**A blue and white logo

Description automatically generated**

**VILNIUS GEDIMINAS TECHNICAL UNIVERSITY**

**FACULTY OF FUNDAMENTAL SCIENCES**

**DEPARTMENT OF INFORMATION SYSTEMS**

**Ansu Mary Jacob (20240045)**

**AI-Enabled Requirements Engineering Specification using RAG**

**Master Graduation Thesis**

**Guided by: Assoc. prof. dr. Jolanta Miliauskaitė**

**VILNIUS 2025**

**TABLE OF CONTENT**

[1 INTRODUCTION 7](#_Toc207261726)

[1.1 Investigation Object 9](#_Toc207261727)

[1.2 The Aim of the Thesis 9](#_Toc207261728)

[1.3 The Tasks of the Thesis 9](#_Toc207261729)

[1.4 Novelty of the Topic 9](#_Toc207261730)

[1.5 Relevance of the Topic 10](#_Toc207261731)

[1.6 Research Methodology 10](#_Toc207261732)

[1.7 Scientific Value of the Thesis 10](#_Toc207261733)

[1.8 Main Results of the Thesis 11](#_Toc207261734)

[1.9 Structure of the Work 11](#_Toc207261735)

[2 Related Works Analysis 12](#_Toc207261736)

[2.1 Main Concept of Requirements Engineering Specification 12](#_Toc207261737)

[2.2 Artificial Intelligence for Requirement Engineering Specification 14](#_Toc207261738)

[2.3 Overview of Requirements Engineering Practices for AI in Literature 16](#_Toc207261739)

[2.4 Summary of research papers 20](#_Toc207261740)

[2.5 Summary of the 2nd section 25](#_Toc207261741)

[2.6 AI Algorithm Comparison Table for NLP-based RES 26](#_Toc207261742)

[2.7 Open Dataset Assessment & Review 27](#_Toc207261743)

[2.8 Summary of AI algorithms and methods 28](#_Toc207261744)

[3 Proposed Approach 30](#_Toc207261745)

[3.1 Describing the Process using BPMN Diagrams 31](#_Toc207261746)

[3.1.1 Main Process of Workflow 31](#_Toc207261747)

[3.1.2 Subprocess of Text Normalization & Cleaning 32](#_Toc207261748)

[3.1.3 Subprocess of Knowledge Base setup 33](#_Toc207261749)

[3.1.4 Subprocess of LLM – based Extraction 34](#_Toc207261750)

[4 Result Measure Metrics 36](#_Toc207261751)

[5 Initial Experiment & Analysis 37](#_Toc207261752)

[5.1 Domain Selection & Justification 37](#_Toc207261753)

[5.2 Technical Feasibility Analysis Based on the Dataset 38](#_Toc207261754)

[5.3 Initial Experiment Results 39](#_Toc207261755)

[6 General Conclusion 40](#_Toc207261756)

[7 Future Work Plan 42](#_Toc207261757)

[8 References 43](#_Toc207261758)

**LIST OF IMAGES**

[Figure 2.1 Requirement Engineering Process 12](#_Toc207261759)

[Figure 3.1 General Schema of Proposed approach 30](#_Toc207261760)

[Figure 3.2 Main Process of Workflow 32](#_Toc207261761)

[Figure 3.3 Subprocess of Text Normalisation & Cleaning 33](#_Toc207261762)

[Figure 3.4 Subprocess of Knowledge Base Setup 34](#_Toc207261763)

[Figure 3.5 Subprocess of LLM - Based Extraction 35](#_Toc207261764)

[Figure 5.1 Banking Chatbot Dataset Sample 38](#_Toc207261765)

[Figure 5.2 Extracted Functional Requirement 39](#_Toc207261766)

**LIST OF TABLES**

[Table 2.1 Summary of Articles 21](#_Toc207261767)

[Table 2.2 Review of AI Algorithm 26](#_Toc207261768)

[Table 2.3 Review of the Open Datasets of RES 27](#_Toc207261769)

**ABBREVIATIONS**

RE – Requirements Engineering

RES – Requirements Engineering Specification

R&D – Research and Development

SLR – Systematic literature review

QoS – quality of service

CPS – Cyber-Physical Systems

VR – virtual reality

SVM – Support Vector Machine

CNN – Convolutional Neural Network

SE – Software Engineering

NA – Not Applicable

GDPR - General Data Protection Regulation

DL – Deep Learning

RAG - Retrieval-augmented generation

BERT - Bidirectional Encoder Representations from Transformers

TF-IDF - Term frequency-Inverse Document Frequency

LLM – Large Language Model

FAISS - Facebook AI Similarity Search

# INTRODUCTION

The Requirement Engineering Specification (RES) is the process of preparing clear, consistent, and unambiguous documentation of the requirements identified in the analysis step. This step also includes assigning priority and grouping requirements into manageable chunks (Maalej et al., 2023; Zowghi et al., 2023).

