involved in the use of requirements engineering for machine learning and artificial intelligence.
2. What are the challenges and gaps highlighted by the requirements engineering literature for machine learning and artificial intelligence? - Verify the challenges and possible paths for future research on requirements engineering for machine learning and artificial intelligence.

B.1.2 PICOC Criteria

Population: Secondary studies published about requirements engineering **Intervention:** Methods, tools, techniques, process, approach **Comparison:** Not applicable **Results:** State of art about Requirements Engineering **Context:** Applied to artificial intelligence and machine learning search

B.1.3 Identification of Studies

Search Strategy:

1. automatic search in scientific database;
2. forward snowballing

List of Sources:

1. ACM DL: https://dl.acm.org/
2. Engineering Village: engineeringvillage.com
3. IEEE Xplore: http://ieeexplore.ieee.org/
4. Scopus: https://www.scopus.com/
5. Web of Science: webofscience.com
6. Wiley: https://onlinelibrary.wiley.com

Reasons for choosing of search sources:
Automatic search

- The source of studies must provide a search engine via the Web;
  that allows, at least, a search for the abstract or summary;
  that allows the use of strings with support for Boolean expressions;
  that produces the same results whenever the same string is entered;
  that is capable of exporting search results to popular formats for tools to support systematic studies.
- The source of studies must index publication vehicles that have Computing as an end or means.

Keywords:

- systematic mapping, systematic review, systematic literature mapping, systematic literature review, literature survey
- requirements engineering, requirements elicitation, requirements analysis, requirements specification, requirements validation, requirements management
- machine learning, artificial intelligence, data driven

String:
(("machine learning"OR "artificial intelligence"OR "data-driven") AND ("requirements engineering"OR "requirements elicitation"OR "requirements analysis"OR "requirements specification"OR "requirements validation"OR "requirements management ") AND ("systematic review"OR "systematic literature review"OR "systematic mapping"OR "systematic literature mapping"OR "literature survey"))
Language: English

B.1.4 Selection and Evaluation of Studies

Inclusion and exclusion criteria
IC: The secondary study identifies proposals, use or evaluation of requirements engineering research focused on machine learning or artificial intelligence;

EC1: Full text not available for free on the Web or on the CAPES Periódicos Platform; EC2: Not written in English; EC3: Not a secondary study; EC4: Does not address requirements engineering for machine learning or artificial intelligence; EC5: It is a preliminary or summarized version of another study already included;

Quality criteria

QC1: Are the inclusion criteria properly described? 2 - the criteria are explicit 1 - the criteria are implicit 0 - criteria are not defined or cannot be easily identified QC2: Are the exclusion criteria properly described? 2 - the criteria are explicit 1 - the criteria are implicit 0 - criteria are not defined or cannot be easily identified QC3: Does the search cover all relevant studies? 2 - (Uses four or more relevant databases) or (one search engine AND no extra search strategy) 1 - ((Uses three relevant sources) OR (a search engine)) AND (an extra search strategy) 0 - (uses a maximum of two relevant sources) E (no extra search strategy) QC4: Is the quality of included primary studies assessed? 2 - quality criteria are explicit and associated with each primary study 1 - the secondary study research questions or selection criteria address the quality of the primary study 0 - quality assessment is not carried out or is not made explicit QC5: Are primary studies adequately described? 2 - the details of each primary study are explicit 1 - there is only one summary of each study 0 - there are no results from the primary studies analyzed QC6: Is the justification for the study adequately described? 2 - the justification is duly substantiated and relevant studies are cited as related 1 - the justification is not clear or relevant studies are not cited 0 - justification is not cited and without relevant related studies QC7: Is the protocol validation properly described? 2 - a - strings are derived from the research questions; b - the data to be extracted is related to the research questions and c - the data analysis procedure is appropriate to answer the questions 1 - a, b or c are absent or not treated adequately 0 - validation is not described or cannot be easily inferred QC8: Is the extraction properly described and appropriate? 2 - a - extraction form is explicitly described; b - there is a strategy for extracting and recording data duly described; and c - how the data extraction process is carried out and validated 1 - a, b or c are absent or not treated adequately 0 - validation is not described or cannot be easily inferred

Strategy to select studies

At this stage, the articles obtained are selected by applying the inclusion and exclusion criteria defined in this protocol. Each document will be assigned a status (included, excluded or doubtful) that will be analyzed by the supervisor. The selection will be carried out by three people, one of whom will be called in case of disagreement between the other two.

Evaluation of Quality Studies

The evaluation table with the appropriate criteria, values and justifications is found at the end of this protocol. The evaluation strategy will be the application of the criteria during extraction, that is, in the phase in which the articles are read in full.

B.1.5 Synthesis strategy

Strategy of Extraction data

After fully reading the studies, we seek to identify answers to the research questions and document them in an extraction table. The extraction will be carried out by three people, one of whom is a specialist in machine learning and artificial intelligence and supervised by a fourth person.

Strategy to summarization of results

Using the extraction table, analyzes must be carried out by listing the answers, calculating probabilities or odds ratios and constructing graphs to check the relevance of the results found. The analysis will be carried out by three people, two of them with experience in RE and one in ML and AI.

Strategy of publication

Once the results are all summarized and all readings and syntheses completed, a document will be created that summarizes and documents the results found that answer the questions in the format of a scientific article to be submitted to events and journals.

B.1.6 Extraction Form

- Title
- Authors
- Affiliation
- Publication vehicle
- Year of publication
- Publication type
- Search source
- Number of citations
- How many search sources
- What search strategies
- How many primary studies did you analyze?
- What is the objective of the secondary study Search string used in each secondary study What are the contributions of each secondary study
- What is the state of the art of requirements engineering for machine learning systems and what are the most researched topics in requirements engineering for machine learning? What are the contributions of RE in AI or ML studies (techniques, processes, tools, metrics)
- What are the RE steps explored in AI or ML studies?

