Defining AI Roles in Requirements Elicitation and Analysis

Master's Thesis in Computer science and engineering

Yuhan Li
Keyu Ren

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2026

Master's Thesis 2026

Defining AI Roles in Requirements Elicitation and Analysis

Yuhan Li

Keyu Ren

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2026

Defining AI Roles in Requirements Elicitation and Analysis
Yuhan Li & Keyu Ren

Supervisor: Richard Berntsson Svensson & Lekshmi Murali Rani, Department of Computer Science and Engineering
Examiner: Eric Knauss, Department of Computer Science and Engineering

Master's Thesis 2026
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2026

Yuhan Li & Keyu Ren
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Artificial intelligence is increasingly used to support Requirements Engineering (RE), especially in text-intensive activities such as summarizing stakeholder input, extracting candidate requirements, and organizing requirements-related information. Yet in early RE, where requirements elicitation and requirements analysis depend on contextual interpretation, incomplete information, and stakeholder negotiation, it remains unclear how responsibilities should be allocated between human practitioners and AI systems.

This thesis investigates how human–AI roles are configured in requirements elicitation and requirements analysis, and how practitioners evaluate the benefits, risks, and trust conditions associated with these configurations. A sequential exploratory mixed-methods design was used, combining six semi-structured interviews with software engineering practitioners and a follow-up survey with 67 valid responses. Interview data were analyzed thematically and used to inform the survey design, while survey data were analyzed descriptively.

The findings show that AI involvement in early RE is task-contingent rather than general. Practitioners were most willing to use AI as an assistant or preliminary processor in bounded, information-heavy, and verifiable tasks, including summarizing, extracting, organizing, and structuring requirements-related material. By contrast, tasks involving stakeholder interaction, interpretation of implicit needs, prioritization, trade-off decisions, and accountability remained strongly human-led. AI-assisted configurations were perceived as the most useful, whereas AI-only configurations were rare and associated with the highest perceived risk. Trust in AI was therefore conditional rather than general: it increased when outputs were easy to verify and tasks were low in decision criticality, and decreased when work depended on contextual judgment, sensitive information, or consequential decisions.

Based on exploratory mixed-methods evidence, this thesis proposes a practitioner-informed role-boundary framework at the task level for AI involvement in early RE. The framework clarifies where AI can productively support elicitation and analysis work and where human judgment should remain dominant. Overall, the study argues that AI should be integrated into early RE as a calibrated collaborative support tool rather than as a substitute for human practitioners.

Keywords: Requirements Engineering, Artificial Intelligence, Human–AI Collaboration, Role Boundary, Requirements Elicitation, Requirements Analysis, Trust.

Acknowledgements

We would like to express our sincere gratitude to our supervisors, Richard Berntsson Svensson and Lekshmi Murali Rani, for their guidance, support, and constructive feedback throughout this thesis project. Their advice helped us refine the research focus, improve the study design, and strengthen the structure and clarity of this thesis.

We would also like to thank our examiner, Eric Knauss, for his valuable comments and feedback during the thesis process. His feedback helped us strengthen the connection between the research questions, methodology, results, and discussion.

We are also grateful to the Department of Computer Science and Engineering at Chalmers University of Technology and the University of Gothenburg for providing the academic environment and resources that supported this thesis work. We would also like to thank the teachers and faculty members who guided us during the past two years of study, as their courses and feedback helped us build the knowledge and skills needed to complete this thesis. We also thank the programme coordinators and programme managers of our respective master's programmes for their guidance and support regarding the thesis process and administrative procedures.

We would like to express our gratitude to all interview participants and survey respondents who generously shared their time, experiences, and perspectives. Their contributions were essential to this study and provided the empirical foundation for understanding human–AI role allocation in early Requirements Engineering.

Finally, we would like to thank our classmates and friends for their encouragement, discussions, and support during the thesis period. We also extend our heartfelt thanks to our families for their patience, understanding, and continuous support throughout this thesis journey.

Yuhan Li & Keyu Ren
Gothenburg, June 2026

Contents

List of Figures

List of Tables

1

Introduction

1.1 Motivation

Requirements Engineering (RE) plays a foundational role in software development, as early requirements decisions shape system functionality, architectural direction, and long-term project outcomes. Among the different phases of RE, requirements elicitation and requirements analysis are particularly critical, as they involve interpreting stakeholder intentions, negotiating meanings, resolving ambiguities, and organizing incomplete information prior to formal specification [1]. Decisions made during these early stages strongly influence downstream development activities, meaning that misunderstandings or premature structuring may propagate throughout the software lifecycle.

At the same time, rapid advances in artificial intelligence, particularly large language models, are increasingly influencing early RE activities [2, 3]. AI systems can support tasks such as automated summarization of stakeholder input, clustering of requirement statements, identification of potential conflicts, detection of ambiguity, and categorization of requirements [2, 3]. These capabilities suggest that AI may provide analytical support in text-intensive RE tasks, for example in organizing and preliminarily processing diverse stakeholder expressions [2, 3]. The increasing use of AI tools in early RE activities therefore raises important questions regarding how such tools should be integrated into human-centered requirements practices.

1.2 Problem Statement

Early Requirements Engineering is not merely an information-processing activity but an interpretive and socially situated practice that relies heavily on human judgment, experience-based reasoning, and negotiation skills [4, 1]. Compared with later RE phases that aim to formalize and precisely document already negotiated requirements, ambiguity in early RE often stems from stakeholders not yet fully clarifying their needs or not knowing how to articulate them clearly. At the same time, the team's understanding of the problem evolves through ongoing discussion and revision.

In this context, introducing AI raises a central question: what role should AI assume in early RE tasks, and how should responsibilities be allocated between humans and AI?

Existing research on AI-supported Requirements Engineering has primarily focused on technical capabilities, tool performance, or algorithmic feasibility [2, 3]. Although some studies have examined human–AI collaboration and levels of automation [5, 6], there remains a lack of systematic frameworks that define AI roles at the task level within the highly ambiguous context of early requirements elicitation and analysis.

1.3 Significance of the Study

The absence of clear role boundaries between humans and AI during early Requirements Engineering activities, particularly during requirements elicitation and preliminary requirements analysis, presents both practical and theoretical challenges [6]. Without explicit delineation of responsibility, AI may assume unclear or disproportionate decision authority in tasks that require contextual judgment, while in tasks suitable for human–AI collaboration its supportive potential may remain underutilized. Such role imbalance may lead to misinterpretation of stakeholder needs, overlooked conflicts, reduced traceability, and blurred accountability. Given the long-term impact of early RE decisions, unclear role allocation may further amplify project risks.

Clarifying how responsibilities should be distributed between humans and AI is therefore essential for enabling responsible and effective AI integration in early Requirements Engineering. A clearer understanding of role allocation may also contribute to the broader discussion of human–AI collaboration in complex, judgment-intensive software engineering activities.

1.4 Research Focus

In human–AI collaborative settings, the concept of a “role” refers to the distribution of responsibilities, decision authority, and task execution between human actors and intelligent systems. Research on human–automation interaction has long emphasized that the allocation of functions between humans and automated systems strongly shapes system performance, user trust, and accountability [5]. Similarly, human–AI interaction guidelines stress careful task sharing between humans and AI systems in collaborative workflows [6].

Building on these perspectives, this study adopts a task-level perspective to examine how responsibilities between humans and AI are distributed in early Requirements Engineering activities, particularly during requirements elicitation and requirements analysis. Focusing on individual tasks allows a more precise examination of how different forms of human–AI collaboration may emerge in practice.

Within this task-level perspective, four typical forms of human–AI role configuration are considered: human-only, AI-assisted, AI-led, and AI-only. These configurations represent different ways in which decision authority and task responsibility may be distributed between humans and AI, ranging from tasks performed entirely by humans, to situations where AI provides assistance, takes a leading role under human supervision, or performs tasks independently.

The study also considers the broader context in which AI is deployed in practice, including current AI capabilities, organizational compliance requirements, and differences in how teams understand and approach AI. Within these real-world constraints, the study examines how roles and responsibilities are interpreted, assigned, and justified in human–AI collaboration.

1.5 Research Questions

Using an exploratory mixed-method design combining semi-structured interviews and a follow-up survey, the study investigates two research questions.

RQ1. How are roles and responsibilities between humans and AI delegated in requirements elicitation and requirements analysis tasks?

This question is descriptive in nature and aims to understand how practitioners currently allocate responsibilities between humans and AI across concrete early RE tasks in real-world settings.

RQ2. What benefits, risks, and trust considerations are associated with these human–AI role configurations?

This question is exploratory and focuses on how practitioners evaluate the implications of different human–AI collaboration arrangements.

Both research questions are addressed through the sequential qualitative–quantitative design. The interview phase provides in-depth insights into practitioners' reasoning about human–AI role configurations, benefits, risks, and trust considerations, while the survey phase complements these insights by examining how these patterns appear across a broader group of practitioners.

1.6 Research Paradigm and Methodological Orientation

This study is guided by a pragmatic research paradigm and adopts a mixed-methods design, with an interpretivist orientation in the qualitative interview phase. This combination was necessary because the research problem requires both practical and interpretive forms of understanding. The study aims to examine how responsibilities between humans and AI can be allocated across early Requirements Engineering (RE) tasks in ways that are meaningful, context-sensitive, and useful for practitioners. This required a design that could both identify broader patterns of role allocation and interpret how practitioners understand their work, explain their decisions, and reason about trust, risk, and responsibility in specific task contexts. Therefore, a pragmatic mixed-methods research design was appropriate because it allowed the study to combine interpretive interview evidence with broader descriptive survey evidence in response to the research problem and the type of knowledge required [7, 8].

Within this overall pragmatic design, the interview phase provided the interpretive basis for the study. The semi-structured interviews were used to explore how practitioners understand requirements elicitation and analysis in their own work, how they make sense of AI involvement, and how they explain the boundaries between human and AI responsibilities. The interview data were therefore treated as situated accounts of practitioners' perceptions, judgments, and experiences, rather than as objective measurements of practice.

The follow-up survey complemented the interviews by capturing broader descriptive patterns across tasks, role configurations, perceived benefits, risks, and trust considerations. Combining the two forms of evidence allowed the study to connect practitioners' situated explanations with wider patterns in human–AI role allocation, which is consistent with the logic of mixed methods research [9]. In this way, the pragmatic paradigm provided the overall rationale for combining qualitative and quantitative evidence, while the interpretivist orientation supported the qualitative phase by enabling a deeper understanding of how practitioners reason about human–AI collaboration in early RE.

At the same time, the quantitative phase was designed to use descriptive rather than inferential statistics. The survey was used to examine whether patterns identified in the interviews also appeared across a broader practitioner group, not to test predefined hypotheses, compare groups statistically, model variable relationships, or estimate population parameters. Given the study's exploratory aim, non-probability recruitment strategy, modest and uneven subgroup sizes, and mainly categorical or ordinal survey items, inferential tests would have implied a degree of statistical generalizability that the design was not intended to support. Descriptive analysis was therefore the more methodologically defensible choice for summarizing response patterns in practitioners' role allocations and evaluations of AI involvement.

By addressing the research questions through this design, this thesis makes three contributions.

First, it develops a task-level conceptualization of AI roles in requirements elicitation and analysis, distinguishing four human–AI role configurations: human-only, AI-assisted, AI-led, and AI-only.

Second, it provides an empirically grounded mapping of how these human–AI role configurations are currently distributed across early Requirements Engineering tasks from the perspective of practitioners.

Third, it offers structured guidance for responsible and context-sensitive human–AI collaboration in early Requirements Engineering tasks characterized by ambiguity and interpretive complexity.

1.7 Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 reviews related work on AI support in Requirements Engineering, human–automation role allocation, and trust in automated systems. Chapter 3 presents the research methodology, including study design, data collection, and analysis procedures. Chapter 4 presents the empirical results of the study. Chapter 5 discusses the implications of these results for AI role configuration in early Requirements Engineering. Finally, Chapter 6 concludes the thesis and outlines directions for future research.

2

Related Work

2.1 Background

Requirements Engineering (RE) refers to activities for identifying, analyzing, documenting, and managing system requirements so that software systems can satisfy stakeholder needs and constraints [4]. Classical RE processes typically include requirements elicitation, analysis, specification, and validation. In this study, the early phases of RE refer mainly to elicitation and analysis. Elicitation focuses on capturing stakeholder needs, expectations, and constraints, while analysis involves interpreting, organizing, and refining requirement information to support later design and implementation decisions [4].

Requirements elicitation commonly involves interactions with stakeholders through interviews, workshops, and analysis of existing documentation. Because stakeholders rarely provide complete and consistent requirement statements, elicitation often requires analysts to interpret partial information, clarify ambiguous expressions, and consolidate inputs from multiple sources. Prior research has shown that requirement statements frequently contain implicit assumptions and context-dependent meanings that must be interpreted by analysts [10].

Requirements analysis builds on elicited information and focuses on structuring and refining requirement statements. This includes identifying inconsistencies, clarifying assumptions and constraints, and organizing requirements so that they can support traceability and further refinement. Early phases of Requirements Engineering are widely recognized as challenging. Because requirements are often expressed informally and stakeholders may have incomplete or evolving understandings of the problem domain, requirements gathered during elicitation are frequently ambiguous, inconsistent, or incomplete.

As a result, practitioners must repeatedly interpret stakeholder input, negotiate different perspectives, and iteratively refine requirements during the analysis process [4]. In practice, requirements elicitation and analysis are therefore often interleaved, particularly in early Requirements Engineering where information remains incomplete and problem understanding continues to evolve. Prior empirical studies have shown that communication gaps between stakeholders and development teams, inconsistent terminology, and ambiguity in natural-language requirements remain persistent challenges in practice [1, 10]. These characteristics make early RE activities highly interpretive and dependent on human judgment, while also motivating research on tools and techniques that can assist practitioners in processing and structuring requirements information [3, 11]. Recent advances in artificial intelligence, especially LLM-based approaches, have created new opportunities to support text-intensive requirements work, including requirements classification, drafting, and generation from informal inputs [3, 2, 12]. This motivates the following review of AI-supported approaches in Requirements Engineering.

2.2 AI Support for Early-Stage Requirements Engineering

In recent years, the application of artificial intelligence in Requirements Engineering (RE) has expanded significantly, particularly in requirements elicitation and analysis. Some studies have explored structured automation approaches, such as multi-agent system frameworks, to organize and process inputs from multiple stakeholders during elicitation and preliminary analysis [13]. These approaches emphasize automated mechanisms for structuring and consolidating requirement information, but pay limited attention to how responsibilities should be allocated between human engineers and AI systems at the task level.

Natural language processing (NLP) techniques have also been applied to requirements analysis. Earlier work has shown that NLP methods can detect ambiguity and inconsistencies in requirement statements, helping analysts identify potentially problematic formulations at an early stage [10]. This line of work demonstrates that AI can support pattern recognition and anomaly detection in textual requirements. However, it mainly treats AI as an analytical support tool, while final interpretation and judgment remain the responsibility of human engineers.

With the emergence of large language models (LLMs), generative AI has increasingly been applied to text-intensive RE tasks, including requirement summarization, classification, clustering, and conflict detection [2]. Prompt-based approaches have also been proposed to integrate generative models into RE workflows, particularly for supporting early-stage analysis [3]. More recent LLM-oriented RE work also shows that LLMs can support requirements drafting or generation from informal inputs, but that the quality of LLM-generated requirements depends on input clarity and prompt design and still requires human oversight for refinement and validation [12]. Work at the intersection of formal requirements engineering and large language models also emphasizes that LLM-generated requirements artifacts raise concerns about correctness, fairness, and trustworthiness [14]. Industry-oriented studies indicate that AI-supported RE practices are increasingly adopted in real-world settings, with human–AI collaboration representing the dominant mode of use [11].

Taken together, existing research shows that AI can technically support several activities in early-stage Requirements Engineering, and that such support is increasingly relevant in practice. However, much of this work remains capability-oriented, focusing on what AI systems can support or generate rather than on how responsibilities should be distributed between humans and AI across different types of RE tasks.

2.3 Human–Automation Delegation and Role Allocation

While technical studies primarily focus on AI capabilities, another stream of research examines how responsibilities should be distributed between humans and automated systems. Research suggests that when intelligent tools are integrated into workflows without clearly defined role boundaries, they may lead to over-reliance or reduced critical evaluation [15]. This indicates that AI adoption is not merely a technical issue but fundamentally a matter of role allocation and responsibility distribution.

For the purposes of this thesis, role allocation is understood at the level of concrete RE tasks rather than at the level of a system as a whole. Roles refer to the function each actor performs in the task, whereas responsibilities refer to who retains interpretive authority, verification duties, final decision-making power, and accountability for outcomes. This distinction is important in early RE because a task may be partially automated while responsibility still remains primarily human.

Automation research conceptualizes human–automation interaction as a continuum characterized by varying levels of automation [5]. These models describe a spectrum from full human control to high degrees of automation, with each level reflecting different allocations of authority, decision-making power, and responsibility. They help explain how control can be distributed between humans and automated systems.

However, automation-level models were primarily developed in structured task environments, such as aviation control, industrial systems, or decision-support contexts, where objectives are stable and performance can be evaluated against predefined criteria. Early-stage Requirements Engineering differs substantially from such contexts. These activities involve interpretive judgment, negotiation of stakeholder intentions, context-dependent meaning construction, and evolving problem understanding.

This limitation is particularly important in early RE because its tasks are internally heterogeneous. Activities such as summarizing stakeholder input, identifying conflicting statements, interpreting ambiguous expressions, or negotiating requirement meanings vary significantly in cognitive complexity and responsibility sensitivity. As a result, different tasks may warrant different degrees and forms of AI participation. Yet existing research has not systematically examined how automation-level distinctions can be translated into task-level role configurations within ambiguity-intensive RE environments.

Recent work on task delegability further supports the need to move beyond capability-oriented views of AI. Lubars and Tan argue that delegation to AI should be examined not only in terms of what AI can do, but also in terms of what AI should do and how much control humans prefer to retain [16]. Their framework identifies motivation, difficulty, risk, and trust as factors shaping delegation decisions, and their empirical findings show little preference for full AI automation and a stronger preference for machine-in-the-loop arrangements, where humans retain the leading role while AI provides support. This perspective is relevant to early RE because the suitability of AI involvement cannot be judged only by technical capability. It also depends on whether the task is verifiable, whether it involves stakeholder interaction, and whether the outcome requires contextual interpretation or accountable decision-making.

2.4 Trust and Appropriate Reliance in Interpretive Requirements Tasks

Closely related to role allocation is the issue of trust in automated systems. Trust is a critical factor shaping human interaction with automation, and classical work on trust in automation argues that system design should promote appropriate reliance by avoiding both overtrust and undertrust [15]. When users' perceptions of system capability diverge from actual performance, inappropriate reliance may occur. Parasuraman and Riley distinguish between use, misuse, disuse, and abuse of automation, where misuse refers to overreliance on automation and disuse refers to the underutilization or rejection of automation despite its potential value [17]. This distinction is relevant to AI-supported early RE because both extremes are problematic: practitioners may rely too strongly on AI-generated outputs without sufficient contextual verification, or they may avoid AI support even in bounded and verifiable tasks where it could reduce effort and improve coverage.

More recent theoretical work on human–AI trust argues that trust should not be understood as a general attitude toward an AI system, but as trust that the system will uphold a specific expected behavior or "contract" in a given context [18]. This view is useful for this thesis because trust in early RE is also task-specific. Practitioners may trust AI to support a limited and verifiable contract, such as summarizing a meeting, extracting candidate requirements, or organizing existing material, while withholding trust when the expected contract involves interpreting stakeholder intent, resolving ambiguity, or making accountable decisions. This also helps explain why over-reliance can become problematic: trust is appropriate only when it is grounded in the AI system's actual capability to maintain the task-specific contract.

Empirical reviews of human trust in AI further show that trust is shaped by AI characteristics, task context, and users' perceptions rather than by technical capability alone. Glikson and Woolley identify transparency, reliability, task characteristics, and perceived machine intelligence as important factors influencing cognitive trust in AI [19]. This is relevant to early RE because practitioners may trust AI more in bounded, text-processing tasks where outputs can be reviewed, but remain cautious in tasks requiring social intelligence, contextual interpretation, or accountable judgment.

In the context of AI, it is also necessary to distinguish between trust in AI and trustworthy AI. Trust in AI refers to users' willingness to rely on AI systems, while trustworthy AI concerns whether the system has properties that justify such reliance. Toreini et al. relate trust in AI to trustworthy machine learning technologies and identify fairness, explainability, auditability, and safety as key categories that can support trustworthiness [20]. This distinction is relevant to early RE because practitioners' trust in AI should be grounded in whether the system can produce outputs that are reliable, reviewable, and appropriate for the task, rather than in general confidence in AI as a technology.

Building on these classical and AI-specific perspectives, trust in AI-supported early RE should be understood as calibrated and task-dependent rather than unconditional. Recent human–AI trust research strengthens this view by showing that appropriate trust should be warranted by the AI system, the task, and the interaction context, rather than simply increased [21]. This is especially important because users may overrely on AI advice in uncertain situations, or overestimate the accuracy of LLM-generated answers when explanations appear detailed or confident [22, 23]. For early RE, these insights suggest that practitioners' reliance on AI should be aligned with task context, output verifiability, error risks, and accountability requirements, rather than with general confidence in AI as a technology.