The RES is used to generate a formal software requirements model. All requirements, including functional and non-functional requirements and constraints, are largely specified in these models. In the specification process, more knowledge about the problem may be required, and this will trigger the elicitation process. Models used at this stage include ER diagrams, data flow diagrams (DFD, etc.(Liu et al., 2022; Sofian et al., 2022).

One of its main contributions was to establish clarity and consensus among stakeholders in order to ensure that everyone knew what the system should achieve. This common understanding minimizes confusion and reduces the risk of expectations not being met. RES is also the basis for design and development, as a blueprint for architects, developers, and testers, and guides technical implementation in accordance with user requirements. By clearly defining what is included and excluded in the project, RES supports effective scope management and prevents uncontrolled expansion of project objectives, commonly known as scope creep. In addition, it increases quality by identifying errors, inconsistencies, or missing information at the beginning of the life cycle when it is less costly to resolve. RES also facilitates testing by creating test cases that correspond directly to the specified requirements to ensure that the system works as expected. Through the proper documentation, traceability is made easier to manage and evaluate changes. Finally, RES supports system maintenance and becomes an essential component of long-term software sustainability by providing a reference for future updates, improvements, or audits (Ahmad et al., 2023; Maalej et al., 2023; Sporsem, 2024).

The main problems of the RES have a significant impact on the success of software projects. First, unintended customer requirements often lead to vague or incomplete specifications and make it difficult for engineers to provide the right solution. Changes in requirements due to changes in business requirements or client insights have led to scope creep and rework. The lack of technical knowledge among engineers prevents proper interpretation and validation of the requirements. In addition, unclear and inconsistent requirements force developers to make assumptions, leading to systems that do not meet actual needs. Long feedback cycles delay validation and slow down the entire development process. Re-use of outdated designs without considering current contexts brings about mismatches and design flaws (Zowghi et al., 2023).

The rapid integration of Artificial Intelligence (AI) into several sectors emphasizes the need to adapt requirements engineering (RE) procedures to the unique requirements of AI systems. AI systems behave in a non-deterministic manner, evolve over time, and depend heavily on large datasets, unlike traditional software systems (Solomonides et al., 2022). The conventional RE originally designed for static and predictive systems faces major challenges.

The main challenge is that traditional R&D methods cannot address the dynamic, data-driven inherent in the development of AI systems. As pointed out in the analysis of current RE practices, AI systems require specialized approaches to ensure ethical compliance, manage evolving requirements and align technical capabilities with the expectations of stakeholders (Maalej et al., 2023). The lack of robust tools and frameworks exacerbates problems such as AI system bias, lack of transparency and inconsistencies due to repeated development processes (Zowghi et al., 2023).

Therefore, the main problem analysed in the thesis is the following:

**How can requirement engineering (RE) Specification processes be improved using AI-based development?**

Traditional RE practices lack flexibility and scalability to handle the iterative nature and evolution of AI systems, which are continuously influenced by changes in data sets and algorithms. This raises key questions about how AI systems should align with stakeholder expectations while maintaining robustness and transparency over time (Taj et al., 2025).

This research seeks to address these challenges by studying new methodologies and frameworks specifically adapted for the development of AI systems. Analysing the existing gaps in the RE process, the study proposes methods to promote consistency, reduce bias and improve the trackability of AI decision-making processes. In addition, the aim of the research is to develop three key tools and strategies that help RE practitioners manage the unique requirements of AI systems and ensure reliability and transparency (Ahmad, Abdelrazek, Arora, Bano, et al., 2023b).

## Investigation Object

The object of the investigation is the methods to adapt requirements engineering (RE) specification practices to effectively address the specific needs of AI systems, with a focus on data-driven requirements and iterative development (Maalej et al., 2023; Zowghi et al., 2023).

## The Aim of the Thesis

The aim is to improve the requirements engineering (RE) specification processes by applying AI-based development.

## The Tasks of the Thesis

To achieve the aim of the thesis, the following tasks have been defined:

1. To analyse existing methods for RE and AI-based development.
2. To propose a novel RES approach that addresses the specific needs of AI systems, incorporating stakeholder involvement and iterative requirements management.
3. To develop a prototype implementation of the proposed method, integrating tools and methodologies for managing technical and data-centric requirements, and to test the effectiveness of the proposed framework through practical case studies and experimental validation. Implement & validate the framework through case studies and experimental analysis.