Table B.1: *Quality score of the secondary studies.*

Study	QC1		QC2		QC3		QC4		QC5		QC6		QC7		QC8		Score
	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	
S1	2	2	2	2	2	2	2	2	2	1	2	2	1	2	0	0	**2,0**
S3	2	2	2	2	2	1	0	0	1	2	2	2	2	2	2	2	**2,0**
S4	2	2	2	2	1	2	0	0	0	0	2	1	2	2	2	2	**2,0**
S5	2	2	2	2	1	2	0	0	0	0	2	1	2	2	2	2	**2,0**
S9	2	2	2	2	2	2	2	2	0	1	2	2	1	1	1	1	**2,0**
S2	0	1	0	1	2	2	1	2	2	1	2	2	2	2	1	1	**1,5**
S7	1	1	0	0	1	2	1	2	0	1	1	2	0	1	0	0	**1,0**
S8	0	2	0	0	0	1	0	0	0	1	2	1	2	1	1	0	**1,0**
S6	0	0	0	0	0	1	0	0	1	1	1	1	0	0	0	1	**0,5**

- What are the elicitation techniques explored in AI or ML studies?
- What are the analysis techniques explored in AI or ML studies?
- What are the most explored specification techniques in AI or ML studies?
- What validation techniques are explored in AI or ML studies?
- What are the management techniques explored in AI or ML studies?
- What are the tools that support RE in AI or ML studies?
- Do secondary studies point to specific data-driven RE approaches?
- What are the challenges and gaps highlighted by the requirements engineering literature for machine learning and artificial intelligence?
- What is the reported maturity of primary RE studies applied to AI or ML?
- What are the challenges identified in each secondary study?
- Future work highlighted in each secondary study?

B.2 Results

B.2.1 Quality Evaluation

B.2.2 Metadata & RE Core

Identification, quality assessment, RE phases, techniques, modeling languages, and stakeholder and NFR constraints.

Table B.2: *Tertiary Literature Mapping: Studies Identification*

ID	Title
S1	Requirements Engineering for Machine Learning: A Systematic Mapping Study
S2	Requirements engineering for artificial intelligence systems: A systematic mapping study
S3	Whats'up with Requirements Engineering for Artificial Intelligence Systems?
S4	Landscape of Requirements Engineering for Machine Learning-based AI Systems
S5	Multilayered review of safety approaches for machine learning-based systems in the days of AI
S6	Requirements engineering for machine learning: a review and reflection
S7	XAIR: A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development Process
S8	How can we develop explainable systems? insights from a literature review and an interview study
S9	Artificial intelligence project success factors—beyond the ethical principles

Table B.3: *Summary of Extraction Data and Bibliometric Metadata: J for journal, C for conference, A and M mean automatic and manual search, and S for snowballing.*

ID	Country	Year	Venue	Source	Type	Strategy	Citations
S1	Australia	2023	SEAA	Scopus	J	A, S	104
S2	Australia	2021	IST	Scopus	C	A, S	77
S3	Brazil	2021	RE	Scopus	C	A, S	71
S4	Japan	2021	APSECW	Scopus	C	A, S	11
S5	Rep. of Korea	2021	JSS	Scopus	J	A, S	42
S6	Germany	2021	REW	Scopus	J	A	88
S7	USA	2023	MAKE	Scopus	C	A, S	2
S8	Japan	2019	ICSSP/ICGSE	IEEE Xplore	J	A, S	97
S9	Turkey	2021	ITM	ScienceDirect	J	A, S, M	213

B.2.3 AI-Specific Engineering Requirements

Detailed mapping of AI-specific engineering challenges (Data, Models, Ethics, Explainability, Traceability, and Integration).

Table B.4: *Tertiary Literature Mapping: Core RE Phases, Modeling, and Stakeholder/NFR Challenges*

ID	RE Phase	Modeling & Tools	Stakeholder & NFR Challenges
S1	Elicitation, Analysis	Brainstorming	Gap in NFR knowledge and proposals; significant difficulties managing stakeholder expectations.
S2	Elicitation, Analysis	UML, DSM, GORE, SysML, ontoUML, jUCMNav, Sirius	Misalignment and communication gap between Data Scientists and Software Engineers; ethical challenges.
S3	Upstream Discovery	UML, GORE, STL, TSC, CM	Severe lack of integration between Requirements Engineers and Data Scientists; reliability constraints.
S4	Elicitation (46.9%)	i*, KAOS, UML, Safety-case	Core domain understanding bottlenecks; requires explicit tripartite collaboration (Engineers, Scientists, Domain Experts).
S5	Full Cycle, V&V, Safety	Goal/Evidence-Driven Modeling	High cognitive demands on legal and ethics integration; difficulty in defining quantifiable and measurable robustness.
S6	Problem Formalization, Elicitation	Ad-hoc	Complete neglect of stakeholder roles and responsibilities; lack of collaborative domain knowledge integration.
S7	Analysis, Specification	Model/Data Cards, Datasheets, System Cards	Managing technical trade-offs between system usability, algorithmic performance, and stakeholder background knowledge.
S8	Elicitation, Analysis	Quality trade-off analysis	High friction when adapting traditional legacy techniques; necessity to characterize stakeholders based on explicit XAI needs.
S9	Complete Lifecycle	Responsibility Matrix	Broad socio-technical scope; controlling accountability boundaries, parameter configuration, and privacy protections.

APPENDIX C

Planning for Validation

C.1 Protocol

C.1.1 Objetive

The overall objective of the research is to evaluate how much an RE process assists in the development of ML systems in innovation projects.

To achieve the overall objective, we seek the following specific objectives: evaluate effectiveness evaluate efficiency evaluate effectiveness evaluate ease of use evaluate utility evaluate usability evaluate intended use

C.1.2 Research Question

QP1 - Does the process assist in the development of ML systems in innovation projects? QP1.1 - Did the process help the project be carried out by optimizing time and resources? QP1.2 - Did the process help the project achieve its objective? QP1.3 - Did the process help the project be carried out in the best way? QP1.4 - Are the process and artifacts complete and do they cover the project development cycle?

QP2 - Does the process contribute to improving the quality of RE in AI innovation projects? QP2.1 Did the completeness, documentation, and organization of the RE process artifacts assist in the development of AI innovation projects?

QP3 - Are participants interested in accepting and using the process in the development of AI systems in innovation projects? QP3.1 - Is the process easy to use? QP3.2 - Is the process useful? QP3.3 - Is the process usable in practice? QP3.4 - Are you interested in using it in practice?

C.1.3 Hypothesis

H01 - The process does not assist in the development of innovation projects in AI. H02 - The process does not contribute to improving the quality of RE in innovation projects in AI. H03 - The process is not an acceptable technology for the development of innovation projects in AI.