Trust calibration becomes particularly complex in environments characterized by ambiguity and outcome uncertainty. Requirements elicitation and requirements analysis represent such environments [4]. These phases involve interpreting stakeholder intentions, clarifying vague expressions, identifying conflicting requirements, and integrating context-dependent information. Unlike structured decision tasks, these activities rarely have a single objectively correct solution; instead, they rely on interpretive coherence, contextual sensitivity, and negotiated meaning.

In such interpretive contexts, AI-generated outputs may exhibit high linguistic fluency and structural coherence while lacking a fully grounded understanding of stakeholder intent. If engineers over-rely on seemingly plausible AI analyses, misunderstandings may become prematurely stabilized and propagate into later development stages. Consequently, trust in early RE is not merely a matter of technology acceptance; it directly relates to how much analytical authority and interpretive responsibility should be assigned to AI across different task types.

Research on human-centered and trustworthy AI emphasizes the importance of transparency, accountability, and explicit human oversight in AI-assisted processes [24]. These principles suggest that role boundaries should be made explicit when AI systems participate in decision-relevant activities, especially in contexts of interpretive uncertainty. While much of this work has focused on safety-critical or decision-support domains, its governance perspective is also relevant to early RE tasks, where misinterpretations may have significant downstream consequences.

Despite these theoretical insights, existing research has not systematically examined how principles of trust calibration and human-centered oversight can inform task-level AI role configurations in requirements elicitation and analysis.

2.5 Research Gap

Taken together, existing literature demonstrates the growing technical feasibility of AI support in requirements elicitation and analysis. Prior research has shown AI's potential in structuring stakeholder input and detecting ambiguities in requirements [13, 10]. More recent studies explore the use of large language models in text-intensive RE tasks [2], while industry-oriented research confirms increasing adoption of AI-supported practices in real-world settings [11].

At the same time, automation-level theories provide abstract models for distributing authority between humans and automated systems [5], while task-delegability research highlights that AI involvement should be considered not only in terms of capability, but also in terms of risk, trust, and preferred human control [16]. Research on trust in automation further emphasizes the importance of calibrated reliance and the risks of both overreliance and underuse [15, 17]. Empirical research on trust in AI also shows that trust is shaped by AI characteristics, task context, transparency, reliability, and users' perceptions [19]. In AI contexts, this further requires distinguishing between users' trust in AI and the properties that make AI systems worthy of trust [20]. These theoretical foundations offer valuable perspectives for understanding human–AI collaboration.

However, these research streams remain largely fragmented. Technical studies predominantly focus on what AI systems can do, automation research provides generalized authority models developed mainly in structured task environments, and trust research often examines reliance mechanisms in safety-critical or decision-support contexts. Few studies systematically integrate these perspectives within the specific setting of requirements elicitation and analysis.

Early RE tasks are characterized by ambiguity, interpretive judgment, and evolving stakeholder intentions. They also vary substantially in cognitive complexity and responsibility sensitivity. Activities such as summarizing stakeholder input, identifying conflicting statements, interpreting ambiguous expressions, or negotiating requirement meanings do not carry the same interpretive risks or accountability implications. Yet current research lacks a task-level conceptual framework explaining how AI roles should be configured across different types of early RE activities.

This study addresses this gap by conceptualizing task-level AI role configurations in requirements elicitation and analysis. This gap motivates RQ1, which examines how task-level human–AI role configurations are allocated across elicitation and analysis tasks. It also motivates RQ2, which examines the benefits, risks, and trust considerations associated with these configurations.

3

Methodology

3.1 Research Design

This study adopts a mixed-methods research design to investigate how practitioners allocate roles between humans and AI models in early Requirements Engineering activities, and how they perceive the potential benefits and risks of AI involvement. Mixed methods research combines qualitative and quantitative approaches within a single study in order to provide a more comprehensive understanding of a research problem than either approach alone [7, 9]. In software engineering research, mixed methods are particularly useful for studying complex socio-technical phenomena, where neither qualitative nor quantitative methods alone are sufficient to capture both contextual depth and broader empirical patterns [8].

The choice of a mixed-methods design is motivated by the exploratory nature of the research topic. The role of AI in Requirements Engineering is still evolving, and existing research has primarily focused on technical capabilities rather than practitioner perspectives or task-level role allocation. Therefore, relying on a single research approach may not sufficiently capture both the depth of practitioner experiences and the broader patterns of AI usage across industry contexts.

In this study, qualitative and quantitative perspectives are integrated to address the research questions from complementary angles. The qualitative component enables the exploration of practitioners' experiences, perceptions, and reasoning regarding the use of AI in Requirements Engineering tasks. At the same time, the quantitative component allows the study to examine whether similar perceptions and patterns appear across a broader group of practitioners.

The study follows a sequential exploratory design in which qualitative insights are first developed and then complemented by a quantitative investigation [7]. This design is appropriate because the initial interviews are used to (i) surface and refine the task-level role configurations of AI involvement in requirements elicitation and analysis (e.g., human-only, AI-assisted, AI-led, AI-only) and (ii) identify the main benefit–risk–trust considerations associated with these configurations. These qualitative findings then directly inform the design of the subsequent questionnaire survey, including the selection of tasks, the wording of role options, and the survey items used to assess perceived benefits, risks, and trust.

To operationalize this design, the study combines semi-structured interviews and a questionnaire survey. The interviews serve as the exploratory phase (primarily supporting RQ1), while the survey provides descriptive evidence on how these role configurations and trust considerations appear across a broader participant group (primarily supporting RQ2).

Figure 3.1 summarizes the sequential exploratory design and shows how findings from the interview phase informed the subsequent survey design.

The following sections describe the data collection procedures and data analysis methods used in this study.

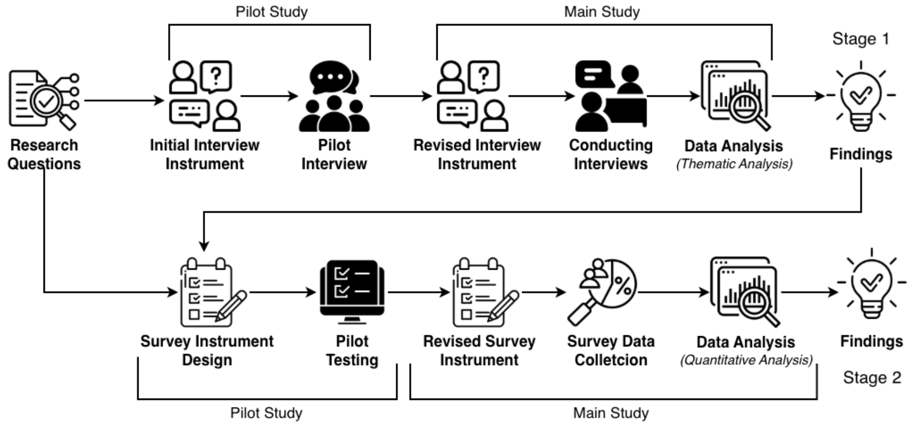

Figure 3.1: Sequential exploratory mixed-methods design used in this study

3.2 Research Questions

This study is exploratory and descriptive in nature, aiming to understand how human and AI capabilities can be combined in the early phases of Requirements Engineering, specifically requirements elicitation and requirements analysis. Rather than testing predefined hypotheses, the study investigates practitioners' reasoning, expectations, and evaluations regarding the configuration of roles between humans and AI in these tasks. Therefore, the study is guided by two exploratory research questions.

The core constructs used in this chapter build on the conceptual framing introduced in the Background and Related Work chapters. In this study, roles and responsibilities refer to the task-level allocation of execution, decision-making, and accountability between human practitioners and AI systems, while trust considerations refer to the conditions under which practitioners regard AI-supported outputs as sufficiently reliable for practical use.

RQ1. How are roles and responsibilities between humans and AI delegated in requirements elicitation and requirements analysis tasks?

Rationale. RQ1 examines how roles and responsibilities are delegated between humans and AI across different Requirements Engineering tasks. Requirements elicitation and requirements analysis activities involve varying levels of ambiguity, interpretive complexity, and contextual dependency, making them suitable for different forms of human–AI role configuration. By investigating how task-level role configurations emerge in practice—such as human-only work, AI-assisted support, AI-led contributions, or limited AI autonomy—this question seeks to identify how responsibilities are assigned across elicitation and analysis activities.

Rather than prescribing ideal role assignments, this question aims to empirically identify how practitioners conceptualize the functional contributions of humans and AI in these tasks. Understanding task-level role configuration provides a foundation for characterizing AI roles in Requirements Engineering and distinguishing tasks where AI participation is considered appropriate from those where human judgment remains central.

RQ2. What benefits, risks, and trust considerations are associated with these human–AI role configurations?

Rationale. RQ2 examines how practitioners evaluate the consequences of delegating specific Requirements Engineering tasks to AI. This includes potential benefits such as improved efficiency or analytical coverage, as well as risks related to misinterpretation, loss of contextual nuance, or reduced accountability. In addition, the question considers how trust influences practitioners' willingness to rely on AI-supported outputs in Requirements Engineering tasks.

The purpose of this question is to assess whether particular human–AI role configurations can support reliable outcomes in requirements elicitation and requirements analysis. Understanding how benefits, risks, and trust considerations interact helps clarify under what conditions AI-supported collaboration may be considered dependable in practice.

Since the study aims to (i) identify how human–AI role configurations are allocated across requirements elicitation and analysis tasks and (ii) assess the benefits, risks, and trust concerns associated with these configurations, a sequential mixed-methods design is appropriate. By combining qualitative and quantitative processes, the study seeks to obtain both detailed experiential insights and a broader understanding of current practices related to the use of AI in Requirements Engineering [9].

Stage 1 (Qualitative – interviews) is used as an exploratory step to surface and refine the set of task-level role configurations (e.g., human-only, AI-assisted, AI-led, AI-only) and the reasoning behind how responsibilities are allocated. This phase addresses both research questions through in-depth qualitative evidence, and its output is used to shape the survey constructs and item wording.

Stage 2 (Quantitative – survey) is then used to examine how widely the role configurations identified in Phase 1 are supported and to assess perceived benefits, risks, and trustworthiness across tasks. This phase primarily supports RQ2, while also offering descriptive evidence that helps validate and contextualize patterns emerging from RQ1.

Finally, findings from both phases are integrated in the interpretation stage to provide a coherent account of role allocation and its implications for feasible and trustworthy AI involvement in early elicitation and analysis.

3.3 Participant Selection

Participants in this study were software engineering practitioners with experience in requirements-related activities and exposure to AI tools. The recruitment targeted roles directly involved in Requirements Engineering processes, including requirements engineers, business analysts, software developers participating in requirements-related tasks, and technical leads.

Interview participants. Interview participants were recruited using purposive sampling to select practitioners with direct experience in Requirements Engineering activities and familiarity with AI-assisted tools. The inclusion criteria for the interview phase were: (i) current or recent involvement in requirements elicitation, requirements analysis, product discovery, or closely related stakeholder-facing work; (ii) practical exposure to AI-supported tools in requirements-related or adjacent software engineering activities; and (iii) the ability to discuss concrete work practices and examples from industry settings. Practitioners were excluded if they had no substantive involvement in requirements-related tasks, no practical exposure to AI-supported tools, or only a purely academic/student perspective without recent industry practice. Recruitment was primarily conducted through professional networks, the researchers' personal networks, and social media platforms. Efforts were also made to include participants from diverse organizational contexts, such as companies of different sizes, organizations operating in consumer-oriented, business-oriented, and government-oriented contexts, as well as practitioners involved in international or cross-cultural collaborations. Purposive sampling is commonly used in qualitative research to obtain rich insights from participants with relevant domain expertise [7].

Survey participants. The questionnaire survey targeted a broader group of software engineering practitioners with exposure to requirements-related work. Participants were recruited through professional communities and online networks to capture a wider range of perspectives across different organizational contexts. Such broader recruitment is commonly used in the quantitative phase of mixed-methods studies to identify patterns across a larger participant group [9].

3.4 Stage 1: Qualitative Interviews and Analysis

The interview phase was designed to provide in-depth insights into practitioners' experiences with the use of artificial intelligence in Requirements Engineering activities, with particular attention to how roles, responsibilities, benefits, risks, and trust considerations are understood in practice.

3.4.1 Interview Instrument Design

The interview instrument was developed based on the research questions and prior literature on Requirements Engineering, AI-supported Requirements Engineering, and human–AI interaction. Overall, the interview guide was organized into five main sections. The first section collected participants' professional background, including their role, work experience, organizational context, RE-related responsibilities, and previous experience with AI tools. The second section focused on participants' current practices in requirements elicitation and requirements analysis, with AI-related questions embedded within these two parts. The third section examined task-level role allocation between humans and AI. Participants were asked to reflect on specific elicitation and analysis tasks mentioned earlier in the interview and to discuss how humans and AI should divide responsibilities in these tasks. This discussion was guided by the four role configurations defined in this study: human-only, AI-assisted, AI-led, and AI-only. When AI involvement was discussed, follow-up questions addressed tool choice, perceived strengths and weaknesses, integration with existing methods, benefits, and risks. The fourth section focused on trust and verification. Depending on the extent to which trust-related issues had already appeared in earlier answers, one to three questions were asked about trust in AI-generated results, reasons for trust or distrust, verification practices, output stability, and previous experiences with serious AI errors. The final section invited participants to reflect on future expectations for AI use in requirements elicitation and analysis, including differences between the two phases, tasks that should remain human-led, possible future AI capabilities, and concerns about deeper AI involvement. The complete interview guide contained 29 core guiding question areas, together with optional and conditional follow-up prompts. The full interview guide is provided in Appendix A.

More specifically, prior literature informed not only the general focus of the interviews, but also the design of specific question areas. Literature on requirements elicitation and analysis informed the questions about practitioners' current RE practices. For example, since requirements elicitation involves identifying, collecting, and consolidating needs from different stakeholders and information sources, the interview guide included questions about requirement sources, elicitation methods, the most frequently used methods, and the most time-consuming or critical elicitation tasks [1, 4]. Since requirements analysis involves interpreting, structuring, prioritizing, and resolving problems such as ambiguity, conflict, overlap, and inconsistency, the guide also included questions about how participants analyze requirement conflicts, overlaps, inconsistencies, prioritization, and task difficulty [4, 10].

Literature on AI and Generative AI in Requirements Engineering informed both the design of AI-related question areas and the way these questions were embedded in the elicitation and analysis sections of the interview guide. Existing studies discuss the potential of AI and LLM-based tools for tasks such as summarizing information, extracting requirements from textual input, organizing requirements, identifying ambiguities, detecting inconsistencies, and supporting prioritization or workflow automation [3, 25, 13, 11]. Based on these topics, the interview guide included questions asking whether participants used AI in requirements elicitation or requirements analysis, which tools they used, which specific tasks AI supported, why AI was or was not used, and what problems AI helped solve. These AI-related questions were not treated only as a separate technical tool section. Instead, they were connected to participants' descriptions of their actual elicitation and analysis practices, such as requirement sources, commonly used methods, major challenges, and task importance. Because participants had different levels of involvement in elicitation and analysis, three to five questions from each sub-section were selected and followed up depending on their answers. In addition, one question about the possibility of integrating AI into the broader RE workflow, from requirement collection to classification and analysis, was asked of all participants.

Literature on automation, human–AI interaction, and trust informed the questions about role allocation, human oversight, verification, and responsibility. Prior work on automation suggests that human-machine work should not be treated as a simple binary distinction between using or not using automation, but should instead be examined in terms of different degrees of human involvement, machine involvement, and decision control [5]. This perspective aligns with the focus of this study on four human–AI role configurations, namely human-only, AI-assisted, AI-led, and AI-only, and provides support for examining role allocation at the task level. Human–AI interaction and trust-related literature further informed follow-up questions about appropriate reliance, including what makes participants trust or distrust AI outputs, how they would verify AI-generated results, whether AI output quality is stable, and whether they had experienced serious AI errors [6, 15, 24]. These questions were used to understand how practitioners judge the suitability of AI involvement in different RE tasks and how they manage the risks associated with AI-supported work.

Before the main data collection, one pilot interview was conducted to evaluate the clarity, relevance, and flow of the interview guide. The pilot interview was used to check whether the questions were understandable, whether the distinction between requirements elicitation and requirements analysis was clear, whether the four role configurations were easy to distinguish, and whether the interview could be completed within the expected time. Based on observations from the pilot interview and subsequent discussion within the research team, several revisions were made. First, explanations of requirements elicitation and requirements analysis were added at the beginning of the corresponding section to ensure that participants interpreted the two concepts consistently. Second, the human–AI role allocation section was adjusted so that participants first discussed concrete tasks before selecting or explaining a suitable role configuration, rather than discussing AI roles only in general terms. Third, conditional follow-up prompts were added for participants with different AI usage experiences, including those who used AI, those who did not use AI, and those who considered possible future AI use. These revisions improved the clarity and flexibility of the interview instrument and strengthened its alignment with the research questions.

3.4.2 Interview Procedure

The qualitative data for this study were collected through interviews with software engineering practitioners. The interviews followed a semi-structured format guided by an interview protocol, allowing the researcher to explore predefined topics while also enabling participants to elaborate on their experiences and viewpoints. This interview format is widely used in exploratory qualitative research because it enables the collection of rich and detailed insights while maintaining a certain level of consistency across interviews [7].

The interviews focused on practitioners' experiences with the use of artificial intelligence in Requirements Engineering activities. Participants were invited to describe how AI tools are used in requirements-related tasks, how responsibilities between humans and AI are understood and allocated in practice, and how practitioners perceive potential risks or limitations when AI is used to support requirements work. These discussions were intended to capture practitioners' perspectives on how responsibilities between humans and AI are understood in early Requirements Engineering activities and to provide empirical insights for answering the research questions regarding the role of AI in Requirements Engineering and the allocation of responsibilities between human practitioners and AI tools.

The interview phase was initially planned to include 8–12 practitioners. In qualitative research, however, sample size is not determined by statistical representativeness, but by the richness and relevance of the information obtained. The concept of information power suggests that when the research aim is focused and participants have relevant professional experience, a relatively small interview sample can still provide sufficient insights [26]. Empirical studies on qualitative sample sizes also suggest that thematic saturation can often be reached with smaller samples when participants share relevant expertise and the topic is clearly defined [27].

In the completed study, six formal interviews were included in the final analysis, in addition to one pilot interview that was used only to refine the interview guide. Recruitment was stopped when the later interviews no longer produced substantively new codes related to task-level role configurations, perceived benefits, perceived risks, and trust considerations. Specifically, the fifth and sixth interviews primarily reinforced and confirmed codes and patterns that had already emerged from earlier interviews, without introducing new themes or sub-themes. While it is not possible to guarantee full saturation with six interviews, the focused scope of the study and the relevance of participants' professional experience supported the decision to proceed to the survey phase. Given the focused scope of the study and the relevance of the participants' professional experience, this was considered sufficient to support the thematic analysis.

All interviews were conducted online in order to accommodate geographically distributed practitioners and to make scheduling and recording more feasible across organizations. Each interview was conducted as a single semi-structured session following the same interview guide. Interview length varied depending on the depth of the participant's examples and reflections, but most interviews lasted approximately 1 to 1.5 hours and none exceeded 2 hours. With participants' consent, the interviews were audio recorded and transcribed verbatim into text. The resulting transcripts were reviewed against the recordings and anonymized before the coding process began.

3.4.3 Qualitative Data Analysis

The qualitative data collected through semi-structured interviews were analyzed using thematic analysis, following the approach proposed by Braun and Clarke [28]. This method was selected because of its flexibility and suitability for identifying patterns across participants' experiences in requirements elicitation and analysis tasks. As this study follows a sequential exploratory design, the qualitative findings were also used to inform the subsequent survey phase [7].

The analysis process was conducted in several iterative phases. First, all interview recordings were transcribed, and the transcripts were anonymized and reviewed to become familiar with the data. During this phase, initial observations were noted, especially those related to task execution, AI involvement, human responsibilities, perceived benefits and risks, and trust or verification practices.

Both researchers were involved in the coding process, but with different roles. One researcher first conducted the initial coding of the interview transcripts, focusing on meaningful segments related to RE tasks, AI involvement, human responsibilities, perceived benefits and risks, and trust or verification practices. After the initial coding was completed, both researchers reviewed and discussed the coding results together. Differences in code names, interpretations, or coding boundaries were discussed and used to refine the initial code list. Ambiguous or overlapping codes were further examined by checking the coded segments against the original transcript context. Through this process, the researchers reached agreement on how the codes should be interpreted and applied.

A formal inter-rater reliability statistic, such as Cohen's kappa, was not calculated, because the analysis was exploratory and interpretive rather than based on independent coding against a fixed coding scheme. Instead, coding consistency was supported through joint review of the coding results, repeated discussion between the researchers, refinement of the code list, and comparison of coded segments with the original interview context.

