A Qualitative Study on Requirements Engineering Practices in an Artificial Intelligence Unit of the Brazilian Industrial Research and Innovation Company

Mariana Crisostomo Martins, Taciana Novo Kudo, Renato F. Bulcão-Neto

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

Abstract. *In recent years, there has been a focus shift from software development in general to the construction and training of machine learning (ML) models integrated into a software product. This movement has raised challenges in ML systems' requirements engineering (RE) theory and practice. This paper investigates RE practices in ML systems research, development, and innovation projects carried out by an Artificial Intelligence (AI) Unit of the Brazilian Industrial Research and Innovation Company. Our methodology includes semi-structured interviews with leaders of 21 projects and data analysis through the grounded theory method. We identified the predominance of RE methods, techniques, and tools applied ad hoc and uncoordinatedly. This result corroborates the literature reports on RE for ML systems, especially those involving innovation projects.*

1. Introduction

Recent research has demonstrated a shift in focus from artificial intelligence (AI) algorithms supporting requirements engineering (RE) activities to RE approaches for AI-based software systems [Villamizar et al. 2021]. However, the literature has also pointed out the complexity of the RE process for these systems. For instance, RE professionals convey unrealistic expectations in AI-based software projects [Alves et al. 2024], it is not clear which RE techniques and tools have been used in practice, and whether well-established RE methods should or could be adapted [Villamizar et al. 2021].

In this paper, we report our findings about RE practices in research, development, and innovation projects carried out by an Artificial Intelligence (AI) Unit of the Brazilian Industrial Research and Innovation Company, called CEIA (Center of Excellence in Artificial Intelligence)[1]. We planned and conducted semi-structured interviews with leaders of 21 projects implementing AI algorithms for several application domains, such as health, entertainment, marketing, and the automobile industry. We performed a qualitative analysis based on Grounded Theory [Stol et al. 2016] and the method proposed by [Glaser and Strauss 2017].

As a result, we conclude that in this particular scenario, with such heterogeneous AI projects, there is a lack of a methodology for RE activities. The majority of coordinators who participated in the research reported that they do not possess any RE specialists on their team, and acknowledge the significance of the activity and the necessity of implementing a systematic process for it.

[1]https://ceia.ufg.br/

This paper is organized as follows: Section 2 outlines related work; Sections 3 and 4 describe our method and present our results, respectively; Section 5 discusses our results; Section 6 summarizes our study's threats to validity; and Section 7 brings our conclusions and future work.

2. Related Work

Interviews with 278 software engineering (SE) professionals with machine learning (ML) knowledge were carried out in [Ishikawa and Yoshioka 2019]. The goal is to verify the current state of ML applications from the perspective of SE activities and identify how ML is perceived by the SE community. The authors sought to understand respondents' experience in SE and ML, identify projects they had previously worked on, and identify perceived obstacles and distinctive features of ML. Among the challenges highlighted are understanding expectations, unrealistic customer expectations, the uncertain nature of the domain, the impossibility of guarantees, and the need for continuous engineering.

In [Silva and Canedo 2022], ER challenges and techniques for chatbot development were identified. The survey obtained 22 responses from Brazilian professionals in the chatbot industry. The authors investigated the nature of the respondents' organizations, the roles they assumed in the projects, the methodologies used for development, the elicitation and documentation techniques, and the tools used for elicitation. Most respondents work for private companies, practice Scrum, elicit requirements through brainstorming, and document using conversational flow. Most of the respondents believe that NFR and the ER process are neglected.

In recent work [Habibullah et al. 2023], the authors interviewed ten engineers specializing in functional and non-functional requirements to gain knowledge about the current state of the art and challenges associated with RE for ML. In addition to demographic factors, such as the professional profile of the interviewees, the authors also identified how non-functional requirements were addressed and evaluated in the projects in which these professionals participated. The authors highlighted among their findings that domain uncertainty and data selection are the central challenges.

Vogelsang and Borg also interviewed data scientists to get RE practice in ML development [Vogelsang and Borg 2019]. The scientists pointed out that they typically use interviews for elicitation. They highlighted the importance of having a legal specialist involved in obtaining needs. They also highlighted the importance of handling data and performance requirements and emphasized the importance of monitoring and the possibility of a checklist.