C.1.4 Other Details

Selection of participants: Undergraduate students in Computer Science, Artificial Intelligence, Software Engineering, and Information Systems who have already completed courses in Software Engineering, Requirements Engineering, or Artificial Intelligence and are involved in projects at the EMBRAPII unit.

Questionnaires: Informed Consent Form, Profile, and Perception.

Treatment: 1 - Presentation of concepts

2 - Presentation of the problem

3 - Presentation of the process to be evaluated

4 - Application of the Informed Consent Form and profile questionnaire

5 - Develop a project according to the process

6 - Application of the perception questionnaire

7 - Data analysis

Preconditions:

- Presentation of basic knowledge of Software Engineering and Requirements Engineering and presentation of the process and its artifacts.
- Presentation of a real problem that can be solved with AI.
- Development of the proposal following the process and developing the artifacts.

Measurement Instruments: Questions from the Perception Questionnaire and Artifacts generated in the proposal.

Data Analysis: For open-ended questions, we will use Grounded Theory and quantitative approaches, and for closed-ended questions, we will use quantitative approaches such as: mean, median, and standard deviation from Likert Scale response values.

C.2 Threats to validity:

Conclusion - The authors may try to manipulate some results -> reduced action with strict adherence to this protocol.

Internal - Participants may not be interested in the study and may not collaborate with effective participation -> student motivation, non-mandatory participation, and evaluation by performing the activity. An approach with a certain degree of complexity in the use of the process may generate fatigue, decreasing commitment -> class on patterns and motivating them regarding the topic and evaluation by performing the activity. The instrumentation of the evaluation may be inadequate -> we tried to minimize this threat by seeking examples in the literature.

Construction - The single-operation bias considered as a threat -> was reduced with the application of the study carried out by the professor of the discipline and accompanied by the author and by a professor of Software Engineering and advisor of this research. The threat of inadequate pre-operational explanation for the study -> was reduced with several weekly explanations in class and in groups.

External - Participants in a given group may have prior knowledge -> evaluate participants through the profile questionnaire.

C.3 Participant Profile

Name: Discipline: University:

Dear student, regarding your profile, please answer the following questions:

1. Are you a student in the following program:
   - B.Sc. in Computer Science
   - B.Sc. in Software Engineering
   - B.Sc. in Information Systems
   - B.Sc. in Artificial Intelligence
   - B.Sc. in another program. Which one?
2. Do you already have a degree in another program?
   - No
   - yes. Which one?
3. How many semesters have you completed?
   - up to two semesters
   - two to four semesters
   - four to six semesters
   - eight to ten semesters
   - more than ten semesters
4. Do you currently work in the field of Computing/Informatics?
   - no
   - yes
5. If you answered yes to the previous question, what is your current area?
   - requirements engineer/analyst. For years.
   - business analyst. For years.

- designer. For years.
- software project/product manager. For years.
- programmer. For years.
- tester. For years.
- consultant. For years.
- other. Which? . For years.

6. Have you ever prepared a requirements specification document?
   - Yes, alone.
   - Yes, with assistance.
   - No.
7. What is your current knowledge of requirements engineering?
   - None.
   - I have little knowledge of the subject.
   - I apply it in practice depending on others.
   - I apply it in practice without depending on others.
   - I have considerable knowledge of the subject.
8. Have you participated in CEIA projects?
   - Yes. In projects, in the last semester(s). With the role of .
   - No.

Participant Perception

1. Does the process aid in the development of ML systems in innovation projects?
   - Would the artifact help the project be carried out by optimizing time and resources?
   - Would the artifact help the project achieve its objective?
   - Would the artifact help the project be carried out in the best way?
   - Would the artifact help a person who is participating in identifying a problem in an ML project for the first time?
2. Regarding the EASE OF USE of the requirements standards catalog:
   - My interaction with the artifact is clear and understandable.
   - Interacting with the artifact doesn't require much mental effort from me.
   - I find the artifact easy to use.
   - I find it easy to make the artifact be used for what I want.
3. Regarding the UTILITY of the process:
   - Using the artifact would improve my performance in identifying problems in projects.

- Using the artifact would increase my productivity.
- Using the artifact would increase my assertiveness in identifying problems to meet client expectations.
- I find the system useful for problem identification.

4. Regarding your ABILITY to identify problems USING the ARTIFACT, I could complete the work. . .
   - ...if there was no one around to tell me what to do.
   - ...if I only had one assistance resource or a brief explanation.
   - ...if someone showed me how to do it first.
   - ...if I had already used artifacts similar to this one to identify problems.
5. Regarding your perception of CONTROLLING the artifact to perform the problem identification activity:
   - I have control over the artifact.
   - The artifact provides the necessary resources to identify the problem.
   - Given the resources, opportunities, and knowledge needed to use the artifact, it would be easy to use it.
   - The artifact is... Compatible with other systems and artifacts used in CEIA.
6. Regarding the PLEASURE of using the artifact:
   - I find the artifact pleasant to use
   - The process of using the artifact is enjoyable
   - I have fun using the artifact
7. Regarding the RELEVANCE of the artifact to your work:
   - In my work, the use of the artifact is important
   - In my work, the use of the artifact is relevant
   - The use of the artifact is pertinent to my tasks in the project
8. Regarding the QUALITY of the OUTPUT and DEMONSTRABILITY OF THE RESULT of the artifact for your work:
   - The quality of the artifact's output is high
   - I rate the artifact's results as excellent
   - I would have no difficulty telling others about the results of using the artifact
   - I believe I could communicate to others the consequences of using the artifact
   - The results of using the artifact are apparent to me
   - I would have difficulty explaining why using the artifact may or may not be beneficial
9. Regarding the INTENTION when using the artifact:

- If I have access to the artifact, I intend to use it
- Given that I had I have access to the artifact; I anticipate using it.
- I plan to use the artifact in future projects.

10. In the projects you participated in, what difficulties did you encounter in identifying project problems? Could these be mitigated by using this artifact?
11. What limitations, suggestions, or opportunities for improvement are there for this process?
12. Do you consider that any questions or fields are missing from the artifact?
13. Do you believe that the organization of the artifact is clear and adequate?