After the coding process, similar codes were grouped into candidate themes by identifying similarities and relationships between them. For example, codes related to AI-supported summarization, documentation, and information organization were grouped into themes concerning AI as a support tool in RE work. Codes related to human judgment, stakeholder communication, and final decision-making were grouped into themes concerning human responsibility and task-level role allocation. Codes related to verification, unstable outputs, trust, and concerns about errors were grouped into themes concerning trust and risk.

The candidate themes were then reviewed against the original interview data and the research questions. Themes related to task performance and responsibility distribution were used to address RQ1, while themes related to benefits, risks, trust, and verification were used to address RQ2. This process helped ensure that the final themes remained grounded in the empirical data while also being aligned with the purpose of the study.

3.5 Stage 2: Quantitative Data Collection and Analysis

Following the interview phase, a survey was conducted to collect quantitative data from a broader population of software engineering practitioners. The purpose of the survey was to complement the qualitative findings obtained from the interviews and to examine whether the perspectives and patterns identified during the interview phase also appear among a wider group of practitioners. The survey also aimed to explore whether differences exist across participants with different professional roles or experience levels.

3.5.1 Survey Instrument Design

The survey questionnaire was designed after the interview phase and was informed by both the research questions and the qualitative findings. Since this study follows a sequential exploratory mixed-methods design, the survey was used to extend the interview findings to a broader group of practitioners and to examine whether the patterns identified in the interviews were also reflected in a wider sample [7, 9]. In particular, themes related to task-level AI involvement, human responsibility, perceived benefits and risks, trust, and verification practices were translated into structured survey items.

The transformation from interview findings to survey items followed a step-by-step process. First, the interview codes and candidate themes were reviewed to identify recurring issues across participants, such as AI use for summarization and documentation, AI support for structuring requirements, concerns about missing context or unstable outputs, the need for human verification, and the view that human judgment remains important in sensitive or decision-related tasks. Second, these themes were grouped according to the two research questions. Themes related to task execution and responsibility distribution informed the role allocation items, while themes related to efficiency, documentation support, risks, trust, and verification informed the perception-based items. Third, the themes were operationalized into different survey question formats. For example, interview insights about AI-supported summarization, extraction, ambiguity identification, and conflict detection were transformed into the task-use question in Q6 and the role allocation matrices in Q7a and Q7b. Interview insights about time saving, better documentation, improved consistency, faster draft generation, better organization of dispersed information, and reduced workload were transformed into the benefit rating matrix in Q9. Concerns about hallucination, loss of context, over-reliance, confidentiality, unclear responsibility, inconsistent results, and biased assumptions were transformed into the risk rating matrix in Q10. Finally, interview findings about trust, reliability, and verification practices informed Q8, Q11, Q12a, Q12b, and Q13. In this way, the survey was designed to operationalize the qualitative findings into structured items that could be answered by a broader group of practitioners.

The survey instrument was organized into several sections. The first section collected participants' professional background, including their primary role in software development, years of professional experience, organization type, and involvement in different Requirements Engineering (RE) stages. The RE stages included requirements elicitation, requirements analysis, requirements specification, and requirements validation. This section was used to characterize the respondents and to ensure that the survey reached practitioners with experience relevant to RE-related work.

The second section asked whether participants used AI tools, such as Generative AI, in requirements elicitation or requirements analysis tasks. Participants who reported using AI, either regularly or occasionally, were asked to specify the tools they used and to indicate the tasks for which they used AI. These tasks included summarizing meetings or documentation, extracting requirements from raw data, identifying ambiguities or missing information, structuring or organizing requirements, detecting conflicts or inconsistencies, supporting background research or information search, supporting prioritization or decision-making, and breaking down requirements into tasks. Participants who reported that they did not use AI were directed to follow-up questions about their reasons for not using AI and their general perception of AI-generated outputs in requirements-related tasks.

The third section focused on task-level role allocation between humans and AI. Before answering this section, participants were given definitions of the four role configurations used in this study: human-only, AI-assisted, AI-led, and AI-only. They were then asked to select suitable role configurations for specific requirements elicitation and requirements analysis tasks. The elicitation tasks included summarizing elicitation sessions, extracting requirements from raw inputs, identifying ambiguities or missing information, and suggesting clarification questions. The analysis tasks included structuring and standardizing requirements, structuring and organizing requirements, detecting conflicts or inconsistencies, and supporting prioritization or trade-off decisions. This section was designed to address RQ1 by examining how practitioners distribute responsibilities between humans and AI across different early RE tasks.

The fourth section examined perceived benefits, risks, and trust considerations related to AI involvement in early RE tasks. These items were informed by the interview findings, where participants discussed issues such as time saving, documentation support, improved consistency, faster draft generation, organization of dispersed information, better coverage, ideation support, reduced workload, and faster information access. To capture perceived benefits, the survey used a five-point rating matrix, where 1 indicated "No benefit" and 5 indicated "Strong benefit". Participants rated how strongly each benefit applied under the four role configurations: human-only, AI-assisted, AI-led, and AI-only.

A similar matrix format was used to examine perceived risks. Participants rated how likely each risk was to occur under each role configuration, using a five-point scale where 1 indicated "Very unlikely" and 5 indicated "Very likely". The risk items included misinterpretation of stakeholder intent, hallucinated or fabricated details, loss of context or nuance, difficulty in understanding implicit requirements or unexpressed stakeholder intentions, reduced quality of stakeholder communication, over-reliance, confidentiality or data leakage concerns, lack of explainability, accountability ambiguity, inconsistent results, bias or unfair assumptions, and no major risk. This design allowed the survey to compare how perceived benefits and risks varied across different human–AI role configurations.

The trust and verification part of the survey included several types of items. Participants were asked under which conditions they considered AI outputs reliable, how they typically used AI-generated outputs, in which situations they verified AI outputs, and how they typically verified them. These questions used single-choice and multiple-choice formats. In addition, the survey included a six-point agreement scale ranging from "Strongly disagree" to "Strongly agree". This scale was used to measure agreement with statements about trust in AI for documentation-related tasks, requirements analysis tasks, decision-making processes, verification before use, confidence in using AI outputs, perceived reliability, possible inaccuracy or inconsistency, reliance on human judgment in critical decisions, and whether AI is better suited as a supporting tool than as an independent actor. The six-point format did not include a neutral midpoint, in order to encourage participants to indicate a direction of agreement or disagreement.

The survey included a combination of question types. Single-choice questions were used for demographic information, AI usage screening, and usual practices when using AI-generated outputs. Multiple-choice questions were used for RE stage involvement, AI-supported tasks, reasons for not using AI, conditions for reliability, verification situations, and verification methods. Matrix-style questions were used for task-level role allocation and for rating benefits and risks across the four role configurations. Open-ended questions were included to allow participants to explain why they chose not to use AI in some tasks and whether they had experienced problems when using AI. Conditional logic was implemented in Qualtrics so that participants received questions relevant to their AI usage experience. The questionnaire contained 15 numbered questions, including sub-questions and optional open-ended items. It is provided in Appendix B.

Before the survey was distributed, a pilot test of the questionnaire was conducted to evaluate its clarity, structure, and technical functionality in Qualtrics. The pilot test was used to check whether the questions were understandable, whether the four role configuration definitions were clear, whether the benefit and risk rating scales were easy to interpret, whether the six-point agreement scale was understandable, and whether the conditional logic directed participants to the appropriate follow-up questions. Based on the pilot feedback, several revisions were made. Some question wording was clarified, the explanations of human-only, AI-assisted, AI-led, and AI-only were refined, and the order of some items was adjusted to improve the flow of the questionnaire. The conditional logic was also checked to ensure that participants who did not use AI were directed to the relevant non-user questions, while AI users continued to the task-level and perception-related sections. These revisions helped ensure that the final survey instrument was clearer, more consistent, and better aligned with the research questions.

3.5.2 Survey Procedure

The survey targeted software industry practitioners who were involved in Requirements Engineering (RE) or RE-related software development activities, particularly requirements elicitation and requirements analysis. The intended participants included requirements engineers, product managers or product owners, software developers, project managers, business or system analysts, UX designers, architects, and QA or test engineers, as these roles may participate in collecting, interpreting, structuring, validating, or communicating requirements in practice.

The questionnaire was implemented and distributed online using Qualtrics. Recruitment was conducted through several channels, including LinkedIn, Facebook groups related to software engineering, project management, and product management, the researchers' personal and professional networks, and direct outreach to practitioners whose public profiles or project experience suggested possible involvement in RE-related work. On LinkedIn, the survey was shared through posts and direct messages to practitioners in relevant software development or requirements-related roles. In Facebook groups, the researchers approached potential participants in groups related to software engineering, project management, and product management, where members' profiles or posts indicated possible relevance to software development or RE-related work. In addition, the researchers contacted personal and professional connections with experience in software development, product work, project management, or requirements-related activities.

The invitation message briefly introduced the study topic, explained that participation was voluntary and anonymous, and included the survey link. The initial target was to collect approximately 50 to 100 responses from practitioners with software development or RE-related experience. The survey was open from April 10 to May 10, 2026. During this period, 73 responses were collected in total. After checking for incomplete or invalid responses, 67 responses were retained for analysis. Responses were considered valid if participants completed the key survey sections and had relevant software development or RE-related experience.

The final survey sample was used to provide descriptive insights into practitioners' views on AI involvement, human–AI role allocation, perceived benefits and risks, and trust considerations in early RE tasks. Since the survey was distributed through the researchers' personal contacts, professional networks, and online communities rather than probabilistic sampling, the results are interpreted as exploratory and descriptive rather than statistically representative of the wider software industry population.

3.5.3 Quantitative Data Analysis

The quantitative data collected through the survey were analyzed using descriptive statistics. The survey included several types of questions, including participant background information, multiple-choice items, task-based role configuration matrices, and Likert-scale evaluations. Descriptive analysis was selected because the survey was exploratory in purpose, used non-probability recruitment, and was intended to summarize patterns within the respondent group rather than to support population-level inference or formal hypothesis testing.

Inferential statistical procedures were not used. More specifically, the study did not apply tests such as chi-square tests, t-tests, ANOVA, regression models, p-values, or confidence intervals, because these techniques are primarily used to test group differences, associations between variables, or to draw inferences from a sample to a wider population. That was not the purpose of this survey. The quantitative phase was designed to describe how the 67 valid respondents answered the questionnaire and to complement the interview findings with broader response patterns, not to estimate population parameters for the wider software industry.

Several methodological considerations further supported this choice. First, the respondents were recruited through non-probability channels, including professional networks, online communities, and direct outreach, so the sample cannot be treated as statistically representative of a defined population. Under these conditions, inferential outputs such as p-values and confidence intervals could easily suggest a degree of precision and generalizability that the sampling strategy does not justify. Second, several potentially interesting subgroup comparisons, for example by professional role or AI use pattern, would involve small and uneven cell sizes, making significance tests unstable and difficult to interpret. Third, many survey items were categorical, multiple-choice, or ordinal Likert-type measures designed to capture exploratory judgments about task-level role configurations, perceived benefits, risks, and trust, rather than to test a set of pre-specified explanatory hypotheses. Conducting multiple post hoc inferential tests across these items would therefore have increased the risk of over-interpreting incidental differences in a dataset that was not designed for confirmatory analysis.

For participant background and multiple-choice questions, frequency counts and response distributions were used to summarize participants' professional roles, years of experience, organizational contexts, and reported AI usage practices. For task-based role configuration questions, response patterns were analyzed to examine how different roles (human-only, AI-assisted, AI-led, AI-only) are distributed across requirements-related tasks.

Likert-scale responses related to benefits, risks, and trust were analyzed by examining the distribution of responses across categories. These response distributions were used to identify overall tendencies in participants' perceptions.

For Q9 and Q10, mean values were also used as descriptive summaries of relative tendencies across role configurations and survey items, including visual comparison through heat maps. Because these Likert-type ratings are ordinal, the mean values were not used for inferential testing or treated as precise interval-scale measures, but only as descriptive indicators within the sample. For Q13, the response distributions were visualized using a diverging stacked bar chart in order to preserve the distributional structure of the agreement scale and to show how agreement and disagreement were distributed across the trust-related items.

Missing survey data were handled on an item-by-item basis. If a respondent skipped a question, that response was treated as missing and excluded only from the analysis of that specific item, while the respondent's remaining answers were retained for all other analyses. No imputation of missing values was performed; instead, percentages were calculated using the number of valid responses for each item.

The quantitative analysis aims to complement the qualitative findings by providing a broader overview of patterns across participants and supporting the interpretation of task-level role configurations and associated perceptions.

3.5.4 Ethical Considerations

Ethical considerations were carefully addressed throughout the data collection process. Before the interviews or the survey began, all participants were informed about the purpose of the study, the voluntary nature of their participation, and their right to withdraw at any time. Informed consent was obtained from all participants before their participation in the study.

To protect participants' privacy, all collected data were anonymized during transcription and analysis and were used only for the purposes of this research. Interview recordings and transcripts were stored securely and were accessible only to the research team.

3.6 Threats to Validity

This study is subject to several threats to validity that should be considered when interpreting the findings. Following common empirical software engineering practice, the threats are discussed in terms of construct validity, internal validity, external validity, conclusion validity, and reliability [29].

3.6.1 Construct Validity

Construct validity concerns whether the key concepts in this study, such as human–AI role allocation, task-level AI involvement, perceived benefits, risks, trust, and verification practices, were appropriately defined and measured. These constructs may be interpreted differently by different practitioners. A specific concern is the interpretation of the role configuration categories. For example, "AI-led" may have meant an AI-generated first draft to some participants and a stronger form of autonomy to others. Similarly, "AI-assisted" may cover a wide range of practices, from simple text polishing to substantial analytical support. Participants may also understand trust, responsibility, and risk differently depending on their organizational context and previous experience with AI tools.

To reduce this threat, the interview and survey instruments were grounded in prior literature on Requirements Engineering, AI-supported RE, automation, human–AI interaction, and trust. This helped ensure that the key constructs were not defined only based on the researchers' assumptions, but were linked to established concepts in related research. For example, literature on automation and human–AI interaction informed the distinction between different levels of AI involvement, while trust-related literature informed questions about verification, reliability, and appropriate reliance.

The four role configurations used in this study, namely human-only, AI-assisted, AI-led, and AI-only, were explicitly defined in the survey before participants answered the role allocation questions. The survey also asked about specific elicitation and analysis tasks rather than only asking about AI use in general. This helped connect the constructs to concrete RE activities. In addition, both the interview guide and the survey questionnaire were piloted and refined to improve clarity and reduce potential misunderstanding. Nevertheless, future studies could use more concrete scenarios or examples to define role configurations more precisely.

Another survey design-related threat concerns the interpretation of the role-allocation items in Q7. Although the items were intended to capture respondents' selected role configuration for each task, some respondents may have interpreted the question more broadly as asking which configurations were possible or acceptable for a given task rather than which configuration or configurations they perceived as most suitable. If such differences in interpretation occurred, the resulting percentages should be understood as approximate descriptive tendencies in perceived role allocation rather than as precise measurements of fixed preference.

Another construct validity threat is that the survey instrument was developed from themes identified in the interview phase. This design supports alignment between the qualitative and quantitative phases, but it may also mean that issues not surfaced in the interviews were underrepresented in the questionnaire. To partly mitigate this limitation, the interviews ended with an open question asking participants whether there was anything important that had not been covered but that they wanted to add. This allowed participants to raise additional issues beyond the planned interview questions. In addition, the survey included optional open-ended questions, allowing participants to provide explanations, examples, and concerns beyond the predefined response options.

A related measurement issue concerns the constructs of trust, benefits, and risks. These concepts are complex and context-dependent. Likert-scale items can capture general tendencies, but they may not fully represent how practitioners make trust decisions in real work situations. To address this limitation, the study combined Likert-scale results with open-ended responses and interview findings, allowing the interpretation to include both numerical patterns and participants' reasoning.

3.6.2 Internal Validity

Internal validity is mainly related to the credibility of the interpretations drawn from participants' self-reported accounts. Since this study is exploratory and does not aim to test causal relationships, internal validity threats are mainly related to interpretation rather than causal inference. One threat is that participants' responses were self-reported. Their descriptions of AI use, trust, verification, and risk may be affected by recall bias, selective reporting, or social desirability bias. For instance, participants may overstate how carefully they verify AI-generated outputs or may simplify how AI is actually used in their daily work.

To reduce this threat, both interviews and surveys were designed around concrete tasks and work practices. Instead of asking only about general opinions toward AI, participants were asked about specific RE activities such as summarizing elicitation sessions, extracting requirements, identifying ambiguities, detecting inconsistencies, and supporting prioritization. This task-oriented design helped make the responses more grounded in actual work situations. The use of both qualitative interviews and survey data also allowed the study to compare detailed individual accounts with broader response patterns.

3.6.3 External Validity

External validity is mainly related to the extent to which the findings can be transferred to other software industry contexts. The qualitative phase was based on a small purposive interview sample, and the survey relied on non-probability recruitment. Survey participants were recruited through the researchers' personal contacts, professional networks, LinkedIn, Facebook groups related to software engineering, project management, and product management, and direct outreach to practitioners with potentially relevant experience. Therefore, the sample cannot be considered statistically representative of the wider software industry population.

The survey collected 73 responses in total, of which 67 were retained for analysis after checking for incomplete or invalid responses. Although this provided a broader empirical basis than the interview phase alone, the findings should still be interpreted as exploratory and descriptive. The results indicate patterns among the participants reached by this study, rather than population-level claims about all Requirements Engineering practitioners. In addition, the sample included more Product Managers/Product Owners, Software Developers, and Business Analysts/System Analysts than dedicated Requirements Engineers. The findings should therefore be read as practitioner-informed evidence about current role perceptions, rather than as a definitive measurement of all RE contexts. To support transferability, the study reports participants' roles, experience, organization types, and

RE-related involvement, so that readers can judge the relevance of the findings to other contexts.

A further external validity threat concerns the current maturity and availability of AI tools. The low preference for AI-only configurations may partly reflect the tools that participants knew or had used, rather than a permanent boundary of AI involvement in early RE. If future systems provide stronger context retention, better traceability, secure integration with internal knowledge bases, and clearer explanations, practitioners may become more willing to let AI lead some preliminary tasks. However, the findings still suggest that human review is needed when tasks involve stakeholder intent, trade-offs, and accountability.

3.6.4 Conclusion Validity

Conclusion validity is mainly related to whether the conclusions remain within what the qualitative and descriptive survey data can reasonably support. In this study, the main conclusion-related threat is the risk of over-interpreting exploratory and descriptive findings. The interview phase provided contextual explanations of practitioners' reasoning, while the survey phase described broader tendencies in participants' role allocation choices and perceptions across different RE tasks. However, the study was not designed to test causal relationships or support statistical generalization to the entire software industry.

Because the quantitative analysis used descriptive rather than inferential statistics, percentages and mean values should be interpreted as descriptive indicators of relative tendencies in the sample, not as statistically tested differences, causal relationships, or estimates of population-level effects. In particular, the study does not claim that apparent differences between role configurations, tasks, or respondent subgroups are statistically significant. Such claims would have required a different survey design, including probability-based sampling or a clearly bounded sampling frame, stronger statistical power for subgroup comparisons, and pre-specified inferential hypotheses. In the absence of these conditions, inferential outputs such as p-values or confidence intervals would risk overstating what the dataset can support. The reported comparisons were therefore used only to support cautious interpretation together with interview and open-ended survey findings.

An additional threat related to the conclusions concerns the interpretation of the role-allocation percentages derived from Q7. If some respondents understood the role-allocation items as asking for acceptable or possible configurations rather than their preferred single configuration, the resulting percentages may overstate the stability of the observed differences between the role categories. For this reason, the role-allocation results are interpreted as descriptive indications of response patterns rather than as exact measurements of exclusive preference.

To reduce this threat, the study avoids making claims beyond what the data can support. Survey results are interpreted together with the interview findings, rather than as standalone evidence of broader population effects. This allows the study not only to describe which role configurations practitioners tended to prefer for different tasks, but also to interpret possible reasons behind these preferences based on the qualitative data. The mixed-methods design therefore provides both contextual depth and broader descriptive evidence. Conclusions are drawn by considering both phases together, with the findings framed as exploratory insights into practitioner perceptions, task-level role allocation tendencies, and future expectations for human–AI collaboration in early RE tasks.

3.6.5 Reliability

Reliability is mainly related to the transparency and consistency of the data collection and analysis process, especially because the qualitative analysis involved interpretive coding and theme development. In qualitative analysis, a key reliability-related threat is that thematic analysis is interpretive by nature and can be influenced by researcher judgment. In this study, one researcher first conducted the initial coding of the interview transcripts. After the initial coding was completed, both researchers reviewed and discussed the coding results together. Points requiring clarification or adjustment, such as code names, interpretations, or coding boundaries, were discussed and used to refine the code list. Ambiguous or overlapping codes were further examined by checking the coded segments against the original transcript context.