We advocate that our work complements those mentioned above since we aim to discover how RE practices have been carried out in 21 AI-based research, development, and innovation projects by interviewing project coordinators from different domains.

3. Method

3.1. Interviews

Our main goal is to obtain the state of RE practice in ML-based RDI projects in the CEIA Unit. We used the GQM (Goal-Question-Metric) goal definition template to define the following research objectives:

- **Object of study.** [What will be analyzed?]: the RE practice in CEIA innovation projects.
- **Purpose.** [Why will the object be analyzed?]: to characterize the state of RE practice in CEIA AI innovation projects.
- **Quality focus.** [Property of the object that will be analyzed]: the process, methodologies, and artifacts characteristics related to the RE phase.
- **Point of view.** [Who will use the collected data?]: CEIA innovation projects leaders (researchers).
- **Environment.** [What is the context in which the analysis is carried out?]: an Artificial Intelligence (AI) Unit of the Brazilian Industrial Research and Innovation Company at the Federal University of Goiás (CEIA-EMBRAPII-UFG), Brazil.

3.1.1. Hypotheses

- H01 - It does not carry out good traditional or agile elicitation practices in its projects.
- H02 - It does not implement good traditional or agile analysis practices.
- H03 - It does not implement good traditional or agile specification practices.
- H04 - It does not carry out good traditional or agile validation practices.
- H05 - It does not implement good traditional or agile management practices.
- H06 - Non-functional requirements are neglected.
- H07 - Coordinator's profile affects the adoption of good RE practices.
- H08 - Team's profile affects the adoption of good RE practices.

3.1.2. Research Questions

1. How are RE phases performed in AI projects?
2. How are non-functional requirements handled in AI projects?
3. Are better RE practices in projects achieved because of their leaders with more experience or greater knowledge of RE?
4. Are better RE practices in projects achieved because of their teams with more experience or greater knowledge of RE?
5. What RE challenges are faced in these projects?
6. What has generally been neglected from the RE phase in these projects?

The research questions guided the questions in the interview guide available in the repository of this article.

3.1.3. Characteristics

We conducted semi-structured interviews, in person or with recording via web conferencing for transcription (1h, on average). Interviews have three dimensions:

1. Temporal dimension: synchronous and interactive;
2. Spatial dimension: remote or in-person;
3. Structural dimension: scripted, classified as a semi-structured interview.

In this protocol, the interview candidates are RDI project coordinators. Initially, any coordinator could be invited, later we gave preference to people who had more than one project under their coordination. This is an exploratory study about the state of RE practice in AI/ML projects carried out at CEIA located at the Informatics Institute (INF) of the Federal University of Goiás (UFG).

This work was submitted to the Ethics Committee on 02/28/23 and approved under CAAE: *67531823.4.0000.5083*.

3.1.4. Data Collection

Coordinators of 21 AI systems projects were invited to answer the interview. More details about the interview setup are available at the survey link on Zenodo[2]. The research context and the application of the Free and Informed Consent Form (ICF) were explained during the interview. The questions organized in a script were asked based on the participant's answers. An interview was then carried out with authorized recording. The script was divided into 9 profile questions, 15 questions about the project, and 41 remaining about RE practice in these projects.

3.2. Data Analysis

Data analysis was carried out based on the Grounded Theory (GT) method [Glaser and Strauss 2017] that aims to generate theories rather than test or validate the existing theory. Widely used in qualitative data analysis and more recently widely used in software engineering research [Stol et al. 2016].

We follow the Straussian approach for GT [Glaser 1992]. In this approach, the research question is open and defined based on the literature. Literature is consulted throughout the process to understand concepts and information that may be useful throughout data analysis. The coding process is determined by the open, axial, and selective phases, which generate categories and relationships. Finally, the results are analyzed.