C.4 Results

C.4.1 Academic Validation

Profile

Table C.1: *Participants' Academic Background and Professional Roles Summary*

ID	Education Degree	Undergraduate Degree	Works in IT?	Primary Professional Roles Deslipped
1	Master	Comp. Science / Multi-degree	Yes	Requirements, Business, Project Management, Dev, QA, Data Science, Consulting
2	Master	Mathematics	No	Python Developer (Internal Projects)
3	Graduate	Computer Networks Technology	Yes	Consultant / Tech Specialist
4	Master	Environmental Management	No	None
5	Graduate	Software Engineering	Yes	Programmer, Tester, Data Science (Up to 1 year)
6	Graduate	Information Systems	Yes	Business Analyst, Project Manager, Programmer, Tester, Data Scientist, Consultant
7	Graduate	Tourism	Yes	Requirements Analyst, Business Analyst
8	Master	Dentistry	No	Founder, Programmer
9	Graduate	Theology and Design	No	Project Manager
10	Graduate	Computer Science	Yes	Requirements Analyst, Business Analyst, Project Manager, Programmer, Tester
11	Graduate	Dentistry	No	Consultant
12	Graduate	Artificial Intelligence	Yes	Requirements Analyst, Business Analyst, ML Developer, Team Leader
13	Graduate	Electrical Engineering	Yes	Requirements Analyst, Business Analyst, Project Manager, Programmer, Consultant
14	Graduate	Electrical Engineering	No	Project Participant
15	Ongoing Master	Information Systems	Yes	Requirements Analyst, Business Analyst, Designer, Data Scientist, Consultant
16	Master	Electrical Engineering	Yes	Automation Dev, Model Training, Prompt Engineer

Table C.2: *Participants' Experience in Software Engineering, AI, and ML Problem Identification Self-Assessment*

ID	SE Exp. (Years)	AI Exp. (Years)	ML Knowledge Level	Autonomy Level	Experienced in ID Problems?	Skilled in ID Problems?
1	> 10 years	> 10 years	High Knowledge	Supervises others	Strongly Agree	Strongly Agree
2	1 to 3 years	1 to 3 years	None	Needs supervision	Partially Agree	Partially Agree
3	< 1 year	< 1 year	Low Knowledge	Needs supervision	Partially Disagree	Partially Disagree
4	< 1 year	< 1 year	None	Needs supervision	Partially Disagree	Partially Disagree
5	1 to 3 years	< 1 year	Low Knowledge	Works with supervision	Strongly Disagree	Strongly Disagree
6	> 10 years	< 1 year	Knowledgeable	Works without supervision	Partially Disagree	Partially Disagree
7	1 to 3 years	< 1 year	Low Knowledge	Needs supervision	Neutral	Neutral
8	3 to 5 years	1 to 3 years	Low Knowledge	Works without supervision	Strongly Disagree	Strongly Disagree
9	< 1 year	–	None	Needs supervision	Strongly Disagree	Strongly Disagree
10	> 10 years	< 1 year	Low Knowledge	Supervises others	Partially Disagree	Partially Disagree
11	< 1 year	< 1 year	Low Knowledge	Needs supervision	Partially Disagree	Partially Disagree
12	1 to 3 years	3 to 5 years	High Knowledge	Supervises others	Partially Agree	Partially Agree
13	> 10 years	1 to 3 years	Low Knowledge	Needs supervision	Partially Disagree	Partially Agree
14	–	–	Low Knowledge	Needs supervision	–	–
15	1 to 3 years	1 to 3 years	Low Knowledge	Needs supervision	Neutral	Partially Disagree
16	> 10 years	1 to 3 years	Knowledgeable	Supervises others	Neutral	Partially Agree

Evaluation

Table C.3: *Participants' Quantitative Responses Matrix (Likert 1-7 Scale)*

ID	1a	1b	1c	1d	1e	2a	2b	2c	2d	3a	3b	3c	3d	4a	4b	4c	4d	5a	5b	5c	5d	6a	6b	6c	7a	7b	7c	8a	8b	8c	8d	8e	8f	9a	9b	9c
P1	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6
P2	7	7	7	7	7	7	7	7	7	6	5	5	5	4	7	7	7	5	7	7	7	4	4	4	4	4	4	4	5	6	6	6	6	7	7	7
P3	5	6	6	6	6	6	5	7	6	4	3	2	2	4	4	5	5	4	4	4	4	3	3	1	4	4	4	5	6	6	5	3	5	4	4	4
P4	6	6	6	6	6	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
P5	6	7	6	6	7	5	5	6	7	7	6	7	6	7	7	7	7	5	7	7	6	5	4	4	4	4	6	5	4	6	6	6	2	5	5	5
P6	6	6	6	6	6	7	7	7	7	7	6	7	6	5	6	6	7	6	5	5	3	5	5	5	6	4	6	6	6	6	6	6	7	7	7	7
P7	7	7	6	6	7	5	6	6	6	4	3	4	4	5	6	6	6	6	6	7	6	5	5	3	6	6	6	6	5	7	6	5	1	7	7	7
P8	7	7	7	7	7	7	7	7	7	5	5	5	5	7	7	7	7	5	7	7	5	7	7	7	7	7	5	5	6	4	7	7	7	6	6	6
P9	7	6	7	6	7	7	6	7	7	6	5	7	6	6	7	7	6	6	7	7	7	7	7	7	7	7	7	6	6	7	7	7	1	7	7	7
P10	6	7	7	7	7	7	7	7	7	7	6	7	6	5	6	6	7	7	7	7	7	5	5	7	6	6	6	7	7	7	7	7	5	6	6	6
P11	6	6	7	6	7	7	6	7	7	4	4	4	4	3	4	6	7	4	3	6	4	4	4	4	4	4	4	5	5	5	5	5	4	5	5	5
P12	7	7	7	7	7	7	7	7	7	6	7	7	7	6	6	7	7	6	6	6	6	7	3	3	5	5	6	7	7	5	6	6	6	6	6	6

Table C.4: *Synthesized Qualitative Feedback from Participants (Optimized for Page-Width)*