A formal inter-rater reliability statistic, such as Cohen's kappa, was not calculated, because the analysis was exploratory and interpretive rather than based on independent coding against a fixed coding scheme. Instead, coding consistency was supported through joint review of the coding results, repeated discussion between the researchers, refinement of the code list, and comparison of coded segments with the original interview context. The use of an interview guide, survey instrument, documented coding process, and appendices containing the research instruments also improves the transparency and traceability of the study procedure.

3.7 Disclosure of AI Use

AI-based tools were used in a limited and supportive way during the preparation of this thesis, primarily for language refinement, rephrasing, and improving the clarity and structure of draft text. They were not used as a substitute for the authors' own literature reading, empirical analysis, interpretation of findings, or final academic judgment. All research design decisions, data collection, coding, analysis, interpretation, and conclusions remained the responsibility of the authors.

The use of AI in thesis writing may introduce risks such as oversimplification, inaccurate wording, unsupported statements, or phrasing that appears academically polished without being sufficiently grounded in the actual study. To reduce these risks, any AI-assisted output was reviewed, revised, and validated by the authors before inclusion in the thesis. AI-generated suggestions were treated only as drafting support, not as evidence or authoritative content. The authors therefore retain full responsibility for the accuracy, integrity, and final presentation of the thesis.

4

Results

4.1 Introduction

This chapter presents the empirical findings from the interview and survey phases. The results are organized to address the two research questions of this study. RQ1 focuses on how roles and responsibilities between humans and AI are delegated in requirements elicitation and requirements analysis tasks, while RQ2 focuses on the perceived benefits, risks, and trust considerations associated with different human–AI role configurations.

The chapter first presents the qualitative findings from the semi-structured interviews. These findings provide detailed insights into practitioners' experiences, reasoning, and perceptions regarding AI involvement in early Requirements Engineering (RE) work. The interview results are organized according to the themes identified through thematic analysis and linked to the research questions, with illustrative examples or quotations from participants where relevant.

The chapter then presents the survey results. The survey findings provide a broader descriptive view of practitioners' role allocation choices, perceived benefits and risks, and trust-related attitudes across different RE tasks and role configurations within the respondent group. The survey results are reported mainly through descriptive statistics, as the purpose of the survey was to describe response patterns rather than to test causal relationships or make population-level statistical claims.

Finally, the chapter integrates the interview and survey findings by examining areas of convergence, extension, and divergence between the two phases. This integration shows how the qualitative and quantitative findings together support, refine, or extend the interpretation of human–AI role allocation in early RE. The relationship between these findings, the research questions, and prior literature is further discussed in the following discussion chapter.

4.2 Interview Results

4.2.1 Participants

A total of seven participants were involved in this study, including one pilot interview conducted to refine the interview protocol. Data from the pilot interview were not included in the final analysis. The results presented are therefore based on six participants (P1–P6).

The participants represented diverse professional roles within software development and related domains, including technical, business-oriented, and managerial positions. This diversity allowed the study to capture perspectives from different levels of organizational responsibility and areas of expertise. Participants had between approximately 7 and 21 years of professional experience, indicating a relatively experienced sample. They were employed in organizations of varying sizes, including both small and medium-sized enterprises (SMEs) and large companies. Participants' specific roles, years of experience, and organizational context are summarized in Table 4.1.

Table 4.1: Participant Overview

Participant	Role	Years of Experience	Organization Size
P1	Full Stack Software Developer	∼15 years	SME
P2	Content Manager	∼8 years	SME
P3	Product Manager	∼7 years	Large company
P4	Industry Director	∼7 years	Large company
P5	Project Manager	∼7 years	Large company
P6	Project Manager	∼21 years	Large company

4.2.2 Thematic Analysis of Interview Data

This section reports the outcome of the thematic analysis conducted on the six formal semi-structured interviews included in the study. Following the approach outlined in Chapter 3, the interview recordings were transcribed and read repeatedly to achieve familiarity with the data before the analytical coding process began.

The transcripts were then coded with attention to three recurring aspects of participants' accounts: how AI was used in early Requirements Engineering tasks, which responsibilities remained human-led, and what benefits, risks, or trust concerns participants associated with AI involvement. Related codes were clustered into candidate themes and sub-themes and then iteratively reviewed against the full dataset to ensure internal coherence and clear distinctions between themes. The final thematic structure was organized around the two research questions. RQ1 captures how roles and responsibilities are allocated between humans and AI, whereas RQ2 captures how participants evaluate these configurations in terms of benefits, risks, and trust. The following sections present the resulting themes and sub-themes together with illustrative quotations from participants.

Themes and sub-themes related to RQ1

For RQ1, the analysis shows that participants distinguish clearly between information-oriented support activities, stakeholder-facing elicitation work, and requirements analysis. *Across these activities, AI is primarily described as a supportive resource rather than an autonomous actor. At the same time, the more a task depends on contextual interpretation, negotiation, or responsibility for decisions, the more strongly participants position humans at the centre of the work.*

I. **AI as a support tool for information-related tasks**

1. **AI supporting information-gathering tasks:** This sub-theme captures the role of AI in preparatory work, such as searching for domain-related terminology, clarifying unfamiliar concepts, or helping practitioners form an initial understanding before validating it with stakeholders. In such tasks, AI mainly serves a supportive preparatory role rather than directly participating in elicitation itself.

   Participant P1: *"We basically use it as a search engine to ask it fixed questions, for example professional terminology."*

2. **AI supporting documentation and summarization tasks:** This sub-theme shows that AI is also used to process already collected information, for example by extracting key points from meetings, generating lists of questions, or helping organize notes and documents. In tasks primarily concerned with organizing existing material, the advantages of AI can be leveraged more fully, while the additional burden placed on practitioners remains relatively low. As a result, compared with tasks that require independent interpretation of stakeholder intent, AI can be more naturally integrated into this kind of workflow.

   Participant P2: *"Now, in meetings, everything is basically recorded, and then AI extracts the key points..."*

II. **Elicitation driven by human interaction:**
This theme highlights that stakeholder communication is regarded as the core of elicitation work. Participants pointed out that elicitation is not only about collecting statements, but also about negotiating meaning, understanding what stakeholders actually need, and responding flexibly during interaction. Accordingly, humans remain primarily responsible for this part of the work.

Participant P1: *"It is still mainly a job centred on dealing with people."*

III. **Conditional AI involvement in requirements analysis**

1. **AI-assisted requirements analysis:** This sub-theme captures situations in which AI supports analysis-related activities, such as identifying duplicate or conflicting requirements, comparing different pieces of content, or helping uncover issues that may otherwise be overlooked in existing material. These uses suggest that, in tasks involving larger amounts of documents or textual information, AI can participate more directly in analytical work and provide support for subsequent human judgment.

   Participant P6: *"If you give it a document, the large model can very quickly identify duplicate parts and conflicting parts for you."*

2. **Human judgment retained:** This sub-theme shows that even when AI is involved in requirements analysis, final interpretation and decision-making remain the responsibility of humans. Participants emphasized that determining which requirements are valid, handling uncertainty, and deciding how to act on analytical results still require human control.

   Participant P3: *"Once it generates a text like that, they may use AI to polish it, but they would not use AI to make this kind of decision, such as whether this is something they actually need."*

3. **Task-contingent role allocation:** This sub-theme emphasizes that the allocation of roles between humans and AI depends on the characteristics of the task itself. AI was regarded as more suitable for repetitive, high-volume, and lower-complexity analytical tasks, whereas humans remained indispensable in tasks requiring deeper contextual understanding, customization, or sensitivity to stakeholder concerns. This suggests that role allocation is determined not only by the phase of work, but also by the specific demands of the task.

   Participant P4: *"The larger and heavier the project is, and the more customized it is, the greater the human role becomes. The smaller and more standardized the project is, the smaller the human role becomes."*

Overall, the themes related to RQ1 indicate that ***role allocation between humans and AI is shaped both by the nature of the task and by the phase of early Requirements Engineering in which the task occurs.*** AI is mainly positioned as a support tool in information-oriented work and in some document-heavy analytical activities, whereas ***humans remain central in elicitation and in those analytical tasks that require contextual interpretation, stakeholder interaction, and accountability***. This thematic structure is summarized in Figure 4.1, which also marks the evolving role of AI in Requirements Engineering as a cross-cutting theme relevant to both RQ1 and RQ2.

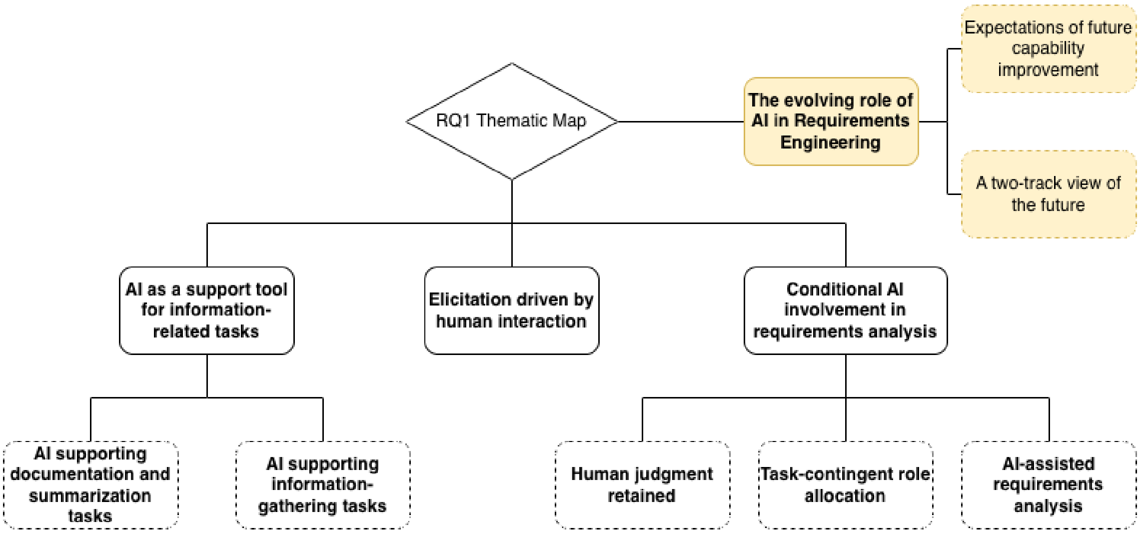

Figure 4.1: Themes and sub-themes for RQ1: *These themes capture different aspects of role allocation between humans and AI in early RE tasks. The yellow-highlighted theme and sub-themes are relevant to both RQ1 and RQ2 and are therefore discussed further in the subsequent section.*

Themes and sub-themes related to RQ2

For RQ2, the thematic analysis identified a second set of themes concerning the perceived benefits, risks, and trust considerations of using AI in requirements elicitation and requirements analysis tasks. Participants did not evaluate AI in uniformly positive or negative terms. The perceived benefits were mainly related to time savings, task support, and reduced manual effort, whereas the perceived risks were mainly associated with limited contextual understanding, unreliable or misleading outputs, hallucination risks, and concerns about data security and information exposure. Trust was therefore not treated as a general attitude toward AI, but as something shaped by whether AI outputs could be verified, controlled, and used safely in the elicitation and analysis tasks discussed by participants.

I. **Barriers to integrating AI into requirements work**

1. **Technical and contextual barriers:** This sub-theme reflects participants' concerns about the limitations of AI in understanding requirements-related context during elicitation and analysis. Participants suggested that AI could support text-level processing, such as organizing or summarizing requirements-related material, but that it was less capable of forming a comprehensive understanding of requirements in relation to business context, stakeholder needs, and project-specific constraints. Therefore, AI was perceived as useful for bounded and text-based support, while human judgment remained necessary for more contextual and interpretive requirements work.

   Participant P5: *"Human judgment in requirements is more comprehensive... currently, AI in my work mainly operates at the text level."*

2. **Organizational barriers:** This sub-theme indicates that AI use in requirements work was not only shaped by individual willingness, but also by whether AI had been formally integrated into organizational workflows. Participants suggested that AI was often used in an ad hoc and individual-driven way, rather than as a standardized part of team-level or process-level requirements work. This limited the extent to which AI could become a consistent component of elicitation and analysis practices.

   Participant P3: *"At each stage, individuals may use AI when needed... it is more of a personal choice."*

3. **Data security-related barriers:** This sub-theme emphasizes that data security and information exposure concerns limited participants' willingness to use AI in requirements work. When early RE tasks involved internal documents, customer information, or sensitive business content, participants were cautious about using external AI tools. Some participants also suggested that AI use would be more acceptable if organizations provided internal or private AI models, because this could reduce the risk of exposing confidential information.

   Participant P6: *"If AI cannot be used, it is most likely due to security concerns... unless the company has an internal private model."*

II. **AI reliability and trust management**

1. **Trust depends on task criticality and output verifiability:** This sub-theme indicates that participants' trust in AI was strongly conditioned by the type of task being performed. AI-generated outputs were seen as more acceptable in low-risk tasks where results could be checked quickly, revised easily, or treated as provisional drafts. In contrast, trust dropped when tasks involved implicit stakeholder needs, strategic interpretation, or decisions with broader project consequences.

   Participant P2: *"The trust level in its outputs is about 60–70%, mainly because summarisation tasks are relatively reliable."*

2. **Human verification remains a condition for practical use:** This sub-theme reflects that participants treated AI outputs as useful but not self-sufficient. Even when AI-based cross-checking could reduce some errors, participants still considered human review necessary before the results were used in practice. This was especially important in requirements-related work, where inaccurate interpretations or small misunderstandings could influence later decisions.

   Participant P5: *"When using multiple AI systems, cross-checking between them can eliminate about 80% of issues... but manual review is still re-*

quired."

3. **Trust is calibrated through bounded and repeated use:** This sub-theme suggests that participants' trust in AI was shaped by both repeated practical experience and the boundaries placed around its use. Rather than trusting AI outputs in general, participants were more willing to rely on AI when they had used it in similar tasks before, when the task scope was clearly defined, and when the system was constrained by a specific knowledge base or available source material. In these situations, AI outputs were perceived as more reliable because they could be linked to a controlled information context, rather than generated from an open-ended or unclear basis.

   Participant P2: *"If you constrain it with a specific knowledge base... the outputs are generally reliable."*

III. **Benefits of AI are concentrated in efficiency-oriented support**

1. **Efficiency gains in routine and preparatory tasks:** This sub-theme captures participants' perception that AI is particularly valuable for repetitive, high-volume, and relatively low-risk work. Participants associated AI's benefits with reducing manual effort in tasks such as producing summaries, generating draft material, organizing notes, or preparing initial structures. In these situations, AI was viewed as a tool for improving efficiency rather than as a system that fundamentally changed decision authority in requirements work.

   Participant P5: *"It's my tool... for tasks that are highly repetitive, large in volume, but low in technical difficulty."*

2. **Analytical support for specific requirements analysis tasks:** This sub-theme highlights that AI was perceived as beneficial in specific analytical tasks where its capability could be clearly observed, such as identifying requirement conflicts or supporting structured comparison. However, this benefit was not interpreted as general trust in AI across all analysis tasks, but as task-specific usefulness.

   Participant P6: *"In this aspect, AI is indeed better than humans in many cases... especially in identifying requirement conflicts."*

3. **Benefits are strongest in bounded and controlled contexts:** This sub-theme emphasizes that participants perceived AI's benefits as dependent on whether its risks could be managed in practice. Although hallucinated or unreliable outputs were recognized as a limitation, participants did not treat these risks as a reason to reject AI altogether. Instead, AI was considered practically beneficial when its outputs could be checked,

> filtered, and selectively used. In such bounded and controlled contexts, the perceived benefits of AI were strong enough for participants to continue using it despite known risks, as long as problematic parts could be identified and human judgment remained involved.
>
> Participant P6: *"If you can clearly identify which parts are hallucinated... you can just ignore them directly."*
>
> Participant P6: *"We do not trust it blindly... we use AI more cautiously rather than abandoning it entirely."*

Overall, the themes related to RQ2 show that practitioners evaluate AI through a conditional logic of usefulness. AI is considered beneficial when it improves efficiency, supports document-heavy work, and operates within tasks whose outputs can be checked and revised. At the same time, trust in AI remains limited and highly situational. The more a task depends on contextual understanding, interpretation of implicit needs, or consequential decision-making, the more participants emphasized the need for human verification, human responsibility, and cautious use. This thematic structure is summarized in Figure 4.2, which also marks the evolving role of AI in Requirements Engineering as a cross-cutting theme relevant to both RQ1 and RQ2.

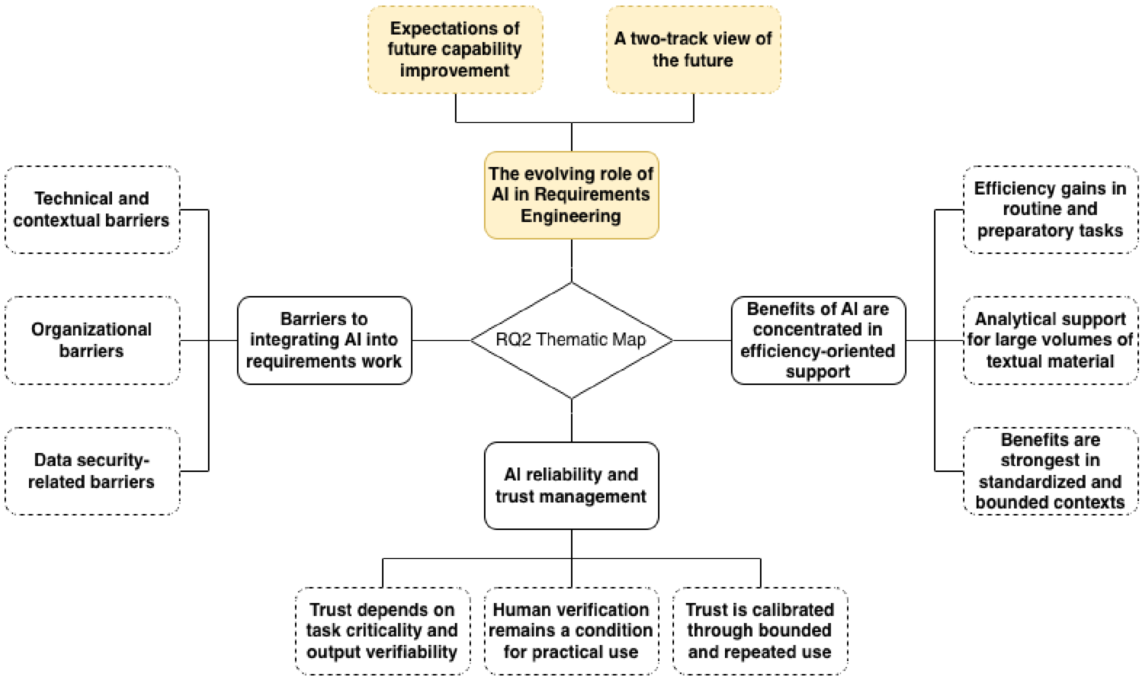

Figure 4.2: Themes and sub-themes for RQ2: *These themes capture the perceived benefits, risks, and trust implications of using AI in early RE tasks. The yellow-highlighted theme and sub-themes are relevant to both RQ1 and RQ2 and are therefore discussed further in the subsequent section.*

Themes and sub-themes related to both RQ1 and RQ2

In addition to the themes that correspond separately to RQ1 and RQ2, the interview data also revealed one theme that is simultaneously relevant to both research questions: the evolving role of AI in Requirements Engineering. This theme was separated out because it concerns not only possible future changes in the allocation of roles between humans and AI, but also participants' broader evaluations of AI capability development, applicability, risks, and trustworthiness. In other words, participants' views of future AI roles were not expressed as abstract technological optimism. Rather, they were grounded in current assessments of where AI is already useful, where its limitations remain visible, and under what conditions its role might reasonably expand.

I. **The evolving role of AI in Requirements Engineering**

1. **Expectations of future capability improvement:** This sub-theme shows that some participants believed AI could become more deeply involved in future early RE work if its capabilities improve further, particularly in areas such as responsiveness, continuous follow-up, context retention, and information integration. This expectation does not imply that AI would simply take on a larger role in all tasks. Rather, participants described future role expansion as conditional: AI may become more involved in certain tasks as its capabilities improve, but such involvement would still be bounded by task type, contextual demands, and practical limits of use.

   Participant P2: *"If AI could respond in real time 24 hours a day and continuously follow the project, it would be more likely to replace human labour."*

2. **A two-track view of the future:** This sub-theme indicates that participants did not imagine a single fixed trajectory for the future use of AI. Instead, they described two parallel directions. One is the continued integration of AI into everyday work as a tool for personal knowledge support and efficiency enhancement. The other is the possibility of a higher degree of end-to-end automation in certain workflows. This view reflects an expectation of diversified future human–AI role configurations rather than a simple assumption of total human replacement. Even in the more automated trajectory, humans were still seen as necessary for understanding, task decomposition, supervision, and final review.