## Novelty of the Topic

In addressing critical gaps in traditional RES with AI systems, such as the inability to effectively address dynamic requirements and data-driven processes (Maalej et al., 2023). The research introduced a customized RE framework incorporating social and technical considerations, providing new knowledge and practical solutions to fill these gaps in the development of AI systems (Zowghi et al., 2023). By taking into account the views of stakeholders, we ensure that the AI system complies with society's expectations and values, fostering trust and user acceptance. The research emphasizes the importance of data governance and quality and addresses the gaps in existing frameworks that do not address the technical challenges associated with large data sets. The practical validation of the framework by case studies shows its application in the fields of various AI, ensuring relevance both in the academic and industrial fields. This research also addresses the gap between theoretical considerations and their practical application in AI development workflows(Ahmad et al., 2023). The new features are the ability to manage non-deterministic behaviours in AI systems, which often lead to unpredictable results in traditional software engineering methods. It provides a structured approach to managing the iterative development cycle and ensures the consistent alignment of the initial requirements despite the development of the system's capabilities. The proposed framework provides the basis for the development of regulatory compliant AI systems to respond to the growing demands of frameworks such as the EU AI Act. Finally, research contributes to a scalable and adaptive solution that integrates technical precision, setting the benchmark for future advances in the engineering of AI requirements.

## Relevance of the Topic

The importance of this topic is underlined by the increasing integration of AI systems into critical areas such as health care, autonomous systems, and finance, which require trust, transparency, and accountability. AI systems are often used in high-risk environments where mistakes or prejudices have serious social consequences, which makes the need for rigorous RE practices critical(Solomonides et al., 2022).

Furthermore, regulatory frameworks such as the EU AI Act require compliance with the standards of AI systems, creating an urgent need for RE methodology that aligns technical development with these regulations (Maalej et al., 2023). As (Zowghi et al., 2023) pointed out, addressing these challenges is not only the key to the development of responsible AI systems, but also to fostering the trust and acceptance of users.

Traditional requirement engineering (RE) specification practices are insufficient to address the complexity of artificial intelligence systems, including the dependency on large-scale data sets, iterative development, and non-deterministic behaviours (Siddeshwar et al., 2024). Therefore, this research proposes RES approach for AI, using AI techniques to directly address these pressing problems and ensure the development of a technically robust and socially responsible system.

## Research Methodology

This study used a systematic literature review (SLR) and comparative analysis to identify gaps and limitations in existing RES approaches and adapted AI system methods(Habiba et al., 2024; Siddeshwar et al., 2024). For the development of the methodology, logical induction and generalization are used to conceptualize a novel RES approach that integrates stakeholder involvement and technical requirements.

## Scientific Value of the Thesis

This research proposes and validates a new approach that addresses the requirements engineering specification of AI development, addresses their dynamic complexity, data-driven behaviour, and non-deterministic nature(Maalej et al., 2023; Zowghi et al., 2023). It provides practical solutions for ensuring transparency, aligning technical development with stakeholder expectations(Solomonides et al., 2022). The approach improves stakeholder involvement, iterative requirements management, and AI-based considerations like data quality and traceability (Nalchigar et al., 2021). It bridges the gap between traditional RES practices and the real-world demands of AI development, contributing both theoretical insights and applied methods.

The effectiveness of the approach must be measured based on accuracy and completeness are essential to ensure the approach successfully captures all relevant FR & NFR requirements. The adaptability must be assessed, focusing on its ability to manage the evolving requirements that arise from changes in data and user expectations(Wang et al., 2024). The effectiveness should be validated through case studies or experiments that demonstrate how well it performs and what outcomes it gives in AI development. The stakeholder satisfaction must be measured to ensure that the approach leads to meet the user needs, build trust, and the support long term system accessibility.

## Main Results of the Thesis

The research results in requirement engineering (RE) specification method for AI development, which addresses dynamic, iterative development and data-based requirements. The approach was validated through practical implementation and experimentation and demonstrated its effectiveness in managing specific challenges associated with AI and aligning with the expectations of stakeholders.

## Structure of the Work

The structure of the paper is as follows:

The second section analyses existing RE approaches for AI development; the third section presents the proposed new RE approach; the fourth section describes the implementation and validation of the method.

# Related Works Analysis

## Main Concept of Requirements Engineering Specification

Requirement Engineering (RE) is a field of software development that includes the identification, registration, verification, and management of system requirements. It ensures that the developed systems meet both the needs of users and stakeholders and align with organizational objectives. RES usually focuses on functional requirements (system functions) and non-functional requirements (performance, security, usability, etc.). In software engineering, RES is the basis for bridging the gap between expectations of stakeholders and technical implementation. However, the nature of the RES process is manual and labour-intensive, and the complexity of modern systems is increasing, often leading to inefficiency and inconsistency in the specifications. This challenge (Liu et al., 2022; Sofian et al., 2022)offers opportunities to automate and improve RES practices using advanced technologies such as artificial intelligence (AI).

The Fig 2.1 RE Process:

A diagram of a system

AI-generated content may be incorrect.