ID	Q10: ML Difficulties & Mitigation	Q11: Limitations & Suggestions	Q12 / Q13: Assessment
P1	Mitigaria, mas cita alternativas mais ágeis.	Tornar artefato mais claro, objetivo e organizado.	No / Needs transparency
P2	Não atua em projetos de software tradicionais.	Nda / Sem observações.	No / Yes
P3	Dificuldade inicial por falta de maturidade técnica.	Nda / Sem observações.	No / Yes
P4	Barreiras no entendimento de requisitos de negócio.	Reduzir número de canvas; focar em prioridades.	Excessive / Yes
P5	Confirmou eficácia na mitigação de problemas.	Sem sugestões de melhoria (Não).	No / Yes
P6	Volatilidade de escopo gera retrabalho no artefato.	Focar em estabilidade de escopo pré-preenchimento.	No / Yes
P7	Escassez de dados e falta de planejamento claro.	Incluir exemplos práticos e estudos de caso reais.	Too early / Yes
P8	Acelera a definição do caminho técnico a seguir.	Migrar estrutura da matriz GUT para formato tabela.	Good start / Yes
P9	Considera o artefato altamente útil para o projeto.	Engenharia é dinâmica; omissões pontuais são normais.	No / Yes
P10	Falta de conhecimento prévio gera barreiras de uso.	Criar guia passo a passo ou manual auxiliar.	No / Yes
P11	Stakeholders focam em sintomas; KPIs mal definidos.	Adicionar viabilidade técnica, infraestrutura e riscos.	Suggests fields / Needs canvas
P12	Coleta de requisitos ambígua e falta de padrões.	Interface pouco intuitiva; criar templates por domínio.	Suggests loops / Needs links

C.4.2 Static Validation

Profile

Table C.5: *Demographic and Academic Profile of Project Coordinators*

ID	Degree	Undergraduate Program	Works in Comp.?	AI Knowledge	Innovation Projects	Autonomy Level
C1	PhD (Doutor)	Computer Engineering	Yes	Advanced	10	Supervises others
C2	PhD (Doutor)	Systems Analysis	Yes	Intermediate	3	Supervises others
C3	PhD (Doutor)	Computer Science	Yes	Basic	8	Supervises others
C4	PhD (Doutor)	Computer Science	Yes	Advanced	10	Supervises others
C5	PhD (Doutor)	Elec., Mech. & Comp. Eng.	Yes	Advanced	6	Supervises others
C6	PhD (Doutor)	Computer Science	Yes	Basic	4	Supervises others
C7	PhD (Doutor)	Computer Science	Yes	Advanced	10	Supervises others
C8	PhD (Doutor)	Computer Science	Yes	Intermediate	5	Works without supervision

Table C.6: *Synthesized Roles and Functions in Innovation Projects*

ID	Q6: Core Professional Background / Expertise	Q11: Roles Performed in Innovation Projects
C1	Req. Eng., Business Analyst, PM, Dev, Data Scientist	Project Coordinator and Team Leader.
C2	Req. Eng., Business Analyst, Project Manager, Tester	Project Coordinator.
C3	Requirements Eng., Project Manager, Consultant	Requirements Analyst, Project Manager, and Coordinator.
C4	Req. Eng., Business Analyst, Project Manager, Consultant	Solution Architect, Engineering Team Lead, Product Manager.
C5	Project Coordinator (CEIA)	Project Coordination and Software Development.
C6	Requirements Eng., Project Manager, Consultant	Project Coordinator and Software Engineering Team Coordinator.
C7	Project Coordinator (CEIA)	General Leader / Project Coordinator.
C8	Req. Eng., Business Analyst, Project Manager	Project Coordinator.

Perception of Use of Framework

Table C.7: *Coordinators' Evaluation of the Framework Across Psychometric Constructs (Likert 1-7)*

	Q1: Auxiliary Potential					Q2: Perceived Usefulness				Q3: Perceived Ease of Use				Q4: Self-Efficacy			
ID	**a**	**b**	**c**	**d**	**e**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**
C1	6	6	6	7	7	4	5	5	6	7	7	7	7	5	6	6	6
C2	5	4	4	5	6	4	2	4	5	7	7	7	7	6	7	7	7
C3	6	6	5	6	6	6	6	6	5	5	5	4	4	5	6	6	6
C4	5	5	6	7	5	6	6	6	7	6	3	6	6	7	3	3	3
C5	7	6	7	7	6	6	6	6	6	7	7	7	7	7	7	7	7
C6	6	5	5	7	6	6	6	6	6	3	3	3	3	3	5	5	6
C7	6	7	7	7	7	5	4	7	7	6	5	5	6	4	5	6	6

	Q5: Perceived Control				Q6: Enjoyment			Q7: Relevance			Q8: Output Quality & Results						Q9: Intention to Use		
ID	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**d**	**e**	**f**	**a**	**b**	**c**
C1	7	6	7	4	5	5	3	6	5	7	7	7	7	7	7	7	6	6	7
C2	6	1	7	7	4	4	4	4	4	4	4	4	5	5	4	3	4	4	4
C3	5	7	6	2	5	5	4	7	7	7	6	6	6	6	6	1	6	6	6
C4	7	6	7	6	5	5	4	6	5	5	6	5	6	6	6	3	5	5	4
C5	7	7	7	5	5	5	5	7	7	7	6	6	6	6	6	6	6	6	6
C6	3	6	6	2	5	5	5	6	6	6	6	5	5	6	6	3	6	6	6
C7	6	7	5	7	5	5	5	7	7	7	7	7	7	7	7	7	6	6	6

Open Responses

Table C.8: *Synthesized Qualitative Feedback and Validations from Coordinators*

ID	Q10: ML Identification Difficulties	Q11: Limitations & Suggestions	Q12: Missing	Q13: Clarity & Layout
C1	Subestimar problemas de ML gera propostas de projetos falíveis.	Mitigação parcial listando soluções viáveis frente aos riscos mapeados.	No	Yes.
C2	Não vivenciou dificuldades prévias na identificação do problema.	Relevante para validar ideias de negócio; menor valor para problemas já bem definidos.	No	Yes, especialmente útil se o problema for ambíguo.
C3	Falta de objetividade do cliente e de processo sistemático reproduzível no CEIA.	Sugere transposição para ferramenta Web colaborativa e melhor visualização das relações dos campos.	No	A organização pode melhorar se integrada a um sistema digital.
C4	*Não informado / Percepção demandaria uso individualizado prévio.*	*Não informado / Percepção demandaria uso individualizado prévio.*	–	–
C5	Não relata dificuldades encontradas nos projetos anteriores.	Sem sugestões de melhorias identificadas no momento.	No	Yes.
C6	Prospecção com profissionais inexperientes e clientes sem maturidade sobre os benefícios da IA.	Sugere desenvolvimento de ferramenta de software para servir de base de conhecimento centralizada.	No	Primeiro canvas possui redundâncias e campos confusos.
C7	Dificuldade crítica em validar a acessibilidade e a qualidade dos dados antes do fechamento da proposta.	Sugere interface interativa guiada por um agente conversacional técnico.	Suggests viability	Yes, mas necessita de exemplos práticos de preenchimento.