   Participant P4: *"Building personalized knowledge bases and full-process AI operation are two different dimensions, and both already exist and are happening in practice."*

Overall, this cross-cutting theme addresses both RQ1 and RQ2 within the scope of early Requirements Engineering, particularly requirements elicitation and requirements analysis. It reflects participants' expectations that the allocation of roles between humans and AI in these early-stage tasks may change as AI capabilities improve. However, these expectations were not expressed as a general prediction that AI would take over requirements work. Rather, participants linked future AI involvement to task suitability, the verifiability of AI outputs, perceived risks, and the continued need for human judgment in interpreting stakeholder needs and requirements-related context. Thus, this theme suggests a conditional expansion of AI roles in early RE, where AI may support or lead some bounded tasks, while human responsibility remains central in tasks involving ambiguity, implicit needs, or consequential decisions.

4.3 Survey Results

The survey was used to complement the interview findings by collecting additional practitioner perspectives on how AI roles are allocated in early RE phases, requirements elicitation and requirements analysis. While the interviews provided detailed explanations of how practitioners reason about human–AI collaboration, the survey offered a wider descriptive view of task-level role configurations, perceived benefits and risks, and trust-related conditions across a larger respondent group within this study. The survey questions covered respondents' professional background, involvement in RE stages, current AI use, preferred human–AI role configurations for early RE tasks, perceived benefits and risks of different configurations, and trust in AI outputs.

The survey data consisted of 67 valid responses. The analysis was exploratory and descriptive, focusing on identifying patterns in the respondent group rather than testing hypotheses. Closed-ended questions were summarized using frequencies, percentages, and mean scores. For matrix questions and Likert-scale items, percentages and means were calculated based on valid responses for each item. Matrix questions were analyzed by comparing four role configurations: human-only, AI-assisted, AI-led, and AI-only. Open-ended answers were examined through a descriptive narrative analysis by grouping recurring reasons and problems mentioned by respondents, in order to interpret why participants avoided AI in some tasks and what problems they had experienced when using AI in requirements-related work.

4.3.1 Participants

The survey participants represented a range of professional roles related to software development and requirements work. Software Developers formed the largest group, accounting for 25.4% of the sample, followed by Product Managers or Product Owners at 23.9% and Business Analysts or System Analysts at 14.9%. Tech Leads or Architects represented 10.4% of the respondents, while Project Managers accounted for 9.0%. The remaining participants included QA or Test Engineers, UX or Interaction Designers, and Requirements Engineers. The professional role distribution of the survey participants is summarized in Table 4.2.

The sample also included respondents with varied levels of professional experience. The largest group had 3–6 years of experience (37.3%), followed by respondents with 1–2 years of experience and respondents with more than 10 years of experience, each accounting for 17.9%. Participants with 7–10 years of experience accounted for 16.4%, while 10.4% had less than one year of experience. This distribution indicates that the survey captured both early-career and experienced practitioner perspectives, although mid-level practitioners were most strongly represented.

In terms of organizational context, respondents came mainly from large enterprises (34.3%) and small and medium-sized enterprises (32.8%). Startups accounted for 23.9%, while research or academic organizations and other organization types accounted for smaller shares. The participants were also involved in multiple RE stages. Requirement analysis was the most frequently selected stage (74.6%), followed by requirement elicitation (71.6%), requirement specification (55.2%), and requirement validation (53.7%). Since the study focuses on elicitation and analysis, the high involvement in these two stages suggests that the survey responses are relevant to the research scope.

Table 4.2: Survey participant profile

Category	Frequency	Percentage
Software Developer	17	25.4%
Product Manager / Product Owner	16	23.9%
Business Analyst / System Analyst	10	14.9%
Tech Lead / Architect	7	10.4%
Project Manager	6	9.0%
QA / Test Engineer	4	6.0%
UX / Interaction Designer	4	6.0%
Requirements Engineer	3	4.5%

4.3.2 Descriptive Analysis of Survey Data

AI use in early requirements work

Most respondents reported that they already use AI in requirements elicitation or requirements analysis work. A total of 49.3% answered that they use AI, and another 38.8% reported using AI occasionally. Only 11.9% reported that they do not use AI in these tasks. The reported tools were mainly general-purpose conversational LLMs, AI coding assistants or AI-enabled development tools, internal company LLMs, and, in a few cases, design, prototyping, or workflow-oriented AI tools. Examples of general-purpose LLMs included ChatGPT or GPT-4, Claude, Gemini, Baidu Wenxin, and Qwen. Code-oriented tools included Cursor, Copilot, and Claude Code, while several respondents also reported using internal LLMs provided by their organizations. A smaller number of responses mentioned tools such as Google Stitch, Modao, MasterGo, and Coze, suggesting some use of AI-supported design, prototyping, or workflow-building tools. Among the respondents who did not use AI, the main reasons included organizational policy restrictions, confidentiality concerns, doubts about output reliability, high verification effort, limited value in direct stakeholder communication, and poor fit with existing workflows. One response also indicated that although industry norms may be relatively unified, implementation processes, details, and workflows vary greatly across companies, making it difficult to identify a suitable entry point for AI use. Overall, AI use was already common among the survey participants, but the distinction between regular, occasional, and non-use suggests that adoption remains selective rather than fully embedded in respondents' early RE workflows.

Among these 59 AI users, the most common AI-supported tasks were identifying ambiguities or missing information (55.9%), summarizing meetings or documentation (54.2%), extracting requirements from raw data (52.5%), structuring or organizing requirements (47.5%), and supporting background research or information search (47.5%). Within this respondent group, these descriptive results suggest that AI is mainly used for text-intensive, preparatory, and organization-oriented activities. More decision-related activities were less common, with 32.2% of AI users reporting the use of AI to support prioritization or decision-making. This pattern is consistent with the interview finding that practitioners are more willing to involve AI when the task is bounded, document-based, and easier to verify.

Respondents' reported use of AI outputs further supports this cautious pattern. Only 13.4% reported using AI outputs without modification. Most respondents either used AI outputs only as reference material (31.3%), modified them slightly (29.9%), or heavily revised them (23.9%). Therefore, AI outputs were rarely treated as final products. Instead, they were more often treated as draft material, input for reflection, or provisional artifacts requiring human review.

Conditions for considering AI outputs reliable

The survey responses indicate that trust in AI outputs was strongly conditional. The most frequently selected condition for considering AI outputs reliable was that sufficient input and context had been provided (68.2%). Similar proportions selected simple or well-structured tasks (63.6%) and tasks that did not involve critical decisions (63.6%) as conditions for reliability. A further 62.1% selected that outputs should be easy to verify. These responses suggest that respondents did not evaluate AI reliability as a general property of the technology. Rather, they linked reliability to task characteristics, available context, and the possibility of checking the output.

Verification practices followed the same logic. Respondents most often reported verifying AI outputs when the outputs were uncertain (63.6%), when important decisions were involved (56.1%), when tasks were domain-specific (50.0%), or when sensitive data was involved (45.5%). The most common verification method was cross-checking with original materials such as transcripts, notes, or documents (66.7%). Other common strategies were using another AI tool or model for cross-checking (59.1%) and consulting domain documentation or standards (56.1%). Asking stakeholders or team members for confirmation was selected by 27.3%. These results suggest that practitioners' trust in AI is not passive acceptance, but is actively managed through comparison, review, and contextual validation.

Task-level role configurations

Tables 4.3 and 4.4 present the exact percentage distributions of preferred role configurations for elicitation and analysis tasks, respectively. These tables show how respondents allocated responsibility between humans and AI across specific early RE tasks.

Table 4.3: Preferred role configurations across requirements elicitation tasks. Percentages are calculated based on the total number of valid role-configuration selections for each task.

Elicitation Task	Human-only	AI-assisted	AI-led	AI-only
Summarizing elicitation sessions	12.1%	56.1%	28.8%	3.0%
Extracting requirements from raw inputs	21.5%	46.2%	29.2%	3.1%
Identifying ambiguities or missing information	48.5%	42.4%	6.1%	3.0%
Suggesting clarification questions	29.2%	41.5%	24.6%	4.6%

Table 4.4: Preferred role configurations across requirements analysis tasks. Percentages are calculated based on the total number of valid role-configuration selections for each task.

Analysis Task	Human-only	AI-assisted	AI-led	AI-only
Structuring and standardizing requirements	26.2%	46.2%	23.1%	4.6%
Structuring and organizing requirements	23.1%	52.3%	23.1%	1.5%
Detecting conflicts or inconsistencies	41.5%	41.5%	15.4%	1.5%
Supporting prioritization or trade-off decisions	67.7%	27.7%	3.1%	1.5%

Figures 4.3 and 4.4 visualize the same distributions to make the task-level comparison easier to interpret.

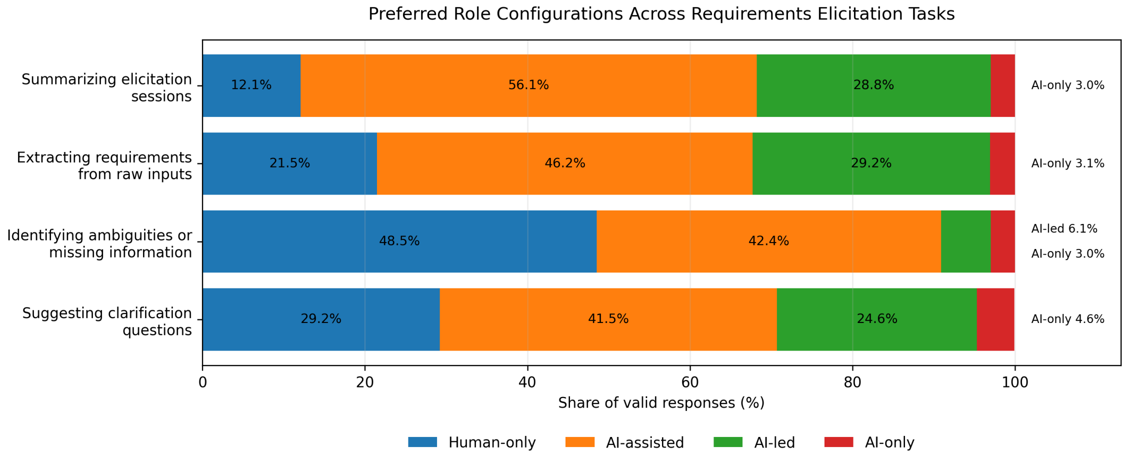

Figure 4.3: Selected human–AI role configurations across requirements elicitation tasks. Each horizontal bar represents one elicitation task, and the segments show the proportion of valid role-configuration selections assigned to each role configuration. Small-percentage labels are shown to the right of the bars for readability.

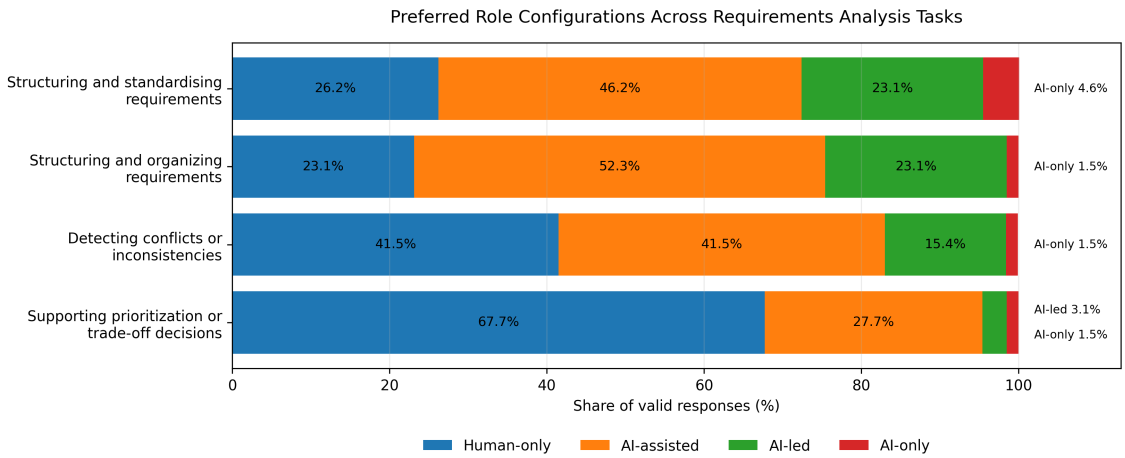

Figure 4.4: Selected human–AI role configurations across requirements analysis tasks. Each horizontal bar represents one analysis task, and the segments show the proportion of valid role-configuration selections assigned to each role configuration. Small-percentage labels are shown to the right of the bars for readability.

Overall, within the survey sample, AI-assisted configurations were more common for summarization, extraction, and structuring tasks, while human-only configurations increased for ambiguity identification, conflict-related analysis, and prioritization decisions.

The task-level results become clearer when the tasks are separated into elicitation-facing tasks and analysis tasks. Within elicitation-facing work, respondents were most open to AI involvement when the task was based on already captured material. Summarizing elicitation sessions showed the highest AI-assisted share (56.1%)

and a relatively high AI-led share (28.8%), while extracting requirements from raw inputs also leaned toward AI-assisted (46.2%) and AI-led (29.2%). Suggesting clarification questions showed a similar but more cautious pattern, with AI-assisted as the largest category (41.5%) but substantial human-only (29.2%) and AI-led (24.6%) selections. In contrast, identifying ambiguities or missing information shifted toward human-only work (48.5%), indicating that once elicitation moves from condensing stakeholder material to interpreting what is unclear, incomplete, or needs follow-up, respondents become less willing to delegate the task to AI.

The analysis tasks show a related but more differentiated pattern. For structuring and standardizing requirements and structuring and organizing requirements, AI-assisted was again the dominant configuration (46.2% and 52.3%, respectively), suggesting that respondents considered AI useful for analytical work that reorganizes and aligns existing requirements artifacts. However, the pattern changed when analysis required stronger interpretive or evaluative judgment. Detecting conflicts or inconsistencies produced an even split between human-only and AI-assisted (both 41.5%), which suggests that practitioners may accept AI as a checking aid but not as the sole analytical authority. Supporting prioritization or trade-off decisions showed the clearest boundary, with 67.7% selecting human-only and only 3.1% selecting AI-led. This indicates that delegation decreases sharply when analysis tasks carry stronger consequences for value judgments, stakeholder negotiation, or accountability.

Taken together, the survey responses suggest a task-specific delegation pattern across early RE. In elicitation, AI appears more acceptable for capture and consolidation tasks based on transcripts, notes, or raw statements, but less acceptable for diagnosing ambiguities that require contextual understanding of stakeholder intent. In analysis, AI is also accepted for structuring work, yet delegation narrows again when the task shifts from organizing requirements to judging conflicts or making trade-offs. This distinction adds more precision to the earlier finding that AI is mainly used for bounded and verifiable work: respondents were not only differentiating between elicitation and analysis in general, but also between lower-interpretation and higher-interpretation tasks within each phase.

AI-only configurations were rare across all tasks, never exceeding 4.6%. Even for tasks where AI-led configurations were relatively visible, such as summarizing elicitation sessions (28.8%) and extracting requirements from raw inputs (29.2%), respondents still rarely selected full AI autonomy. This suggests that the survey participants did not reject AI involvement, but they also did not support removing humans from early RE tasks.

4.3.3 Narrative Analysis of Survey Data

The open-ended survey questions were used to further interpret the reasons behind respondents' cautious and conditional role allocation. The first question asked why respondents choose not to use AI in some tasks. Eighteen respondents provided meaningful non-empty answers to this open-ended question, and one additional non-user response provided relevant explanations in earlier survey fields. Therefore, 19 comments related to AI non-use or avoidance were considered in this part of the narrative analysis. The most common theme was confidentiality or policy constraints, appearing in 8 of the 19 comments. Respondents described organizational rules, client data sensitivity, or concerns about sharing internal information as reasons for avoiding AI tools. This theme aligns with the interview finding that data security can limit AI use even when practitioners see potential benefits.

Another recurring set of comments concerned output quality and contextual reliability. Several comments referred to problems such as hallucination risk, generic outputs, missing domain context, or difficulties in extracting key information from fragmented client conversations. Participants described low confidence in AI correctness when requirements were ambiguous, context-dependent, or based on incomplete stakeholder communication. A related concern was high verification overhead. In these answers, respondents explained that checking and correcting AI-generated output could sometimes take as much time as completing the task manually.

A further group of comments concerned task and workflow fit. Some respondents preferred direct stakeholder communication or considered AI less useful for tasks involving major decisions. One non-user also explained that AI was difficult to apply because existing workflows did not provide a clear entry point, and because implementation processes and details vary substantially across companies even when industry norms are relatively unified. This suggests that non-use was not only caused by distrust in AI itself, but also by practical constraints in organizational workflow, stakeholder interaction, and task context.

The second open-ended question asked whether respondents had experienced problems when using AI. Twenty-six respondents provided meaningful non-empty answers. The recurring problems were hallucinated or fabricated details, inconsistent outputs across runs, misinterpretation caused by missing context, low explainability, and data leakage concerns. These themes show that respondents' reservations were not only abstract worries. They were linked to concrete experiences of AI outputs being plausible but unreliable, unstable between attempts, or difficult to justify.

Taken together, the open-ended responses indicate a bounded-use understanding of AI in early requirements work. Respondents were generally willing to use AI for drafting, summarizing, organizing, and checking material, but they became more cautious when tasks involved stakeholder intent, implicit assumptions, sensitive information, workflow-specific context, or accountability for decisions. In narrative terms, many responses followed a similar structure: AI was first recognized as useful for efficiency, then qualified by a condition such as verifiability, data sensitivity, task fit, or workflow compatibility, and finally positioned as requiring human review. This pattern supports the quantitative finding that AI is more accepted as an assistant than as an autonomous actor in early RE.

4.3.4 Insights Related to Research Questions

RQ1: Role and responsibility delegation between humans and AI

For information-processing and documentation-oriented tasks, AI-assisted configurations were the most common choice. Summarizing elicitation sessions, extracting requirements from raw inputs, structuring and standardizing requirements, and organizing requirements all received high AI-assisted shares. These tasks are closely related to transforming existing material into more usable forms. In such contexts, AI appears to have been perceived as a practical support tool that can reduce manual effort and improve efficiency while still leaving humans in control.

However, when tasks moved closer to interpretation and decision-making, respondents shifted toward stronger human involvement. Identifying ambiguities or missing information requires understanding the meaning and implications of requirements, not only processing text. Detecting conflicts or inconsistencies showed a balanced pattern, with human-only and AI-assisted configurations receiving equal shares, suggesting that AI may be accepted as a checking aid but not as a replacement for human interpretation. Supporting prioritization or trade-off decisions showed the clearest human boundary, with 67.7% selecting human-only. This suggests that respondents associated prioritization with accountability, stakeholder value, and project consequences, making it less suitable for autonomous AI involvement.

The low selection of AI-only configurations across all tasks is also important for RQ1. It indicates that respondents generally did not conceptualize AI as a replacement for practitioners in early RE. Instead, the most common role pattern is collaborative but human-controlled: AI can assist, and in some bounded tasks may lead the initial processing, but humans remain responsible for interpretation, validation, and final decisions.

RQ2: Benefits, risks, and trust considerations in role division

Table 4.5 summarizes the average ratings of benefit and risk items across the four role configurations. Benefit items were rated on a 1–5 scale, where higher values indicate stronger perceived benefit. Risk items were also rated on a 1–5 scale, where higher values indicate stronger perceived risk; the item “No major risk” was excluded from the average risk rating. These mean values are reported as descriptive summaries of relative tendencies across items and role configurations rather than as inferential results or precise interval-scale measurements.

Table 4.5: Average benefit and risk item ratings by role configuration

Role configuration	Avg. benefit rating	Avg. risk rating
Human-only	2.55	1.94
AI-assisted	3.96	3.10
AI-led	3.85	3.56
AI-only	3.59	4.13

Figures 4.5, 4.6, and 4.7 complement these mean values by showing the item-level patterns and response distributions behind the benefit, risk, and trust responses.

Figure 4.5: Heat map of Q9 benefit ratings across the four role configurations. Darker shading indicates stronger perceived benefit for a given benefit item under each human–AI role configuration.

Figure 4.5 indicates that perceived benefits were consistently higher for AI-assisted and AI-led configurations than for human-only work across most benefit items. The strongest concentrations appear in efficiency- and preparation-oriented items, especially time saving, faster information access, reduced workload or cognitive load, and faster generation of initial drafts. This pattern is also reflected in the average ratings: AI-assisted work received the highest average benefit-item rating (3.96), slightly higher than AI-led work (3.85) and clearly higher than AI-only work (3.59), while human-only work received the lowest average rating (2.55). Taken together, these descriptive results suggest that respondents associated AI involvement mainly with efficiency-oriented and preparatory benefits rather than with replacing human responsibility.