Open coding begins with detailed reading and searching for important information in documents. Subsequently, fragments are associated with quotations and codes are assigned that represent a certain phenomenon. For axial coding, the aim is to find categories that bring together several grouped and related codes. Finally, in selective coding, the central category is generated that represents the main finding of the research. This procedure is performed iteratively until theoretical saturation is reached [Corbin and Strauss 2014]. However, we did not reach theoretical saturation because we thought we would need more interviews and investments for different scenarios.

The interview process started in June 2023 and finished in December 2023. Some interviews needed to be divided into two parts due to participants' time availability. While conducting interviews, we realized transcribing and data coding. 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* tool, which automatically transcribes and edits texts with transcription problems. We downloaded the documents and uploaded them to the *Atlas.TI* tool.

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

For each document referring to an interview, we carried out coding. The document was read in its entirety, marking items that referred to a specific question in the answers. The idea was to identify similarities between the coordinators to group the coded items and establish what the RE practice was like in these projects, testing the hypotheses defined for this research. open, axial, and selective coding procedures were carried out by the first author. The second and third authors reviewed the codings weekly and discussed partial results.

The discussion helped to avoid inconsistencies, seeking completeness of responses by checking code nomenclature and relationships. For the 10 interviews, 838 citations were generated from which 249 codes and 39 groups of codes emerged, synthesized into 4 networks that address different aspects of the interviews, such as coordinator, project, practice requirements, and RE to AI.

4. Results

We conducted the interviews with ten coordinators and then transcribed them using the *Reshape* tool. A total of 555 minutes were written down. The interviews in the media lasted slightly less than 60 minutes.

4.1. Characterization

4.1.1. Project

Concerning 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 varied as well; on some teams, only 16% of the team worked on AI-related tasks, while on others, 100% of the team was directly involved with the AI solution. It is imperative to emphasize that teams exhibit volatility throughout a project, contingent upon the requirements of individual teams, project phases, and participant counts.

Technology Readiness Levels (TRL) represent a type of measurement system to determine the degree of maturity (from 1 to 9) of a specific technology [Dunbar 2017]. TRL levels for the projects consulted ranged from 3 to 7, with level 6 being the most typical. Technical coordinators, who oversaw smaller teams tasked with completing particular project tasks, assisted the coordinators under interview. Each project had one to three technical coordinators. About the agile approach used in research, most of the coordinators said they employed adapted Scrum and Kanban and a combination of agile approaches in their projects. No coordinator acknowledged using non-agile methods. While most of the project domains are in health, there is also AI research in business, industrial management, process improvement in organizations, law, and agribusiness.

Most of the projects use supervised and unsupervised learning approaches. We also find examples of reinforcement learning and semi-supervised learning. Prediction is the most common ML task in those projects. But, we also found other tasks such as recognition, diagnosis, regression, signal processing, robot navigation, prescription, and recommendation. There are not many participants in most projects who are requirements engineering experts. Most of the RE and SE experts are project coordinators or technical coordinators. Project information is in Table 1. In the task column, we represent acronyms

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

Tabela 1. Project characterization.

for prediction (P), diagnosis (D), recommendation (R), image recognition (IR), robot navigation (RN), and sentiment analysis (SA). Similarly, in the learning type column, with supervised (S), unsupervised (U), reinforcement (R), and semi-supervised (SS).

4.1.2. Coordinator

The majority of coordinators work on one or two projects. Half of the coordinators have prior experience working on AI projects in the industry with two to eighteen years of experience. Three coordinators reported experience of one to thirteen years of coordination in previous AI projects. Information regarding the profile of each coordinator can be seen in Table 2: the coordinator's identifier (ID), SE and RE knowledge, the number of current projects (CP), AI project development time (in years), industry experience in AI (in years), and SE experience, respectively.

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	alguma experiência	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

Tabela 2. Profile of the coordinators interviewed.

4.2. RE Practice

All coordinators carried out requirements elicitation, seven performed requirements analysis, five applied requirements validation techniques, and only one interviewed claimed he practiced some requirements management technique. Table 3 presents codification obtained with interviews regarding the RE practice in projects.