C.4.3 Dynamic Validation

Profile

Table C.9: *Dynamic Validation Participants' Profile and Professional Experience*

ID	Academic Degree	Major / Field	Period	SE Experience	ML Experience	Roles Played in Projects	ML Autonomy
P1	Bachelor (In progress)	Production Engineering	> 8th	Up to 1 year	Up to 1 year	Scrum Master, Developer	Supervised
P2	PhD (In progress)	Physics	5th-6th	2 to 4 years	None	Developer, ML Modeler	Supervised
P3	PhD (In progress)	Physics	–	> 4 years	Up to 1 year	Data Scientist, Scrum Master, PO	Unsupervised
P4	Bachelor (In progress)	Computer Engineering	> 8th	2 to 4 years	Up to 1 year	Developer, PO	Unsupervised
P5	Bachelor (In progress)	Computer Engineering	> 8th	Up to 1 year	Up to 1 year	Developer	Unsupervised
P6	Bachelor (In progress)	Computer Engineering	7th-8th	2 to 4 years	2 to 4 years	Developer, Data Scientist	Unsupervised
P7	Bachelor (In progress)	Computer Engineering	> 8th	> 4 years	Up to 1 year	Developer	Unsupervised
P8	Bachelor (In progress)	Computer Science	> 8th	> 4 years	1 to 2 years	Developer, Data Scientist	Unsupervised
P9	Bachelor (In progress)	Computer Engineering	3rd-4th	Up to 1 year	None	Product Ideation, Developer	Unsupervised

Perception of Use Framework

Table C.10: *Coordinators' Framework Perception and Evaluation Metrics (Likert 1-7)*

	Q1				Q2				Q3				Q4				Q5				Q6				Q7				Q8				Q9			
ID	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**	**a**	**b**	**c**	**d**
C1	3	4	4	4	4	5	5	4	5	4	4	5	5	5	4	5	5	4	5	4	4	5	5	5	5	5	6	4	5	4	5	4	4	4	5	5
C2	7	6	4	6	7	3	4	7	7	3	2	6	7	6	7	7	4	3	7	6	6	5	5	3	6	7	7	6	5	7	7	7	3	6	5	6
C3	5	6	6	6	7	5	5	6	6	6	6	6	6	4	5	6	6	6	6	6	6	6	6	3	5	6	6	5	5	6	6	5	2	4	4	4
C4	7	7	4	7	3	7	5	7	6	7	7	7	7	5	6	7	7	3	7	7	4	7	7	4	5	7	7	6	7	7	6	7	1	6	7	6
C5	7	6	6	7	6	6	6	6	6	6	3	6	6	5	5	5	5	4	6	5	4	4	4	4	6	6	6	5	5	5	5	5	5	5	5	5
C6	7	7	7	7	5	7	7	7	7	6	6	6	6	5	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	1	7	7	7
C7	7	7	7	7	6	7	6	7	7	5	3	6	6	5	6	7	7	5	6	7	7	6	7	7	7	7	7	7	6	6	5	6	2	6	6	6
C8	6	6	5	6	2	4	3	5	7	3	3	4	2	3	4	6	7	2	6	7	4	3	4	3	7	7	6	5	4	5	6	7	2	4	5	3
C9	5	5	5	5	6	5	5	5	5	4	6	4	4	2	3	6	5	3	4	6	6	4	4	4	4	6	6	4	4	5	5	5	3	4	4	4
C10	6	7	7	7	6	6	6	7	7	6	5	5	6	5	6	7	7	6	7	6	7	5	5	4	6	7	6	7	6	6	6	6	2	5	5	6

Open Responses

Table C.11: *Qualitative Feedback from Dynamic Validation: Difficulties, Mitigations, and Suggestions*

ID	Project Difficulties in ML Task Identification	Suggestions and Framework Criticisms
P1	Dificuldade em entender o ambiente do cliente. O artefato auxilia a organizar ideias, mas não resolve o entendimento por completo.	Nenhuma sugestão apresentada.
P2	Não relatou dificuldades específicas no projeto.	Sugeriu tornar o framework mais enxuto, reduzindo e unindo caixas/campos muito semelhantes entre si.
P3	Falta de espaço inicial para discutir particularidades de ML (dados e modelos). O artefato mitigaria isso se usado no início.	Não apresentou sugestões adicionais.
P4	Falta de clareza na comunicação com o cliente e nos processos. O artefato mitigaria ao guiar a priorização de dados.	Sentiu sobreposição entre alguns campos (ex: Matriz GUT). Sugeriu uma versão Web/App com UX minimalista.
P5	Confirmou a existência de dificuldades e que o artefato as facilitou.	Não apresentou sugestões adicionais.
P6	Não relatou dificuldades específicas no projeto.	Sugeriu adicionar campo para caracterizar o tipo de problema (classificação, regressão, supervisionado, LLMs, etc.). Justificou discordância na adequação para iniciantes e stakeholders.
P7	Falta de foco da equipe ao avançar nas Sprints. O artefato mitigaria o desvio ao ser preenchido em rodadas cíclicas.	Reflexão sobre possível excesso de campos, o que pode causar exaustão e travar o processo criativo. Sugeriu uso como o Kanban.
P8	Falta de clareza nos processos relacionados ao problema e necessidade de estruturar os dados necessários de forma clara.	Sugeriu a criação de um glossário para termos do artefato (como "AM"e "Tendência"da GUT) para evitar confusões.
P9	Problemas bem definidos no projeto atual, mas vê grande utilidade na modelagem de viabilidade e para futuros cenários ambíguos.	Achou mais prático utilizar o artefato de forma virtual.