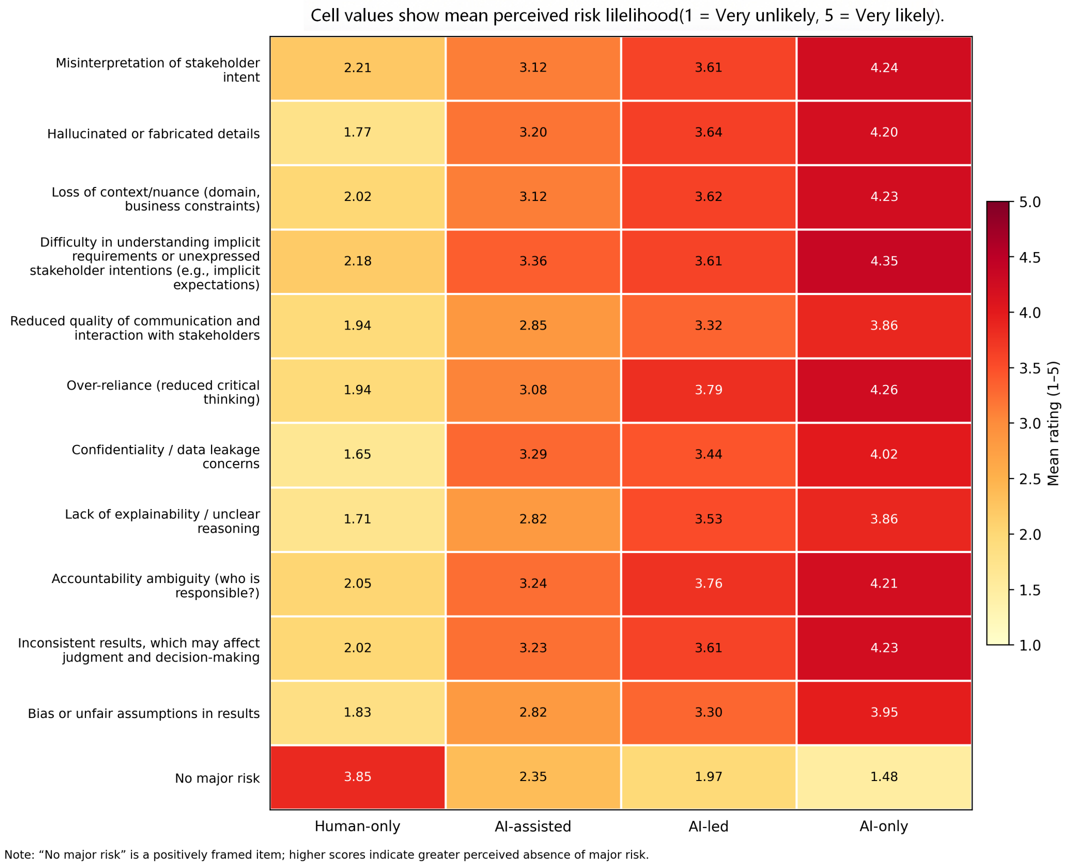

Figure 4.6: Heat map of Q10 risk ratings across the four role configurations. Darker shading indicates stronger perceived risk for a given risk item under each human–AI role configuration.

Figure 4.6 indicates the opposite pattern for risks: perceived risk increases as AI receives more autonomy. Human-only work had the lowest average risk-item rating (1.94), AI-assisted work had a moderate average rating (3.10), the rating increased further for AI-led work (3.56), and it was highest for AI-only work (4.13). Across the heat map, the darkest concentrations for AI-only appear in risks such as hallucinated or fabricated details, loss of context or nuance, over-reliance, accountability ambiguity, and misinterpretation of stakeholder intent. The reversed pattern for the "No major risk" item further supports this interpretation, as it was rated highest for human-only and lowest for AI-only. Together, these descriptive results suggest that practitioners became more cautious as AI moved from supportive use toward more autonomous and accountable roles.

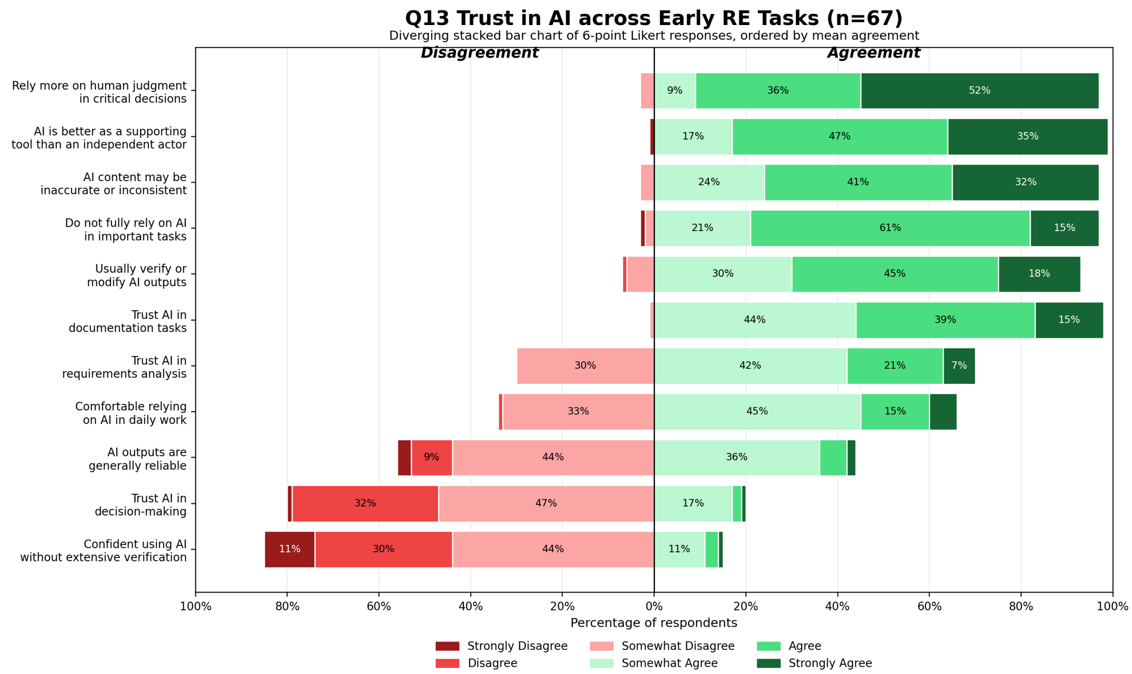

Figure 4.7: Diverging stacked bar chart of Q13 Likert-scale responses on trust in AI-supported Requirements Engineering tasks. Items are ordered by mean agreement score from highest to lowest to highlight the most and least agreed-on statements.

Figure 4.7 is consistent with this interpretation by showing that trust in AI is stronger for documentation-related and analysis-support tasks, but weaker for decision-making. This is consistent with the mean ratings: trust in AI for documentation-related tasks had a mean of 4.65 on a 1–6 scale, while trust in AI for requirements analysis tasks was lower at 4.02 and trust in AI for decision-making processes was much lower at 2.88. The response distribution also shows strong agreement that AI outputs should be verified before use, that critical decisions should rely more on human judgment (mean 5.38), and that AI is better suited as a supporting tool rather than an independent actor (mean 5.14). Together, the mean ratings and the response distribution suggest bounded trust: respondents did not reject AI, but they trusted it more in bounded support tasks than in tasks involving stronger judgment and accountability.

Exploratory descriptive comparisons by role suggest the same cautious tendency across professional groups, although group sizes were small. Mean trust in AI for decision-making was lowest among Business Analysts or System Analysts (2.40), Project Managers (2.50), and Product Managers or Product Owners (2.56), while Software Developers (3.41) and Tech Leads or Architects (3.17) appeared relatively more open to AI involvement in decision-adjacent reasoning. This difference should not be interpreted as a strong statistical claim, but it suggests that roles closer to stakeholder negotiation and product responsibility may be more cautious about delegating decisions to AI.

Overall, the survey findings are consistent with the central interpretation emerging from the study: appropriate AI roles in early RE appear to be defined not by AI capability alone, but by the relationship between task characteristics, expected benefits, risks, and trust conditions. AI is perceived as useful when it accelerates bounded and verifiable work. However, as tasks become more interpretive, context-dependent, sensitive, or decision-oriented, respondents shift responsibility back to humans. The survey therefore reinforces a human-centered model of AI involvement in requirements elicitation and analysis, where AI can support and sometimes lead preliminary processing, but human judgment remains necessary for validation, interpretation, and accountability.

4.3.5 Integration of Interview and Survey Findings

The interview and survey findings were examined together to identify areas of convergence, extension, and divergence. In this study, triangulation was understood not as using one phase to validate the other, but as examining how the two forms of evidence supported, refined, or extended the interpretation of human–AI role allocation in early Requirements Engineering.

The interview and survey findings showed a high degree of convergence regarding the role of AI in early RE. ***Across both phases, AI was mainly positioned as a supporting tool for bounded, text-intensive, and verifiable tasks, such as summarizing, extracting, organizing, and drafting requirements-related material.*** Interview participants described AI as useful for reducing manual effort and producing initial structures, while survey respondents most often selected AI-assisted configurations for summarizing elicitation sessions, extracting requirements from raw inputs, and structuring requirements. This convergence is consistent with the interpretation that practitioners do not reject AI in early RE, but tend to allocate it to tasks where outputs can be reviewed, corrected, and contextualized by humans.

Both phases also converged on the importance of human responsibility in interpretive and decision-related tasks. Interview participants repeatedly emphasized that stakeholder understanding, contextual judgment, and final decisions should remain human responsibilities. Similarly, the survey showed stronger preference for human-only configurations in tasks such as identifying ambiguities or missing information and supporting prioritization or trade-off decisions. This suggests that practitioners draw a boundary between using AI to process requirements information and allowing AI to take responsibility for interpreting stakeholder intent or making consequential decisions.

The survey findings also extended the interview findings by showing how this cautious role allocation appeared across a broader respondent group within the study. While the interviews provided detailed explanations of why practitioners verify AI outputs and avoid full automation, the survey summarized these tendencies numerically through role configuration choices, trust ratings, and benefit-risk comparisons. In particular, AI-assisted work received the highest perceived benefit, while perceived risk increased as AI received more autonomy. This is consistent with the interview-based interpretation that practitioners value AI mainly when it remains under human control.

When the two phases are integrated more explicitly, they support a task-specific mapping of role configurations across elicitation and analysis. A first cluster consists of elicitation capture and consolidation tasks, such as summarizing elicitation sessions and extracting requirements from raw inputs, where AI-assisted and AI-led configurations were relatively acceptable because the expected contribution was speed, draft generation, and transformation of already available material. A second cluster consists of elicitation sense-making tasks, such as identifying ambiguities or suggesting clarification questions, where delegation became more cautious because the work depends more directly on understanding what stakeholders meant, what is still missing, and what should be asked next. A third cluster consists of analysis structuring tasks, such as standardizing, organizing, and structuring requirements, where AI-assisted configurations were again dominant because the work remained comparatively bounded and reviewable. A fourth cluster consists of analysis judgment tasks, such as conflict detection and especially prioritization or trade-off decisions, where human-only configurations became equally common or dominant because these tasks involve stronger interpretive responsibility and project consequences. Taken together, this forms a practical taxonomy of delegation patterns: the closer a task is to transforming explicit material, the more delegation to AI is accepted; the closer it is to interpreting intent or making consequential judgments, the more responsibility is retained by humans.

Based on this integrated interpretation, Table 4.6 summarizes the task-level role-boundary framework derived from the interview and survey findings.

Table 4.6: Task-level role-boundary framework for AI involvement in early Requirements Engineering

Task cluster	Example tasks	Suitable role boundary	Boundary rationale
Elicitation capture and consolidation	Summarizing elicitation sessions; extracting requirements from raw inputs	AI-assisted or AI-led preliminary processing, followed by human review	Verifiability is relatively high because the task is based on already collected stakeholder material. Perceived risk is lower when outputs can be checked against original sources, while accountability remains with humans through review.
Elicitation sense-making	Identifying ambiguities or missing information; suggesting clarification questions	AI-assisted support, with humans deciding what needs clarification	Verifiability is more limited because the task depends on stakeholder intent, missing context, and follow-up needs. Perceived risk increases if AI misunderstands what is implicit, so humans remain accountable for deciding what should be clarified.
Analysis structuring	Structuring, organizing, and standardizing requirements	AI-assisted work under human validation	Verifiability is moderate to high because the task is document-based and reviewable. Perceived risk remains manageable if humans validate relevance, correctness, and contextual fit before using the structured output.
Analysis judgment and decisions	Detecting conflicts or inconsistencies; supporting prioritization or trade-off decisions	AI-assisted checking for conflict detection; final judgment and prioritization remain human-led	Verifiability is lower and perceived risk is higher because these tasks involve interpretation, stakeholder consequences, prioritization, and accountability. AI outputs should therefore support human decision-making rather than serve as final conclusions.

Some differences between the two phases also provided additional nuance. The interviews placed strong emphasis on human judgment in tasks involving ambiguity, conflict, and stakeholder interpretation. The survey generally supported this pattern, but also showed a more balanced result for detecting conflicts or inconsistencies, where human-only and AI-assisted configurations were equally selected. This suggests that some practitioners may accept AI as a checking aid for conflict detection, even if they do not consider it suitable for taking full responsibility for interpretation or decision-making.

The open-ended survey responses further refined the interview findings by highlighting practical reasons for AI non-use. In addition to confidentiality, policy restrictions, and concerns about hallucinated or unreliable outputs, some respondents pointed to workflow fit, lack of clear entry points for AI use, and variation in implementation processes across companies. This indicates that AI role allocation is shaped not only by task characteristics, but also by organizational context and the practical conditions under which AI can be integrated into existing requirements work.

The integrated evidence also suggests that trust mediates willingness to delegate across these task clusters. The survey showed higher trust in AI for documentation-related tasks (mean 4.65) than for requirements analysis tasks (mean 4.02), and much lower trust for decision-making processes (mean 2.88). This gradient aligns with the role-configuration choices: practitioners delegated more readily when expected benefits were efficiency-oriented and outputs could be verified, but withdrew delegation as perceived risk increased around context loss, misinterpretation, inconsistent outputs, accountability ambiguity, and confidentiality concerns. In this sense, benefit, risk, and trust do not operate as separate themes. Rather, they interact to shape delegation patterns: expected benefit creates openness to AI support, perceived risk limits how far autonomy can be extended, and trust reflects whether a task is considered sufficiently bounded and verifiable for AI involvement.

Overall, the integrated findings indicate that appropriate AI involvement in early RE is task-contingent, conditional, and human-centered. AI is perceived as valuable for improving efficiency and structuring requirements information, but human judgment remains central when requirements work involves ambiguity, stakeholder intent, sensitive information, workflow-specific context, or accountability for decisions. The combined interview and survey results therefore contribute not only a descriptive account of current practice, but also a theoretical explanation of delegation in early RE: practitioners assign human–AI role configurations according to the interaction between task type, expected benefit, perceived risk, and task-specific trust.

5

Discussion

The main finding of this study is that AI role allocation in early RE appears to be conditional rather than general. Practitioners are willing to involve AI in bounded, information-heavy, and verifiable work, but they keep humans responsible for stakeholder understanding, contextual interpretation, prioritization, trade-off decisions, and accountability. In line with general guidance that discussion sections should explain the meaning of the findings, relate them to prior work, consider alternative explanations, and acknowledge limitations [30], this chapter discusses what this finding means for the two research questions, how it relates to prior work, what alternative explanations should be considered, and what implications it has for practice, research, and academia.

5.1 RQ1: Task-level AI role configurations

5.1.1 Major findings

RQ1 asked how roles and responsibilities between humans and AI are delegated in requirements elicitation and requirements analysis tasks. The results suggest that practitioners do not assign one fixed role to AI across early RE. Instead, role allocation changes with the nature of the task. AI was most often positioned as an assistant or preliminary processor in tasks that transform existing information, such as summarizing elicitation sessions, organizing notes, extracting candidate requirements, structuring requirement-related material, and preparing draft questions. Human responsibility remained central in tasks that require stakeholder interaction, interpretation of implicit needs, conflict resolution, prioritization, and final decisions. This suggests that practitioners distinguish between tasks where AI outputs can be reviewed against existing material and tasks where the main work involves contextual interpretation, stakeholder understanding, or accountability for decisions.

The interview findings illustrate this distinction. Participants described AI as useful for searching terminology, supporting background understanding, summarizing meeting content, organizing documents, comparing materials, and handling large volumes of text. These activities are valuable but bounded: the input material is usually available, and the AI output can be checked against transcripts, notes, documents, or stakeholder statements. In contrast, participants were cautious about using AI as a replacement for human elicitation or analysis. They emphasized that requirements work involves asking follow-up questions, negotiating meaning, understanding unstated needs, and deciding how requirements fit product goals and organizational constraints.

The survey results provide descriptive support for this pattern within the study sample. AI-assisted configurations were most frequently selected for summarizing elicitation sessions (56.1%), structuring and organizing requirements (52.3%), extracting requirements from raw inputs (46.2%), and structuring and standardizing requirements (46.2%). AI-led configurations were visible in some bounded tasks, especially extracting requirements from raw inputs (29.2%) and summarizing elicitation sessions (28.8%), but AI-only configurations were rare and never exceeded 4.6%. In this context, AI-led responses should be interpreted cautiously, as they may indicate that AI produces the main initial output while humans still remain involved in review, rather than full autonomous responsibility. By contrast, human-only configurations were most common for identifying ambiguities or missing information (48.5%), detecting conflicts or inconsistencies (41.5%), and supporting prioritization or trade-off decisions (67.7%).

Taken together, these results suggest that the dominant configuration in early RE is not full automation, but human-controlled collaboration. AI can contribute to preliminary processing and organization, while humans preserve responsibility for interpretation, validation, stakeholder communication, and decisions with project consequences. The role boundary therefore appears to depend on whether a task is bounded, text-based, and verifiable, or whether it requires contextual judgment, stakeholder-sensitive interpretation, and accountable decision-making.

5.1.2 Meaning of the findings

The findings suggest that task characteristics are more important than a general attitude of either accepting or rejecting AI. Prior RE research has shown that earlier NLP-based techniques can support ambiguity detection in requirements texts [10]. GenAI and LLM-oriented RE studies further examine support for requirements-related tasks such as summarization, classification, requirements analysis, conflict-related reasoning, and requirements drafting or generation [3, 25, 12]. However, recent empirical work also shows that LLM-generated requirements quality depends on input clarity and prompt design, and that human oversight remains necessary for refinement and validation [12]. Work at the intersection of formal requirements engineering and large language models further emphasizes that LLM-generated requirements artifacts raise concerns about correctness, fairness, and trustworthiness [14]. This study is consistent with the technical potential identified in prior RE research, but extends it with a responsibility-oriented interpretation: practitioners do not evaluate AI only by what it can generate, but by whether the task is verifiable, context-dependent, ambiguous, and consequential.

This interpretation explains why AI-assisted configurations were accepted for documentation and structuring work. These tasks are text-intensive and often involve reorganizing existing material into more usable forms. When the original material is available, practitioners can inspect whether the AI has omitted, distorted, or invented information. The possibility of verification makes AI involvement more acceptable. In such settings, AI can reduce manual effort without taking over responsibility for the meaning of the requirement.

The same logic explains why practitioners resisted stronger AI roles in interpretive and decision-related tasks. Early RE is socially situated and ambiguous [1, 4, 10]. Stakeholders may express incomplete needs, use inconsistent terminology, disagree about priorities, or leave important assumptions unstated. The meaning of a requirement often emerges through interaction rather than from the text alone. This study extends prior descriptions of early RE ambiguity by showing how such ambiguity affects human–AI delegation decisions. Therefore, tasks such as identifying missing information, resolving conflicts, and deciding priorities require contextual judgment. AI may identify textual patterns or produce plausible suggestions, but it does not necessarily understand the organizational, political, contractual, or interpersonal context behind those patterns.

This finding also suggests that human-led boundaries in early RE are not only technical but also professional and organizational. The preference for human control in decision-related tasks may reflect professional responsibility as well as task complexity. Practitioners may resist AI-led prioritization not only because AI may lack contextual understanding, but also because prioritization is tied to role identity, accountability, and relationships with stakeholders. Acceptable role allocation therefore depends not only on whether an AI system can produce a plausible recommendation, but also on who remains answerable for the outcome.

The results also refine how automation-level concepts should be applied to early RE. General automation models describe a continuum from human control to high automation [5], while human-centered AI literature emphasizes maintaining meaningful human control and responsibility in high-impact AI use [24]. The findings suggest that this continuum is useful only if it is translated to the task level. A task such as summarizing a meeting can tolerate an AI-led initial output, while a task such as prioritizing requirements remains human-led even when AI provides supporting arguments. Thus, the same RE phase can contain several appropriate levels of AI involvement. This responds to the research gap identified in the related work: role configurations in early RE need to be understood at the level of concrete tasks, not only at the level of broad process phases.

This interpretation confirms prior work on task delegability, which shows that full AI automation is rarely preferred and that human-led AI assistance is often more acceptable [16]. However, this study extends that general task-delegability perspective to the specific context of early Requirements Engineering. In this context, delegation is not shaped only by general perceptions of trust and risk, but also by whether AI outputs are verifiable, whether the task involves stakeholder interaction, and whether the outcome requires contextual interpretation or accountable decision-making. While prior task-delegability research examines broad everyday and professional tasks, this study offers domain-specific evidence suggesting how AI role allocation changes across concrete requirements elicitation and analysis tasks.