4.2.1. Elicitation

One coordinator said not to apply a specific method to elicitation, eight realized meetings, one made business analysis and another one realized documents analysis. Five coordinators realized brainstorming, two had done interviews and the other two realized prototyping. One realized observation and the other one realized paper analysis. Four coordinators consulted clients as stakeholders, two consulted users and one consulted expert domain. *Google Meet*, *Google Docs*, and *Figma* are examples of tools used to support requirements elicitation.

4.2.2. Analysis

In response to questions about the analysis, two coordinators said they had done requirements classification; one coordinator cited modeling, another coordinator said they had done prioritization, and a third coordinator said they had done grouping. According to the coordinator responsible for prioritization, he used *Google Meet* to hold a meeting with the client. Other coordinators emphasized the team's and the client's participation in the analysis stage. One coordinator had done use cases and the other one made use of flowcharts as analysis tools.

4.2.3. Specification

Regarding the specification, four coordinators reported using different approaches: one based on natural language, another on an approach based on ISO/IEC/IEEE 29148:2018, and two coordinators wrote requirements as user stories. One coordinator stated that he will use the ML Canvas, while another will apply team-generated reports.

About the attributes identified for requirements specification by the coordinators, we found the task description, author record, and identifier (ID). One coordinator claimed to be responsible for requirements specification. *Google Docs*, *Discord*, *Google Sheets*, *Clickcopy*, and *Gitscrum* were used for requirements specification. We realized that, even when requirements are not registered in an organized manner, the organization's project management tools already assist in doing so.

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

Tabela 3. RE Practice in the CEIA projects.

4.2.4. Validation

Regarding requirements validation, prototyping is the most mentioned method by five people. Only one coordinator uses simulation, and another uses peer review. Six coordinators affirm they consult the client to validate the requirements. *Google Docs* and *Discord* were mentioned as supporting tools for requirements validation.

4.2.5. Management

Only one coordinator interviewed performed the requirements management activity in his project. For this, he used the *Google Sheets* tool.

4.2.6. Other Questions

Six coordinators reported holding meetings regarding change management. Typically, the coordinator chooses whether or not to include the team in meetings that are with clients. Changes are tracked using the *Gitscrum* and *Google Docs* tools, where requirements are frequently documented.

Regarding the team's comprehension of the tasks that need to be completed and the project's anticipated outcomes, all the coordinators reported holding meetings with the team to align them. Three of them keep a weekly schedule for these meetings. Additionally, one coordinator reported that he validated the team's comprehension with *Gitscrum* tool.

All coordinators realized an evaluation of the project's advancement. Seven coordinators said they conducted the assessment based on macro deliveries, while three coordinators said they held weekly meetings to ensure that the progress was in line with the plan. Three coordinators gauge the project's progress by looking at the accuracy of their models, while two coordinators said they have a work plan and regularly review it to check if the project is not meeting expectations.

It was revealed by six of the coordinators that they failed to document the lessons they had learned from the project. Four coordinators, on the other hand, said they documented the lessons they had learned. The team was involved in this registration by two of these coordinators.

Concerning the evaluation of quality, these activities were found to be described by a few techniques, such as weekly meetings and macro deliveries with client participation. Only one coordinator cited other evaluations of quality activities which included response time, model evaluation, customer acceptance criteria, and satisfaction questionnaire. Two coordinators also said they did not conduct any quality assessments due to a lack of knowledge of software quality metrics.

The coordinators focused on two non-functional requirements: model accuracy and performance. Response time, data security, privacy, and usability were cited more than one time. Sustainability, model scalability, data exploration, ease of use, and data requirements appeared just one time.

Coord.	Better elicitation and spec.	Beter data req. and RNF	Difficulty in project	Reduce difficulties
C1	yes, extension process and methods	yes, estabilsh 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 acoording TRL	obtaining data and lack of experience	process RE4ML, specification forms
C3	no, adapted for RDI	no	team and infraestructure 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 infraestructure 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 infraestructure 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 RE4ML
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, obtining data, expectations clients	document for communication

Tabela 4. RE aspects that can be improved in ML projects at CEIA.

4.3. RE for AI

We summarize responses from coordinators who face difficulties in projects that may be related to the lack of RE practice. About how to mitigate these difficulties. In Table 4, we present codification obtained with interviews about how RE practice can be better in AI projects.