APPENDIX D

Validation in Focus Group

D.1 Protocol

D.1.1 Objective

The overall objective of the research is to evaluate how much an ER process assists in the development of AI systems in innovation projects.

To achieve the overall objective, we seek the following specific objectives: evaluate ease of use, evaluate utility, and evaluate intended use. We also aim to identify positive aspects and areas for improvement in both the framework and the tool (with LLM support).

D.1.2 Research Question

1. QP1 - Does the process assist in the development of ML systems in innovation projects?

   QP1.1 - Did the process help the project be carried out by optimizing time and resources?

   QP1.2 - Did the process help the project achieve its objective?

   QP1.3 - Did the process help the project be carried out in the best way?

   QP1.4 - Are the process and artifacts complete and do they cover the project development cycle?
2. QP2 - Does the process contribute to improving the quality of ER in AI innovation projects?

   QP2.1 Did the completeness, documentation, and organization of the ER process artifacts assist in the development of AI innovation projects?
3. QP3 - Are participants interested in accepting and using the process in the development of AI systems in innovation projects?

   QP3.1 - Is the process easy to use?

   QP3.2 - Is the process useful?

QP3.3 - Is the process usable in practice
QP3.4 - Are you interested in using it in practice?

D.1.3 Hypothesis

H01 - The framework does not assist in the development of innovation projects in AI. H02 - The framework does not contribute to improving the quality of RE in innovation projects in AI. H03 - The framework is not an acceptable technology for the development of innovation projects in AI. H04 - The tool does not helps in the use of framework use and acceptance.

D.1.4 Other Details

Selection of participants: Specialists, coordinators and tech leads in AI development.

Questionnaires: Informed Consent Form, Profile, and Perception.

Treatment:

1. Presentation of concepts
2. Presentation of the problem
3. Presentation of the framework to be evaluated
4. Presentation of the tool to be evaluated
5. It discusses the concept of discovery and understanding the problem with use of framework
6. It discusses how much the tool and use of LLM helps in discovery.
7. Application of the perception questionnaire
8. Data analysis

Preconditions:

- Presentation of basic knowledge of Software Engineering and Requirements Engineering and presentation of the process and its artifacts.
- Presentation of a real problem that can be solved with AI.
- Development of the proposal following the process and developing the artifacts.

Measurement Instruments: Questions from the Perception Questionnaire and Artifacts generated in the proposal.

Data Analysis: For open-ended questions, we will use Grounded Theory and quantitative approaches, and for closed-ended questions, we will use quantitative approaches such as: mean, median, and standard deviation from Likert Scale response values.

D.2 Participant Profile

1. What is your educational level?
2. What role have you held in the market?
    - data scientist
    - AI/ML engineer
    - project manager
    - product manager/owner
    - researcher
    - professor
    - software developer
    - other
3. Indicate the sector of activity and the size of the organization in that sector
    - public
    - private
    - startup
    - consulting
4. What types of innovation projects have you participated in?
    - R&D and research applied
    - digital products development
    - AI products
    - open innovation or partnership
    - prototyping or MVP
5. What is your profile when participating in these projects?
    - leadership and management
    - technical development
    - research
    - consultant
    - software development
    - support
6. What AI technologies do you have experience with?
    - machine learning
    - LLMs
    - computer vision
    - recommendation systems
    - ai agents

7. What is your experience or familiarity with requirements in AI projects?
    - project discovery
    - solution ideation
    - requirements specification
    - prototyping
    - validation and testing

D.3 Perceptions in Focus Group

We asked the experts to complete a basic version of a TAM (Technology Assessment Module) in order to obtain greater commitment to data collection. We asked questions regarding ease of use, usefulness, and intended future use of both the framework and the tool.

Each of the questions could be answered with a value from 1 to 7 on the Likert scale for agreement with: strongly disagree, somewhat disagree, somewhat disagree, neutral, somewhat agree, somewhat agree, and strongly agree.

1. The framework is easy to use.
2. The tool is easy to use.
3. The framework is useful for supporting the proposed activities.
4. The tool is useful for supporting the proposed activities.
5. I intend to use the framework in the future.
6. I intend to use the tool in the future.

D.4 Analysis of Data

In addition to the questionnaires, we recorded the focus group dynamics and performed the analysis based on the transcript. We also analyzed the transcript using thematic coding, following an open and axial coding approach. Then we grouped the codes and identified positive perceptions and areas for improvement for both the framework and the tool, including the use of LLM.

D.5 Results

D.5.1 Profile

Note: The IDs were standardized according to the groups (FG1P1 to FG2P5). Where FG1 is Focus Group 1 and FG2 is Focus Group 2. Same P1 is for participant 1 and so on.

Table D.1: *Participants' Academic Background, Professional Roles, and Organizational Sectors*

ID	Academic Degree	Core Professional Roles (Experience)	Sector of Activity and Organization Size
FG1P1	PhD (Completed)	Data Scientist (3–5y); AI/ML Engineer (3–5y); Data Engineer (1–3y); Researcher (3–5y); Professor (5–10y)	Private (Up to 100 employees)
FG1P2	PhD (Completed)	Data Scientist (3–5y); AI/ML Engineer (1–3y); Researcher (>10y); Professor (3–5y); Software Developer (3–5y)	Public (<200); Private (<100); Startup (<50); Consulting (<100)
FG1P3	PhD (In progress)	Data Scientist (1–3y); AI/ML Engineer (3–5y); Data Engineer (<1y); Researcher (5–10y); Professor (<1y); Software Developer (5–10y)	Consulting (More than 200 employees)
FG2P1	MSc (Completed)	Data Scientist (5–10y); AI/ML Engineer (1–3y); Data Engineer (<1y); Project Manager (<1y); Product Manager (<1y); Researcher (>10y); Professor (5–10y); Software Developer (>10y)	Startup (> 200); Consulting (> 200)
FG2P2	PhD (In progress)	Data Scientist (3–5y); AI/ML Engineer (3–5y); Researcher (3–5y); Professor (5–10y); Software Developer (5–10y)	Public (<50); Private (> 200); Startup (<100); Consulting (> 200)
FG2P3	PhD (In progress)	Researcher (3–5y)	Public (Up to 50 employees)
FG2P4	MSc (Completed)	Data Scientist (5–10y); AI/ML Engineer (5–10y); Data Engineer (5–10y); Project Manager (<1y); Product Manager (1–3y); Researcher (<1y); Software Developer (>10y)	Consulting (More than 200 employees)
FG2P5	PhD (Completed)	Data Scientist (>10y); AI/ML Engineer (5–10y); Researcher (5–10y); Professor (<1y); Software Developer (5–10y)	Public (> 200); Private (> 200); Startup (<200); Consulting (> 200)