From this perspective, human-centered AI involvement does not mean that AI is used only marginally. Respondents were open to substantial AI involvement in selected tasks, as shown by the AI-led shares for summarization and extraction. However, human-centeredness means that humans retain responsibility for validating, contextualizing, and acting on AI outputs. This is especially important in early RE because early interpretations can propagate into specification, design, implementation, and validation. A plausible but incorrect AI-generated interpretation may become costly if it is stabilized too early in the project.

The answer to RQ1 is therefore that role delegation in early requirements elicitation and analysis is best understood as task-contingent human–AI collaboration. AI is most appropriate as an assistant or preliminary processor in bounded, information-heavy, and verifiable tasks. Human responsibility remains dominant when the task depends on stakeholder communication, implicit meaning, conflict resolution, prioritization, trade-offs, and accountability.

5.2 RQ2: Benefits, risks, and trust considerations

5.2.1 Major findings

RQ2 asked what benefits, risks, and trust considerations are associated with different human–AI role configurations. The results suggest a conditional logic of usefulness. Practitioners value AI because it can save time, reduce workload, support information access, produce initial drafts, and structure large volumes of material. At the same time, trust remains bounded by task criticality, verifiability, available context, data sensitivity, and previous experience with AI errors.

The perceived benefits were strongest for AI-assisted work. In the survey, AI-assisted configurations had the highest overall benefit mean (3.96 on a 1–5 scale), followed by AI-led work (3.85), AI-only work (3.59), and human-only work (2.55). The largest perceived benefit gaps between AI-assisted and human-only configurations concerned faster information access, time saving, reduced workload or cognitive load, and faster generation of initial drafts. This suggests that respondents primarily associated AI's value with acceleration and support rather than with independent ownership of requirements work.

Perceived risks increased as AI received more autonomy. Human-only work had the lowest risk mean (1.94), AI-assisted work had a moderate risk mean (3.10), AI-led work had a higher risk mean (3.56), and AI-only work had the highest risk mean (4.13). Under the AI-only configuration, the highest-rated risks concerned difficulty understanding implicit requirements or unexpressed stakeholder intentions, over-reliance, loss of context or nuance, misinterpretation of stakeholder intent, inconsistent results, hallucinated or fabricated details, accountability ambiguity, and confidentiality or data leakage concerns.

Trust followed the same task-sensitive pattern. Respondents reported higher trust in AI for documentation-related tasks (mean 4.65 on a 1–6 scale) than for requirements analysis tasks (mean 4.02), and much lower trust for decision-making processes (mean 2.88). They also strongly agreed that critical decisions should rely more on human judgment (mean 5.38), that AI is better suited as a supporting tool than as an independent actor (mean 5.14), and that AI-generated content may be inaccurate or inconsistent (mean 5.02). These findings suggest bounded trust rather than general skepticism: practitioners trust AI more when the task is easier to verify and less when the task requires contextual judgment or accountability.

The open-ended survey responses explain why this bounded trust emerges. Among respondents who avoided AI in some tasks, confidentiality or policy constraints were the most frequent reason. Other reasons included hallucination, low confidence in ambiguous early RE tasks, and high verification overhead. Respondents who had experienced problems with AI reported hallucinated or fabricated details, inconsistent outputs across runs, misinterpretation caused by missing context, low explainability, and data leakage concerns. These concerns are grounded in concrete experience rather than only in abstract fear of new technology.

5.2.2 Meaning of the findings

The findings suggest that benefits and risks are connected rather than separate. The same AI involvement that creates efficiency can also create interpretive and accountability risks if the system is given too much authority. AI-assisted configurations are attractive because they offer a compromise: practitioners can gain speed and support while preserving human oversight. This helps explain why AI-assisted work received the highest perceived benefit rating and why participants repeatedly described AI as useful but not independently responsible.

This pattern is consistent with research on trust and appropriate reliance in automation, which emphasizes the need to avoid both overreliance and underuse [15]. Parasuraman and Riley's distinction between misuse and disuse helps explain why neither unrestricted AI autonomy nor complete rejection of AI is desirable in early RE [17]. AI-only or strongly AI-led configurations may foster overreliance when outputs are accepted without sufficient contextual checking, while human-only configurations may lead to underuse in tasks where AI can provide useful support, such as summarizing, extracting, or organizing requirements-related material. Recent experimental evidence on AI reliance helps explain the first of these risks: in uncertain situations, users may follow AI advice even when it conflicts with available contextual information or their own assessment [22]. In early RE, where AI outputs often need to be checked against stakeholder intent and project context, stronger AI authority may therefore amplify reliance problems rather than simply improve efficiency.

The same logic also supports the distinction between trust in AI and trustworthy

AI. Practitioners' willingness to rely on AI was not based only on general confidence in AI, but on whether the system appeared suitable for the specific RE task and whether its outputs could be reviewed, corrected, or justified [20]. This is consistent with empirical research on trust in AI, which shows that trust is shaped by factors such as task characteristics, transparency, reliability, and users' perceptions rather than by technical capability alone [19]. In this sense, this study does not argue for more trust in AI as an end in itself. Instead, it argues for appropriate and calibrated trust, where AI involvement is constrained or expanded according to task type, output verifiability, risk, and accountability.

However, trust calibration in early RE is more difficult than in many structured decision-support settings. Requirements elicitation and analysis often lack a single objectively correct answer. A requirement may be incomplete, ambiguous, politically sensitive, or dependent on tacit stakeholder knowledge. This is consistent with prior RE research showing that elicitation involves communication, interpretation, and negotiation of stakeholder needs, rather than simply collecting ready-made requirements [1, 4]. In such conditions, AI-generated outputs can be fluent and well organized while still being wrong in context. Work on LLM calibration helps explain why this is difficult to manage: users may overestimate the accuracy of LLM-generated answers when explanations appear detailed or confident [23]. This problem is especially relevant to RE, where GenAI-for-RE literature warns that coherent AI-generated outputs may still lack sufficient grounding in domain knowledge, stakeholder intent, or project-specific context [3, 25]. As a result, plausible AI-generated interpretations may be accepted before the underlying uncertainty has been resolved. In this context, overreliance can be understood as unwarranted trust: practitioners may rely on AI for an interpretive task even though the system is not sufficiently trustworthy for that task-specific role [18].

The risk of over-reliance is therefore closely related to premature closure. If AI produces a clear summary, classification, or prioritization, practitioners may treat uncertain stakeholder input as more settled than it really is. This is particularly problematic for implicit requirements, unclear expectations, ambiguous terminology, and conflicting stakeholder needs. Prior work on ambiguity in RE has shown that terminology and domain knowledge can create misunderstandings between stakeholders, especially when different domains or professional backgrounds are involved [10]. The survey finding that difficulty understanding implicit requirements was the highest-rated AI-only risk supports this interpretation. It marks a boundary between processing expressed information and understanding what has not yet been fully expressed.

The findings also show that data governance is not a secondary issue but a condition for adoption. Early RE materials often include client information, internal strategy, business rules, domain terminology, or sensitive stakeholder statements. Even when AI is technically capable of supporting a task, practitioners may avoid it if the available tool environment does not satisfy confidentiality or organizational policy requirements. This is consistent with recent practitioner-oriented RE evidence, which emphasizes that responsible AI use in RE requires governance, human oversight, and attention to organizational constraints [11]. This means that role configuration depends not only on task characteristics, but also on whether the AI system is public, organization-approved, integrated with internal knowledge, and governed by secure data-handling practices.

The role-based survey comparison adds a further, cautious interpretation. Business Analysts/System Analysts and Product Managers/Product Owners reported lower trust in AI for decision-making than Software Developers and Tech Leads/Architects. Because the group sizes were small, this should not be treated as a strong statistical claim. Nevertheless, the pattern is meaningful as a descriptive indication: roles closer to stakeholder negotiation, product value, and business accountability may experience the risks of AI decision-making more directly. This interpretation is consistent with prior work showing that requirements elicitation involves multiple roles, communication challenges, and difficulties in understanding and prioritizing stakeholder needs [1]. For these roles, an incorrect requirement interpretation can affect stakeholder trust, contractual scope, customer satisfaction, or product direction.

The answer to RQ2 is therefore that AI-assisted configurations are perceived as the most consistently beneficial and acceptable form of AI involvement in early RE within this study. AI-led configurations can provide value in selected bounded and verifiable tasks, such as summarization and extraction, but they are associated with more mixed evaluations and higher perceived risks than AI-assisted work. Benefits are strongest when AI supports information-heavy tasks under human review, while risks increase when AI is given greater autonomy in tasks involving contextual understanding, implicit meaning, sensitive information, or accountability for decisions. Trust in AI is therefore task-sensitive rather than general, shaped by task criticality, available context, output verifiability, data sensitivity, and prior experience with AI errors.

5.3 Implications for Practitioners, Researchers, and Academia

5.3.1 Implications for Practitioners

For practitioners, the main implication is that AI adoption in early RE should be managed at the task level. The practical question is not whether AI should be used in requirements elicitation and analysis, but which role AI should take in each task and what review is required before the output is used.

For low-risk and information-heavy tasks, practitioners can reasonably use AI as an assistant or preliminary processor. Examples include summarizing meeting transcripts, organizing stakeholder notes, extracting candidate requirements from raw input, generating draft clarification questions, grouping similar feedback, and identifying possible inconsistencies for later review. These tasks can benefit from AI's ability to process and structure text quickly, especially when the original material is available for checking.

For tasks involving stakeholder intent, prioritization, conflict resolution, or trade-off decisions, the findings support maintaining stronger human control. AI may still be useful for preparing options, highlighting possible issues, or producing discussion material, but the final interpretation and decision should remain human-led. This is particularly important when requirements involve implicit needs, organizational politics, contractual scope, regulatory constraints, or sensitive data.

The findings suggest that organizations may benefit from establishing explicit verification practices for AI-supported RE work. The survey results show that practitioners already use strategies such as checking outputs against original materials, comparing outputs across AI tools, consulting domain documentation, and asking stakeholders or team members for confirmation. These practices can be formalized into lightweight guidelines. For example, AI-generated summaries should be checked against transcripts; AI-generated requirement lists should be reviewed by analysts; AI-suggested priorities should be treated as discussion input rather than a decision; and sensitive data should be processed only through approved tools.

Another practical implication concerns knowledge management. Interview participants emphasized internal terminology, domain-specific concepts, product knowledge, and organizational context. AI performed better when such context was provided and worse when it lacked internal knowledge. Organizations aiming to use AI in RE may therefore need curated knowledge bases, domain-specific prompting practices, and secure internal AI environments. These investments should support human judgment rather than justify full automation of interpretive work.

5.3.2 Implications for Researchers

For researchers, the study contributes a task-level perspective on AI roles in Requirements Engineering. Existing work has shown that AI can technically support tasks such as summarization, ambiguity detection, classification, and conflict detection [3, 25, 10, 13]. This study adds that technical capability must be interpreted through responsibility boundaries. A model that lists AI-supported tasks is incomplete unless it also explains how much authority AI should have in each task and under what conditions.

The findings also suggest that future research should investigate role configurations as socio-technical arrangements rather than only as tool features. The same AI capability may be acceptable in one organizational context and unacceptable in another, depending on data sensitivity, domain maturity, stakeholder relationships, and available verification mechanisms. Future studies could therefore compare AI role allocation across domains such as enterprise software, public-sector systems, SaaS products, safety-critical systems, and consumer applications.

The study also points to the need for more precise concepts of trust calibration in early RE. Trust in AI for documentation tasks was much higher than trust in AI for decision-making. This suggests that trust should be studied at the task level rather than as a general user attitude. Future research could examine how practitioners decide whether an AI output is sufficiently reliable, how they detect hallucinations in requirements work, and how verification practices affect trust over time.

Finally, future research should examine how emerging AI environments reshape role boundaries. Interview participants described two possible directions: personal or organizational knowledge support, and more automated end-to-end workflows. These directions are not mutually exclusive. Research on retrieval-augmented generation, agentic workflows, traceability mechanisms, and human review could clarify when stronger AI involvement improves requirements work and when it simply shifts risk to later stages.

5.3.3 Implications for Academia

For academia, particularly in Requirements Engineering education, the findings suggest that AI literacy should be taught as a question of role allocation and professional responsibility, not only as a question of tool use. Students should learn when AI can productively support early RE work, when its outputs must be verified, and why human judgment remains central in tasks involving stakeholder interpretation, prioritization, negotiation, and accountability.

This has implications for how RE is taught. In addition to established topics such as elicitation techniques, stakeholder communication, analysis, specification, and validation, teaching should include concrete human–AI collaboration scenarios drawn from early RE practice. For example, students can compare AI-generated summaries with original stakeholder material, evaluate AI-suggested clarification questions, inspect missing context in AI-supported analyses, and discuss why AI-generated prioritization proposals should not be accepted uncritically. Such exercises would help learners develop reflective rather than passive AI use.

The findings also highlight an important educational distinction between producing well-structured requirements text and understanding stakeholder intent. Although AI can often generate fluent and organized outputs, early RE still depends on contextual interpretation, negotiation, and accountable decision-making. Emphasizing this distinction can help future practitioners avoid over-reliance on plausible AI outputs and develop a more critical understanding of AI's appropriate role in Requirements Engineering.

5.4 Chapter synthesis

The discussion above suggests that AI can support early Requirements Engineering, but its role should be calibrated at the task level rather than generalized across elicitation and analysis. AI appears most useful when it helps practitioners process, organize, and draft from existing information that can be reviewed. Its role becomes more limited when requirements work depends on implicit stakeholder intent, sensitive organizational context, competing priorities, or accountability for decisions. Together, these findings point toward a human-centered understanding of AI involvement in early RE, where AI can strengthen efficiency and analytical support, while humans remain responsible for interpretation, stakeholder understanding, verification, and final decisions.

6

Conclusion

This chapter concludes the thesis by summarizing the main findings of the study and outlining directions for future work. The study investigated how roles and responsibilities between humans and AI are delegated in requirements elicitation and requirements analysis tasks, and how practitioners evaluate the benefits, risks, and trust considerations associated with these role configurations. By combining semi-structured interviews with a follow-up survey, the thesis provides both detailed practitioner reasoning and descriptive survey evidence on AI involvement in early Requirements Engineering (RE).

6.1 Summary of Findings

This study examined AI roles in early Requirements Engineering, focusing on requirements elicitation and requirements analysis. These phases are characterized by ambiguity, incomplete stakeholder input, evolving problem understanding, and a strong need for contextual judgment. For this reason, the study did not treat AI adoption as a binary question of whether AI should or should not be used. Instead, it examined how different human–AI role configurations become more or less appropriate depending on the task.

In relation to RQ1, the findings suggest that roles and responsibilities between humans and AI are delegated in a task-contingent way. AI is mainly considered appropriate as an assistant or preliminary processor in bounded, information-heavy, and verifiable tasks, such as summarizing, organizing, and structuring requirement-related material. In contrast, human responsibility remains central in tasks that require stakeholder interaction, interpretation of implicit needs, prioritization, trade-off decisions, and accountability.

The findings therefore suggest that AI roles in early RE are better understood as task-level configurations rather than as one general role. AI can support or partly lead preliminary processing in selected tasks, but humans remain responsible for validating outputs, interpreting stakeholder intent, managing communication, and making final decisions. The survey provides descriptive support for this conclusion by showing that human-only accounted for 67.7% of valid responses for supporting prioritization or trade-off decisions, while AI-only configurations remained rare across elicitation and analysis tasks.

In relation to RQ2, the findings suggest that AI-supported role configurations are evaluated through a conditional logic of usefulness. Practitioners value AI when it improves efficiency, reduces workload, supports faster access to information, and helps produce initial drafts or structured material. However, perceived risks increase when AI receives more autonomy, especially in tasks involving implicit requirements, missing context, sensitive data, or decision responsibility.

AI-supported configurations therefore appear most beneficial and trustworthy when tasks are bounded, verifiable, information-heavy, and low in decision criticality. As tasks become more interpretive, stakeholder-dependent, sensitive, or decision-oriented, practitioners shift responsibility back to humans. This pattern was visible in both interviews and survey results. For example, AI-assisted work received the strongest perceived benefit overall, while AI-only work had the highest risk mean (4.13 on a 1–5 scale). Trust in AI was therefore not a general attitude toward the technology, but a situated judgment shaped by task type, context availability, output verifiability, data sensitivity, and prior experience with AI errors.

Taken together, the findings suggest that AI has meaningful potential in early RE, but that its role must be calibrated rather than generalized. The thesis contributes to the research gap identified in the introduction by offering a preliminary, practitioner-informed task-level role-boundary framework for AI involvement in early RE. This framework maps AI role configurations to different elicitation and analysis tasks and provides guidance for when AI can support requirements work and when human judgment should remain dominant. In this sense, the central contribution of the study is a human-centered interpretation of AI role allocation: AI can improve efficiency and support preliminary requirements work, but reliable early RE still depends on human interpretation, stakeholder understanding, and accountable decision-making.

6.2 Future Work

Future work can extend this study by validating, observing, and operationalizing the task-level role-boundary framework developed in this thesis. Since AI tools and organizational practices are developing rapidly, further research should focus not only on whether AI capabilities improve, but also on how responsibility boundaries can be made usable in real requirements work.

6.2.1 Future Work for Research

First, future research should externally validate the proposed role configurations with larger and more diverse samples. The present study provides exploratory evidence based on six formal interviews and 67 valid survey responses. Broader studies could examine whether the same patterns hold across different countries, industries, organization sizes, and project types. In particular, comparative studies across domains with different levels of risk and regulation, such as public-sector systems, safety-critical systems, enterprise software, SaaS products, and consumer-facing applications, could clarify how context affects acceptable AI roles.

Second, future research should collect in-situ evidence through case studies and field observations. This study relied on interviews and survey responses, which capture practitioner perceptions and reported practices. Direct observation of AI-supported requirements work could show how practitioners actually use AI during elicitation meetings, document analysis, backlog refinement, prioritization, and stakeholder negotiation. Such studies could also reveal how team culture, tool access, data policies, and internal knowledge bases influence role boundaries in practice.

Third, future research should operationalize this task-level role-boundary framework into practical decision aids and evaluate these aids in real projects. Such decision aids could help teams decide when AI should be excluded, used as an assistant, or used in an AI-led configuration for preliminary processing, while keeping human accountability explicit. This direction could also examine how emerging agentic tools, retrieval-augmented generation, and domain-specific knowledge bases affect acceptable AI involvement. The key question is not only whether these technologies make AI more capable, but whether they improve trust calibration, verification practices, and accountability in requirements work.

6.2.2 Future Directions for Practice

For practitioners, future directions should focus on translating task-level role allocation into concrete organizational practices. One important step is to develop lightweight guidelines for AI use in requirements elicitation and analysis. These guidelines should distinguish between tasks suitable for AI assistance and tasks that require strong human control, especially stakeholder negotiation, prioritization, conflict resolution, and final decisions.

Organizations should also establish clear verification procedures for AI-generated requirements artifacts. The findings show that practitioners already rely on practices such as checking original materials, consulting domain documentation, comparing outputs, and asking stakeholders or team members for confirmation. These practices could be formalized into review checklists, team conventions, or workflow steps, so that AI-generated summaries, candidate requirements, conflicts, or priorities are treated as inputs for human review rather than final conclusions.

Finally, practitioners should invest in secure, context-aware AI environments and AI literacy for requirements work. Many barriers identified in this study concern confidentiality, client data, internal terminology, and missing organizational context. Private tools, approved internal models, curated project knowledge bases, and clear data-handling rules may make AI support more useful and safer. At the same time, practitioners need training to recognize hallucinations, identify missing context, evaluate plausible but incorrect outputs, and decide when human judgment must override AI suggestions.

Overall, the most appropriate path forward appears not to be to replace humans with AI in early Requirements Engineering, but to define clearer task-level role boundaries. Such boundaries can allow AI to improve efficiency and support preliminary requirements work while preserving the human judgment required for stakeholder understanding, contextual interpretation, negotiation, and accountable requirements decisions.

Bibliography

[1] C. Palomares, X. Franch, C. Quer, P. Chatzipetrou, L. López, and T. Gorschek, “The state-of-practice in requirements elicitation: An extended interview study at 12 companies,” *Requirements Engineering*, vol. 26, no. 2, pp. 273–299, 2021, https://doi.org/10.1007/s00766-020-00345-x. [Online]. Available: https://doi.org/10.1007/s00766-020-00345-x

[2] M. A. Zadenoori, L. Zhao, W. Alhoshan, and A. Ferrari, “Automatic prompt engineering: The case of requirements classification,” in *Requirements Engineering: Foundation for Software Quality*, A. Hess and A. Susi, Eds. Cham: Springer Nature Switzerland, 2025, pp. 217–225. [Online]. Available: https://doi.org/10.1007/978-3-031-88531-0_15

[3] H. Cheng, J. H. Husen, Y. Lu, T. Racharak, N. Yoshioka, N. Ubayashi, and H. Washizaki, “Generative AI for Requirements Engineering: A Systematic Literature Review,” *Software: Practice and Experience*, vol. 56, no. 2, pp. 141–170, Feb. 2026, 10.1002/spe.70029.