4.3.1. Difficulties in ML Projects at CEIA

Seven coordinators responded that data gathering was the biggest challenge they encountered. Three coordinators cited the challenge of acquiring the infrastructure required for projects, and four coordinators characterized managing customer expectations as challenging. According to two coordinators, RDI projects have additional challenges, e.g., the coordinator's lack of experience with this kind of project, understanding of the development process, finances, team communication, process management, commit quality, and technical knowledge, etc.

As measures to lessen challenges that arise during project requirements documents, according to five coordinators, would facilitate communication. According to four coordinators, project management can be aided by implementing a RE process for machine learning. According to two coordinators, specifications can aid in project management and help in client negotiation to get additional time for task completion. According to one coordinator, innovative validation techniques can lessen project challenges.

4.3.2. RE Process

To gauge the coordinator's level of importance regarding the existence of a requirements engineering process for machine learning, we provided them with a Likert scale. Only one coordinator indicated that it was important, while eight coordinators said it was very important.

Concerning the lack of RE in AI, two coordinators believe this lack has to be in the ML domain. Others two coordinators believe that the challenges are due to the characteristics of RDI projects. All these four coordinators feel that the RE process needs to be adjusted to be more dynamic. Another coordinator stated that, in his opinion, cyclical reviews are necessary for the RE process in ML to comply with the data-based construction process. In response, three coordinators said they were unsure of how NFR can be adapted to the AI domain.

4.3.3. RE Improvements

We questioned coordinators about how they thought RE could be enhanced to handle NFR and accomplish AI domain-specific elicitation and specification.

According to five coordinators, better elicitation and specification for the AI domain can be achieved through the extension of diagrams, techniques, and tools. According to four coordinators, improving the requirements engineering process can lead to better specifications and elicitation as well as increased usability in real-world industrial settings. Six coordinators are unaware of the adaptability of requirements engineering. Two coordinators and additional RE methods for AI ought to be suggested. Two more think that better elicitation and specification for this domain can be achieved with increased requirements analyst knowledge. While some coordinators think requirements patterns can be helpful in this area, others think projects can use the model cards approach.

Four coordinators think linking data to the goal during the RE process will ensure better NFR handling. Two coordinators emphasized the challenges in defining metrics to improve NFR care. Additionally, two coordinators think implementing a simplified method in projects can ease specifying NFR. According to a coordinator, NFR ought to be addressed differently based on the TRL of the project. Three coordinators don't know how to specify NFR in AI projects. One coordinator thinks that assessing model efficiency can be a useful tool for identifying non-functional requirements for AI. Another coordinator thinks that the model cards approach can assist in managing NFR. Furthermore, according to one coordinator, requirements engineering cannot assist in obtaining improved NFR for artificial intelligence.

All these aspects synthesize the data obtained from interviews with machine learning project coordinators regarding requirements engineering practice. In the next section (see 5) we will present how this data validated our hypotheses and answered the research questions.

5. Discussion

We created a few hypotheses to accomplish the study's goal of characterizing the state of requirements engineering practice in an industrial setting involving RDI projects. Conventional requirements elicitation techniques, including brainstorming, prototypes, and meetings, are presented by the consulted unit. Customers, users, and domain experts are the stakeholders in this situation. There was no indication of any pertinent data scientist participation. Elicitation activities have been supported by the use of online meetings, project management software, and document creation tools.

Concerning the analysis stage, we observed that coordinators lack a systematic method for categorizing requirements. Typically, requirements are often modeled based on categorization and prioritization according to functionality and the project's main delivery. We were unable to locate a shared set of modeling methods and resources.

The coordinators declared documented the requirements for the practice of defining requirements. Usually, requirements include a date, an identifier, and an explanation of the activity. Requirements were recorded using spreadsheets, online document creation tools, and project management software. The project coordinator is the one in charge of registration. The coordinators stated struggled to decide which methods and resources would work best for documenting needs.