Table D.2: *Participants' Experience with AI Technologies and Requirements Engineering Core Stages*

		Familiarity with Requirements in AI Projects				
ID	**Experience with AI Technologies**	**DIS**	**IDE**	**SPE**	**PRO**	**VAL**
FG1P1	**High:** ML, LLMs **Moderate:** Deep Learning, Computer Vision, AI Agents **Low:** Recommendation Systems	Low	High	High	Moderate	Moderate
FG1P2	**High:** ML, Deep Learning, AI Agents **Moderate:** LLMs, Computer Vision, Recommendation	Moderate	High	Low	High	High
FG1P3	**High:** ML, Deep Learning, LLMs, Computer Vision, Recommendation Systems, AI Agents	Low	Moderate	High	Moderate	Moderate
FG2P1	**High:** ML, Deep Learning, Recommendation, AI Agents **Moderate:** LLMs, Computer Vision	Moderate	High	Moderate	Moderate	High
FG2P2	**High:** ML, Deep Learning, LLMs, Recommendation, AI Agents **Low:** Computer Vision	High	High	High	Moderate	Moderate
FG2P3	**High:** Computer Vision, Recommendation, AI Agents **Moderate:** LLMs **Low:** ML, Deep Learning	High	High	High	Moderate	Low
FG2P4	**High:** ML, Deep Learning, Vision, Recommendation, AI Agents **Moderate:** LLMs	Moderate	High	Moderate	High	High
FG2P5	**High:** ML, Recommendation, AI Agents, Vision **Moderate:** Deep Learning, LLMs **Low:** Computer Vision	High	High	High	High	High

note: For the sake of conciseness, the competencies have been abbreviated in the Requirements matrix (DIS: Discovery; IDE: Ideation; SPE: Specification; PRO: Prototyping; VAL: Validation and Testing).

D.5.2 Perception of Use

Focus Group 1

Table D.3: *Synthesis of participant contributions - Focus Group 1*

Participant	Indications regarding the Framework / Canvas	Indications regarding the Tool / LLM Usage
FG1P3	* Validation of the initial focus on common corporate business processes. * Early mapping of sensitive data compliance within the *Business Owner* perspective rather than solely in the final risks section. * Proposal for a simple intermediate validation mechanism (a two-way *handshake* checklist) between steps to align business and data science views.	* Critical risk of the LLM "steering" or biasing the project by auto-filling the entire Canvas based on a single initial response. * Suggestion to block full automatic generation until a minimum percentage (e.g., 50% or 60%) of basic fields is manually filled. * Requirement for visual traceability of sources, indicating which previous cards the LLM used as context.
FG1P2	* Warning regarding industrial secrets and business confidentiality during discovery dynamics to protect client privacy. * Mitigation of technical deficiencies and asymmetries on the client side, as discovery success depends heavily on client maturity. * Recommendation to include a formal discussion on operationalization and technology transfer.	* Inquiry regarding dynamic reprocessing (whether modifying a card triggers recalculations for subsequent cards). * Proposal to allow users to configure and run a local *foundation model* within their own controlled infrastructure to address privacy concerns.
FG1P1	* Warning about the critical importance of compliance and regulatory frameworks (e.g., AI Act in Europe, LGPD in Brazil), keeping them visible outside of general risks. * Expansion of the *baseline* concept to encompass human resources/staffing needs and cloud infrastructure costs.	* Positive evaluation of using AI to mitigate user fatigue caused by manually filling long forms. * Recommendation to document model accuracy, correctness, and hallucination tendencies within the "Threats to Validity" section of the doctoral thesis.

Focus Group 2

Table D.4: *Synthesis of participant contributions - Focus Group 2*

Participant	Indications regarding the Framework / Canvas	Indications regarding the Tool / LLM Usage
FG2P1	* Validation of environmental limitations (infrastructure, cloud vs. on-premises, and financial constraints). * Inclusion of ROI due to the high cost of AI projects. * Stratification of the Canvas view by role/hierarchical level (e.g., data engineer view vs. leadership view). * Use of *Data Mesh* as a high-level architectural reference for data organization.	* Development of a knowledge base using data from previously completed projects (*meta-learning* approach) to improve LLM suggestion accuracy.
FG2P4	* Assessment of discovery methodologies from major consulting firms to combine with the Canvas, avoiding endless questionnaires. * Adoption of the Double Diamond (*Design Thinking*) approach to explore, narrow down, and re-explore before the final step.	* Use of operational KPIs (represented in a radar chart format) to provide clarity to executive boards regardless of their technical background. * Definition of critical manual checkpoints by the user to validate and guide model suggestions, preventing error propagation.
FG2P2	* Mapping of initial explainability requirements directly within the business perspective. * Inclusion of a “Go/No-Go” metric or criterion for the model (e.g., defining tolerable hallucination limits in critical scenarios).	* Validation of the LLM as an excellent resource for initial pre-filling, helping professionals extract a “critical mass” from clients who struggle to provide initial inputs.
FG2P5	* Praise for the inclusion of KPIs as an element to ground corporate expectations. * Concern regarding basic inputs: mapping the core infrastructure required for the AI system to operate. * Identification of the commercial challenge of fitting this discovery phase into the client’s work process.	* Requirement for automatic generation of project scope artifacts or *project charters* (similar to the TDSP method) to deliver real value. * Use of the LLM or autonomous agents to automatically critique internal inconsistencies within the Canvas.
FG2P3	* Assessment of the framework as dynamic and flexible, despite being structured in blocks. * Concern regarding the team’s learning curve: potential difficulties in getting the entire team to understand how each pillar supports daily activities.	* Strict necessity of embedding technological *guardrails* into the LLM workflow to mitigate hallucinations regarding exact framework data. * Praise for the user interface visualization filters.