[4] K. Pohl, *Requirements Engineering: Fundamentals, Principles, and Techniques*. Berlin, Heidelberg: Springer, 2025. [Online]. Available: https://link.springer.com/book/9783662692042

[5] R. Parasuraman, T. Sheridan, and C. Wickens, “A model for types and levels of human interaction with automation,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, 2000, 10.1109/3468.844354.

[6] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz, “Guidelines for human-ai interaction,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13, 10.1145/3290605.3300233. [Online]. Available: https://doi.org/10.1145/3290605.3300233

[7] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. Los Angeles: SAGE Publications,

2018.

[8] M.-A. Storey, R. Hoda, A. Maciel Paz Milani, and M. T. Baldassarre, "Guiding principles for mixed methods research in software engineering," *Empirical Software Engineering*, vol. 30, no. 5, p. 138, 2025, 10.1007/s10664-025-10629-x. [Online]. Available: https://doi.org/10.1007/s10664-025-10629-x

[9] V. Venkatesh, S. A. Brown, and H. Bala, "Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems," *MIS Q.*, vol. 37, no. 1, p. 21–54, Mar. 2013. [Online]. Available: https://doi.org/10.25300/MISQ/2013/37.1.02

[10] A. Ferrari and A. Esuli, "An nlp approach for cross-domain ambiguity detection in requirements engineering," *Automated Software Engg.*, vol. 26, no. 3, p. 559–598, Sep. 2019, 10.1007/s10515-019-00261-7. [Online]. Available: https://doi.org/10.1007/s10515-019-00261-7

[11] L. M. Rani, R. B. Svensson, and R. Feldt, "Ai for requirements engineering: Industry adoption and practitioner perspectives," in *2025 40th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, 2025, pp. 244–251, 10.1109/ASEW67777.2025.00053.

[12] G. Paiva, E. D. Canedo, and G. P. Rocha Filho, "From issue titles to requirements: an empirical study of large language models and prompt engineering strategies," *Requirements Engineering*, vol. 31, no. 1, p. 7, 2026, 10.1007/s00766-026-00462-z. [Online]. Available: https://doi.org/10.1007/s00766-026-00462-z

[13] M. A. Sami, M. Waseem, Z. Zhang, Z. Rasheed, K. Systä, and P. Abrahamsson, "Ai based multiagent approach for requirements elicitation and analysis," 2024. [Online]. Available: https://arxiv.org/abs/2409.00038

[14] A. Ferrari and P. Spoletini, "Formal requirements engineering and large language models: A two-way roadmap," *Information and Software Technology*, vol. 181, p. 107697, 2025, https://doi.org/10.1016/j.infsof.2025.107697. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584925000369

[15] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004, pMID: 15151155. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1518/hfes.46.1.50_30392

[16] B. Lubars and C. Tan, "Ask not what ai can do, but what ai should do: towards a framework of task delegability," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, vol. 32. Red Hook, NY, USA: Curran Associates Inc., 2019, pp. 57–67, 10.5555/3454287.3454293.

[17] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997, 10.1518/001872097778543886. [Online]. Available: https://doi.org/10.1518/001872097778543886

[18] A. Jacovi, A. Marasović, T. Miller, and Y. Goldberg, "Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 624–635, 10.1145/3442188.3445923. [Online]. Available: https://doi.org/10.1145/3442188.3445923

[19] E. Glikson and A. W. Woolley, "Human trust in artificial intelligence: Review of empirical research," *Academy of Management Annals*, vol. 14, no. 2, pp. 627–660, 2020, 10.5465/annals.2018.0057. [Online]. Available: https://api.semanticscholar.org/CorpusID:216198731

[20] E. Toreini, M. Aitken, K. Coopamootoo, K. Elliott, C. G. Zelaya, and A. van Moorsel, "The relationship between trust in ai and trustworthy machine learning technologies," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, ser. FAT* '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 272–283, 10.1145/3351095.3372834. [Online]. Available: https://doi.org/10.1145/3351095.3372834

[21] S. Mehrotra, C. Degachi, O. Vereschak, C. M. Jonker, and M. L. Tielman, "A systematic review on fostering appropriate trust in human-ai interaction: Trends, opportunities and challenges," *ACM J. Responsib. Comput.*, vol. 1, no. 4, Nov. 2024, 10.1145/3696449. [Online]. Available: https://doi.org/10.1145/3696449

[22] A. Klingbeil, C. Grützner, and P. Schreck, "Trust and reliance on ai — an experimental study on the extent and costs of overreliance on ai," *Computers in Human Behavior*, vol. 160, p. 108352, 2024, 10.1016/j.chb.2024.108352. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563224002206

[23] M. Steyvers, H. Tejeda, A. Kumar, C. Belem, S. Karny, X. Hu, L. W. Mayer, and P. Smyth, "What large language models know and what people think they know," *Nature Machine Intelligence*, vol. 7, no. 2, pp. 221–231, 2025, 10.1038/s42256-024-00976-7. [Online]. Available: https://doi.org/10.1038/s42256-024-00976-7

[24] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," *International Journal of Human–Computer Interaction*, vol. 36, no. 6, pp. 495–504, 2020. [Online]. Available: https://doi.org/10.1080/10447318.2020.1741118

[25] M. A. Zadenoori, J. Dąbrowski, W. Alhoshan, L. Zhao, and A. Ferrari,

"Large language models (llms) for requirements engineering (re): A systematic literature review," 2025. [Online]. Available: https://arxiv.org/abs/2509.11446

[26] K. Malterud, V. D. Siersma, and A. D. Guassora, "Sample size in qualitative interview studies: Guided by information power," *Qualitative Health Research*, vol. 26, no. 13, pp. 1753–1760, 2016, 10.1177/1049732315617444. [Online]. Available: https://doi.org/10.1177/1049732315617444

[27] M. Hennink and B. N. Kaiser, "Sample sizes for saturation in qualitative research: A systematic review of empirical tests," *Social Science & Medicine*, vol. 292, p. 114523, 2022, https://doi.org/10.1016/j.socscimed.2021.114523. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0277953621008558

[28] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006, 10.1191/1478088706qp063oa. [Online]. Available: https://doi.org/10.1191/1478088706qp063oa

[29] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering (2024 edition)*. Springer Berlin, Heidelberg, 2024, 10.1007/978-3-662-69306-3. [Online]. Available: https://doi.org/10.1007/978-3-662-69306-3

[30] D. R. Hess, "How to write an effective discussion," *Respiratory Care*, vol. 68, no. 12, pp. 1771–1774, 2023, 10.4187/respcare.11435. [Online]. Available: https://doi.org/10.4187/respcare.11435

A

Interview Guide

This appendix presents the interview guide used in this study. The interviews were semi-structured, meaning that the questions below were used as guiding questions rather than as a fixed script. Follow-up questions were asked depending on the participant's role, experience, and previous answers.

Introduction

This study investigates how software industry practitioners make decisions about the use of artificial intelligence (AI), particularly Generative AI (GenAI) tools, in early Requirements Engineering (RE). The study focuses on requirements elicitation and requirements analysis, and examines how roles and responsibilities are allocated between humans and AI in these tasks. It also explores the perceived benefits, risks, trust, verification practices, and responsibility issues related to AI-supported requirements work.

Participation in this study is voluntary. Interview recordings and transcripts will be anonymized to remove any personally identifiable information. Only the research team will have access to the raw data. Any publications or presentations based on this study will use aggregated findings or anonymized quotations to protect participants' privacy.

1. Participant Background

- Could you briefly describe your current role, years of experience, and main responsibilities?
- What type of company or domain do you work in? Please describe the organizational context without naming the company.
- What requirements-related activities are you usually involved in?
- Have you used AI tools in your work? If yes, for what kinds of tasks and for how long?

2. Requirements Elicitation and Requirements Analysis

Before asking the following questions, the interviewer briefly explained the two concepts used in this study:

- **Requirements elicitation** refers to the process of discovering, collecting, and capturing requirements from different sources.
- **Requirements analysis** refers to the process of understanding, organizing, evaluating, and refining requirements after they have been collected.

2.1 Requirements Elicitation

- What channels or methods do you usually use to collect requirements?
- Which elicitation activities are the most important or time-consuming in your work?
- What are the main challenges you experience during requirements elicitation?
- Do different types of requirements require different elicitation methods? If yes, could you give an example?
- Have you used AI tools in requirements elicitation?
  - If yes: For which tasks? What tools did you use? Why did you use AI, and what problems did it help solve?
  - If no: Why have you not used AI in this stage?
- What factors influence whether you would use AI for a requirements elicitation task?

2.2 Requirements Analysis

- After requirements are collected, how do you usually analyze them?
- How do you identify conflicts, overlaps, ambiguities, or inconsistencies between requirements?
- How do you usually prioritize requirements?
- Which analysis activities are the most important or time-consuming in your work?
- What are the main challenges you experience during requirements analysis?

- Have you used AI tools in requirements analysis?
  - If yes: For which tasks? What tools did you use? Why did you use AI, and what problems did it help solve?
  - If no: Why have you not used AI in this stage?
- What factors influence whether you would use AI for a requirements analysis task?
- Is AI currently integrated into your requirements workflow? If yes, in which parts of the workflow is it used?
- If AI is not yet integrated, or only used informally, under what conditions would your team consider deeper integration of AI into the requirements workflow?

3. Human-AI Role Allocation

Participants were asked to reflect on specific requirements elicitation and analysis tasks mentioned earlier in the interview. For selected tasks, the following prompts were used when relevant.

3.1 Role Configuration

- For this task, how do you think humans and AI should divide responsibilities?
- Would you see this task as more suitable for:
  - human-only work,
  - AI-assisted work where humans remain in control,
  - AI-led work where humans review or approve the result, or
  - AI-only work?

Depending on the participant's answer, the following follow-up questions were used:

- If the task should remain human-only: Why do you think AI is not suitable for this task?
- If AI should assist humans: What could AI help with, and what should still be judged by humans?
- If AI could lead the task: What would humans need to review or approve?
- If the task could be fully handled by AI: Why do you think human review would not be necessary, and how would quality be ensured?

3.2 Tool Choice and Practical Use

For tasks involving AI use or potential AI use, the following prompts were used:

- Which AI tool would you use for this task, and why?
- What are the strengths and weaknesses of this tool in this task?
- Would you combine the AI tool with other methods or tools?
- If you have already used AI for this task, how well did it work?
- What problems or limitations have you experienced?

3.3 Benefits and Risks

- What benefits could this human-AI role allocation bring?
- What risks, concerns, or limitations would you associate with this role allocation?

4. Trust, Verification, and Responsibility

For tasks where AI was used or considered, the following prompts were used when relevant:

- How much would you trust AI-generated results in this task?
- What factors make you trust or distrust AI-generated results?
- How would you check or verify AI-generated results before using them?
- Have you experienced serious errors or unstable outputs from AI tools? If yes, how did you handle them?
- Who should be responsible if AI-supported requirements work leads to errors or misunderstandings?

For tasks where AI was not used or was considered unsuitable, the following prompts were used:

- Why do you think this task is not suitable for AI?
- What conditions would need to be met before you would consider using AI for this task?

5. Summary and Future Expectations

- Do you think the role of AI should be different in requirements elicitation compared with requirements analysis? Why?
- Are there any requirements elicitation or analysis tasks that should remain human-led even if AI improves? Why?
- Could you describe a recent requirements-related project and explain where AI was used or could have been used across the elicitation and analysis process?
- Looking ahead, how do you think the role of AI in Requirements Engineering should develop?

Optional follow-up prompts:

- Which tasks should always remain human-led?
- Which tasks could be delegated more to AI in the future?
- What new AI capabilities would be most useful for requirements work?
- What concerns you most about deeper AI involvement in requirements work?
- Should AI's role in requirements tasks be fixed, or should it change depending on the task and context?

Closing Question

- Is there anything important about AI use in early requirements work that we have not discussed but you think should be mentioned?

B

Survey Questionnaire

The complete survey questionnaire used in this study is included below.

Human–AI Role Allocation in Early Requirements Engineering: Practitioner Survey

Purpose of the Survey

This survey aims to understand how human and AI roles are allocated in early requirements elicitation and analysis tasks, and to examine the benefits, risks, and trust associated with different forms of human–AI collaboration. The results will be used for academic research as part of a master's thesis in software engineering.

Participation

Your participation is entirely voluntary. You may choose not to answer any question and may withdraw from the survey at any time before submission without providing a reason. There are no right or wrong answers. We are interested in your honest and independent perspective.

Confidentiality and Data Protection

All responses will be treated as strictly confidential and will be used only for research purposes. Except for the optional email address provided for follow-up interview coordination, no personally identifiable information will be collected. The data will be stored securely and reported only in aggregated or anonymized form.

Time Required

The survey takes approximately 10–12 minutes to complete.

Consent

By proceeding with this survey, you confirm that you are at least 18 years old and that you voluntarily agree to participate in this study under the conditions described above.

Contact Information

If you have any questions or concerns about this study, you may contact the researchers via the email below. (gusliyup@student.gu.se) (keyu@chalmers.se)

Optional follow-up interview

If you are willing to participate in a follow-up interview, you may optionally leave your email address below.

Your email will be used **only for interview coordination** and **will not be linked to your survey responses**.

Q0-1 Email (Optional)

____

Q0 I have read the information above and agree to participate in this study.

- ◯ Yes
- ◯ No

Participants who selected "No" were directed to the end of the survey.

Q1 What is your primary role in software development?

- ◯ Requirements Engineer
- ◯ Product Manager / Product Owner
- ◯ Software Developer
- ◯ Project Manager
- ◯ UX / Interaction Designer
- ◯ Business Analyst / System Analyst
- ◯ Tech Lead / Architect
- ◯ QA / Test Engineer
- ◯ Other ____

Q2 Years of professional experience

- ◯ Under 1 year
- ◯ 1-2 years
- ◯ 3-6 years
- ◯ 7-10 years
- ◯ More than 10 years

Q3 Organization Type

- ◯ Startup
- ◯ Small and Medium Enterprises
- ◯ Large Enterprise
- ◯ Research/ Academic Organization
- ◯ Other ____

Q4 Which stages of Requirements Engineering are you involved in?

- ☐ Requirement Elicitation *(e.g., gathering needs from stakeholders, interviews, workshops)*
- ☐ Requirement Analysis *(e.g., clarifying, structuring, resolving conflicts in requirements)*
- ☐ Requirement Specification *(e.g., documenting requirements in structured formats such as user stories or use cases)*
- ☐ Requirement Validation *(e.g., verifying and confirming requirements with stakeholders to ensure correctness, completeness, and feasibility)*

Q5 Do you use AI tools (such as Generative AI) in requirements elicitation or analysis tasks? If yes, please specify which tools you use.

- ◯ Yes ____
- ◯ Yes, occasionally ____
- ◯ No

Displayed only if the participant selected "No" in Q5.

Q5a If you do not use Generative AI in your work, what are the main reasons? (Select all that apply)

- ☐ Data security or confidentiality concerns
- ☐ Lack of integration with existing workflows
- ☐ Do not trust AI outputs
- ☐ AI does not provide useful results for my tasks
- ☐ Lack of organizational support or guidelines
- ☐ Do not know how to effectively use AI tools
- ☐ No clear need for AI in my work
- ☐ Other ________________________________

Displayed only if the participant selected "No" in Q5.

Q5b Even though you do not currently use AI, what is your general perception of AI-generated outputs in requirements-related tasks?

____

____

After answering Q5b, participants were directed to the end of the survey.

Q6 Which tasks do you use AI for in your work? (Select all that apply)

- ☐ Summarizing meetings / documentation
- ☐ Extracting requirements from raw data
- ☐ Identifying ambiguities or missing information
- ☐ Structuring / organizing requirements
- ☐ Detecting conflicts or inconsistencies
- ☐ Supporting background research / information search
- ☐ Supporting prioritization or decision-making
- ☐ Breaking down requirements into tasks
- ☐ Other ____

Q7 The following questions (Q7a and Q7b) ask how roles are distributed across different early RE tasks. Use the following definitions when selecting your answers.

Human-only: *The task is performed entirely by a human, without using AI.*
AI-assisted: *AI supports the task, but the human performs the work and makes all decisions.*
AI-led: *AI performs most of the task, while the human supervises and may adjust the results.*
AI-only: *The task is performed entirely by AI, without meaningful human involvement.* You can select multiple role configurations for a single task, if applicable.

Q7a Requirement Elicitation Tasks

	Human-only	AI-assisted	AI-led	AI-only
Summarizing elicitation sessions	☐	☐	☐	☐
Extracting requirements from raw inputs	☐	☐	☐	☐
Identifying ambiguities or missing information	☐	☐	☐	☐
Suggesting clarification questions	☐	☐	☐	☐

Q7b Requirement Analysis Tasks

	Human-only	AI-assisted	AI-led	AI-only
Structuring and standardising requirements	☐	☐	☐	☐
Structuring and organizing requirements	☐	☐	☐	☐
Detecting conflicts or inconsistencies	☐	☐	☐	☐
Supporting prioritization or trade-off decisions	☐	☐	☐	☐

The following questions explore practitioners' perceptions of the ***Benefits, Risks,*** and ***Trust*** related to using AI (such as Generative AI) in early Requirements Engineering (RE) tasks, particularly in requirements elicitation and analysis.

Q8 Under which conditions do you consider AI outputs to be reliable? (Select all that apply)

- ☐ When the task is simple or well-structured
- ☐ When the task is repetitive or high-volume
- ☐ When sufficient input/context is provided
- ☐ When using domain-specific prompts or knowledge
- ☐ When outputs can be easily verified
- ☐ When using multiple tools for cross-checking
- ☐ When the task does not involve critical decisions
- ☐ I generally do not consider AI outputs reliable
- ☐ Other ____

*

Q9 Think about the ***Benefits*** that may arise during ***early RE tasks (elicitation and analysis)*** under different role configurations. For each benefit below, rate how strongly it applies under each configuration. Use the scale: 1 = No benefit – 5 = Strong benefit

	Human only	AI assisted	AI-led	AI only
Time saving				
Better documentation (summaries, notes, traceability)				
Improved consistency (formatting, structure, terminology)				
Faster generation of initial drafts (e.g., first versions, initial structures)				
Better organization of dispersed information sources				

Better coverage (identifying missing information or overlooked aspects)				
Support for ideation (suggesting clarifying questions)				
Reduced workload / cognitive load				
Faster information access (e.g., background research)				

Q10 Think about the ***Risks*** that may arise during ***early RE tasks (elicitation and analysis)*** under different role configurations. For each risk below, rate how likely it is to occur under each configuration. Use the scale: 1 = Very unlikely – 5 = Very likely

	Human only	AI assisted	AI-led	AI only
Misinterpretation of stakeholder intent				
Hallucinated or fabricated details				
Loss of context/nuance (domain, business constraints)				
Difficulty in understanding implicit requirements or unexpressed stakeholder intentions (e.g., implicit expectations)				
Reduced quality of communication and interaction with stakeholders				

Over-reliance (reduced critical thinking)				
Confidentiality / data leakage concerns				
Lack of explainability / unclear reasoning				
Accountability ambiguity (who is responsible?)				
Inconsistent results, which may affect judgment and decision-making				
Bias or unfair assumptions in results				
No major risk				

Q11 When using AI-generated outputs, how do you typically use them in your work? (Select the option that best describes your usual practice)

- ○ Use directly
- ○ Modify slightly
- ○ Heavily revise
- ○ Use only as reference

Q12a In which situations do you verify AI outputs? (Select all that apply)

- ☐ Important decisions
- ☐ Uncertain outputs
- ☐ Sensitive data
- ☐ Domain-specific tasks
- ☐ I will verify under any circumstances
- ☐ I never verify AI output.
- ☐ Other ____

Q12b How do you typically verify AI outputs? (Select all that apply)

- ☐ Cross-check with original materials (transcripts/notes/docs)
- ☐ Ask stakeholders / team members for confirmation
- ☐ Use another AI tool/model to cross-check
- ☐ Consult domain documentation / standards
- ☐ I do not verify
- ☐ Other ____

Q13 Please rate agreement (Strongly disagree → Strongly agree):

	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
I trust AI in documentation-related tasks.	○	○	○	○	○	○
I trust AI in requirements analysis tasks (e.g., categorization, identifying conflicts).	○	○	○	○	○	○
I trust AI in decision-making processes.	○	○	○	○	○	○
Before using AI-generated outputs, I typically verify or modify them.	○	○	○	○	○	○
I feel confident using AI outputs even without extensive verification.	○	○	○	○	○	○
I believe AI outputs are generally reliable.	○	○	○	○	○	○
AI-generated content may sometimes be inaccurate or inconsistent.	○	○	○	○	○	○
I do not fully rely on AI outputs in important tasks.	○	○	○	○	○	○
When making critical decisions, I tend to rely more on human judgment than on AI.	○	○	○	○	○	○

In my daily work, I feel comfortable relying on AI.	○	○	○	○	○	○
I believe AI is better suited as a supporting tool rather than for independently completing tasks.	○	○	○	○	○	○

Q14 Why do you choose not to use AI in some tasks? (Optional)

____

____

____

Q15 Have you experienced problems when using AI? (Optional)

____

____

____