Numerous coordinators claimed to have performed some kind of validation. The interviewees stated that meetings, peer validation, and prototype use are the methods used for validations. There were no found processes, metrics, or tools to assist with the validation process. To validate partial deliveries, the customer is consulted.

Single coordinator is said to oversee and management requirements. Modifying the specification document's requirement is how this kind of management is done. As a result, not a single coordinator mentioned using a management procedure that addresses artifact traceability.

Agile development methods, mostly centered on Scrum, are used on all projects. The coordinators gave evidence that they had taken care of a few non-functional requirements aimed at the model, including usability, accuracy, and response time. We do point out that this is an uncommon and poorly established practice. We also found that coordinators meet with the team to make sure everyone is aware of the tasks and deliverables, and that requirement documentation typically acts as a record of what is done rather than directing the development process.

It has come to our attention that coordinators recognize the value of the requirements engineering process in the development of AI systems. Furthermore, they acknowledge that defining the process and artifacts can aid in development. The coordinators confirmed that traditional requirements engineering would not likely be successful in the unit if it were used, but they also stated that they did not know how this could be accomplished due to the nature of RDI projects and the feature of experimentation-based AI development.

Lastly, coordinators with more experience with requirements engineering and software engineering also showed a greater concern for requirements-related aspects of their projects. The application of these activities in projects is not supported by the involvement of SE or RE specialists unless this specialist is involved in requirements practices, which is not frequent in the projects consulted.

Regarding related work, we identified some similarities with the state of practice in our study. Such as the adoption of the agile scrum methodology, elicitation practices with brainstorming and interviews, and forms of documentation based on natural language. Another similarity is that there is the treatment of non-functional requirements, however it is poorly structured and the biggest concern is with the performance of the models. Concern with the RE process for the development of AI systems is also common in the works.

The null hypotheses of this study were all validated. From H01 to H05 we noticed that the unit does not present an appropriate method or process for practices in each of the requirements engineering phases. Each project adopts a practice and we realize that the analysis, validation and management phases are overcome in almost all projects. Traditional elicitation and specification approaches are typically applied, but without adequate registration or which makes traceability difficult. Regarding hypothesis H06, non-functional requirements are also neglected in projects, normally not addressing aspects of fairness, reliability and transparency that are so important for artificial intelligence systems. We realize that the absence of the requirements engineering professional on the team can impact specific or insufficient team decisions and execution of the requirements engineering process on projects or that affect the quality of the product [Chemuturi 2013, Hussain and Mkpojiogu 2016]. The same goes for when the coordinator has little knowledge about software engineering without adequate support from the requirements engineering professional to direct the requirements engineering process.

6. Threats to validity

Descriptive validity refers to the possibility an interviewer may not gather all pertinent information during the interview. We recorded the interviews in video and storage on a computer and Google Drive to mitigate this threat. We annotated the transcripts using the Reshape tool and we reviewed the transcripts. The interviews were conducted with a strict adherence to a precise and objective protocol. We aimed to entice project coordinators to engage in the research, which was non-obligatory and lacked benefits, but to evaluate the outcomes through the execution of the activities. It is posited that the semi-structured nature may potentially introduce biases in the research. Pilot testing of the questionnaires was conducted with the assistance of a project coordinator to anticipate and address potential issues in the interpretation of the questionnaires. There were no modifications made to the substance of the inquiries. Only the structuring was reformulated, and the issues began to be enumerated. The planning and execution of the evaluation were conducted following the planning presented by Wohlin in his book Experimentation for Software Engineering [Wohlin et al. 2012].

Concerning interpretation validity, i.e., the possibility of misunderstandings between interviewees and researchers, we mitigate this potential risk by communicating the study objective to the participants at the beginning of each interview. We explained the RE concepts to the participants. The interview guide was reviewed to improve its reliability. The concepts of researcher bias and theoretical validity pertain to the researcher's bias towards interpreting interviews in a manner that aligns with their objectives or initial theory. As this is our initial investigation in this particular domain, we do not possess a specific research methodology we wish to promote. We were very open to the outcomes of the interviews.

The term 'reactivity' refers to the possibility that interviewees may exhibit divergent behavior due to the interviewer's presence. It is not feasible to eliminate reactivity, however, we are cognizant of its potential impact on the observed phenomena. We attempt to mitigate this threat by highlighting the significance of the study and the potential for enhancements within the organization by identifying real needs.

Additionally, a threat to validity was the execution of the research in a single

Brazilian unit. We had limitations regarding the feasibility of carrying it out in other places, but we suggest that this research be replicated in other settings so that it can become more complete and broader.

7. Conclusions

To understand how the most complex activity in the development of ML-based AI systems has been developed in practice and what are the opportunities for improvement and research. We have conducted a qualitative interview study to understand the perception of and practices for RE in development projects of AI systems in the RDI scenario.

In general, coordinators believe that:

- requirements activities can be improved in projects,
- traditional RE and already known artifacts would hardly be successful in the unit,
- do not know how to apply RE practices to RDI projects involving AI,
- the treatment of NFR can be improved in projects,
- a requirements document can help in communicating projects
- the RE process adapted to ML development can improvement quality of deliveries and stakeholders satisfaction.

Based on the results of our research, we conclude that it is complementary to what already exists regarding the state of RE practice for ML. Although we do not address specific types of requirements or a specific category in the domain, we interviewed project coordinators with different knowledge and experiences. RE practices change but challenges remain similar.

We think that combining artifacts appropriate for the AI domain with a requirements engineering process can help mitigate challenges found during project development, resulting in higher-quality deliveries and the verification of these qualities. Additionally, it can act as a guide for team members who are unfamiliar with this kind of system's requirements practice.

The primary constraint on this research was its examination within a restricted context at the Embrapii facility. However, we think the findings apply to other organizations that similarly do RDI projects. Additional research is required to look into other RDI units and apply methods, techniques, and tools in these kinds of situations by combining agile approaches with ad hoc approaches and research proposals.

8. Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Referências

Alves, A. P. S. et al. (2024). Status quo and problems of requirements engineering for machine learning: Results from an international survey. In Kadgien, R., Jedlitschka, A., Janes, A., Lenarduzzi, V., and Li, X., editors, *Product-Focused Software Process Improvement*, pages 159–174, Cham. Springer Nature Switzerland.

Chemuturi, M. (2013). *Requirements Engineering and Management for Software Development Projects*. Springer New York, 1st edition.

Corbin, J. and Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications, 4th edition.

Dunbar, B. (2017). National aeronautics and space administration (nasa). technology readiness level. https://esto.nasa.gov/trl/.

Glaser, B. G. (1992). *Basics of grounded theory analysis: Emergence vs forcing*. Sociology Pr.

Glaser, B. G. and Strauss, A. L. (2017). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Routledge.

Habibullah, K. M., Gay, G., and Horkoff, J. (2023). Non-functional requirements for machine learning: understanding current use and challenges among practitioners. *Requirements Engineering*, 28(2):283–316.

Hussain, A. and Mkpojiogu, E. O. C. (2016). Requirements: Towards an understanding on why software projects fail. *AIP Conference Proceedings*, 1761(1):020046.

Ishikawa, F. and Yoshioka, N. (2019). How do engineers perceive difficulties in engineering of machine-learning systems? - questionnaire survey. In *IEEE/ACM Joint International Workshop on Conducting Empirical Studies in Industry and International Workshop on Software Engineering Research and Industrial Practice*, pages 2–9.

Silva, G. R. S. and Canedo, E. D. (2022). Requirements engineering challenges and techniques in building chatbots. In *International Conference on Agents and Artificial Intelligence*.

Stol, K.-J., Ralph, P., and Fitzgerald, B. (2016). Grounded theory in software engineering research: A critical review and guidelines. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, pages 120–131.

Villamizar, H., Escovedo, T., and Kalinowski, M. (2021). Requirements engineering for machine learning: A systematic mapping study. In *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pages 29–36.

Vogelsang, A. and Borg, M. (2019). Requirements engineering for machine learning: Perspectives from data scientists. *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, pages 245–251.

Wohlin, C., , et al. (2012). *Experimentation in Software Engineering*. Springer Berlin Heidelberg.