Figure 2.1 Requirement Engineering (RE) Process

The input and output of the RES process are crucial to ensuring the smooth transition from the analysis of requirements to system design and implementation (Ahmad et al., 2023; Nalchigar et al., 2021) . RES inputs usually contain information collected during the development and analysis stages. These inputs include user requirements and expectations, functional and non-functional requirements, business objectives, domain knowledge, legal and regulatory constraints, and support artifacts such as user cases, interviews, and existing system documentation(Ramachandran et al., 2024). Furthermore, data flow diagrams (DFDs), entity-related diagrams (ERDs), and business process models are structured inputs to facilitate specification clarity.

The output of the RES process is a formalized, structured, and verified document that is the basis for system development. Formal Requirements Specification (FRS) is a mathematically precise way of defining software requirements. Unlike informal or semi-formal methods, formal specifications use logic, set theory, and algebraic expressions to define system behaviour. This approach ensures that the system is unambiguous, consistent, and verifiable before development begins.

The main output is a Software Requirements Specification (SRS) document, which clearly defines all functional and non-functional requirements, system constraints, and assumptions. Other outputs include a requirement model (e.g., a UML diagram), a list of priority requirements, a requirement tracking matrix, and a defined acceptance criteria (Maalej et al., 2023) for validation and testing. These outputs guide designers, developers, and testers to ensure that the system is constructed in accordance with the expectations and objectives of the stakeholders. The (Ahmad et al., 2023) clarity and completeness of these outputs directly affect the quality, traceability, and maintenance of the final system.

One of the main challenges of RES processes is that customers and stakeholders often lack a clear or complete understanding of what they want from the system (Zowghi et al., 2023). They usually express only general goals rather than detailed functional or non-functional requirements (Sporsem et al., 2024). As a result, it is difficult for the requirements engineer to translate vague objectives into precise and actionable specifications. As a result, engineers often have to make assumptions, increasing the risk of misunderstanding between stakeholder expectations and developed systems.

Requirements are rarely fixed. In the development process, they are often changing as a result of new insights, business priorities, or changing market conditions. In the context of AI development (Maalej et al., 2023; Siddeshwar et al., 2024), this challenge is even more pronounced because changes in training data, model behaviour, or regulatory expectations often require system requirements to be updated. Without flexible and adaptive specification processes, such changes can lead to scope creep, rework, and delays in project timelines. (Taj et al., 2025) demonstrate how fine-tuned models like BERT help adapt requirement elicitation to such dynamic settings.

In some cases, software engineers involved in the specification phase of requirements lack sufficient domain knowledge in the application area (such as health care, finance, and autonomous systems). According to (Nalchigar et al., 2021)this gap hinders effective communication with stakeholders and limits the quality of elicited requirements. This lack of expertise makes it difficult to understand and validate the inputs of stakeholders completely or to predict the practical impact. It can result in incomplete or incorrect documentation of requirements, which later leads to system defects or unmet user needs (Habiba et al., 2024).

The requirements are expressed in natural language that is unclear and inconsistent. (Kaur & Kaur, 2024; Liu et al., 2022) report that such unstructured formats introduce inconsistencies and multiple interpretations. A single requirement can be interpreted differently by different readers. Without a formal clarification, developers and testers are obliged to make assumptions that the system does not function as expected.

AI systems are basically non-deterministic and data-oriented, making it complicated to use in traditional RES techniques (Zowghi et al., 2023). It is difficult to ensure their behaviour is consistent with the expectations and intentions of stakeholders. This represents the risks in transparency, accountability, and system reliability.

## Artificial Intelligence for Requirement Engineering Specification

AI is the creation of systems capable of performing tasks that require human intelligence (Sofian et al., 2022). It includes the understanding of natural language, the study of data, reasoning, and decision-making. AI technologies such as machine learning (ML), natural language processing (NLP), and deep learning (DL) have been applied in various fields, including medical, financial, and autonomous systems (Liu et al., 2022; Ramachandran et al., 2024). AI systems are based on algorithms and data to identify patterns, predict, and improve the performance over time. This iterative and data-driven approach makes AI especially suitable for automating complex, repetitive, and data-intensive tasks, such as those described in the RE specification (Ahmad et al., 2023; Nalchigar et al., 2021).

AI and RES are interdependent in several key areas, particularly to improve and automate the processes involved in capturing, analysing, managing, and confirming system requirements. As highlighted by (Kaur et al., 2024), modern AI methods like NLP and ML play a significant role in the classification and refinement of system requirements. Here is how AI and RE are related:

AI is a key part of RES that automates, improves, and optimises many important steps in the life cycle of requirements. One of the main tasks in the RES is to find out the requirements of various stakeholders' needs. It automatically analyses unstructured data sources like conference notes, emails, documents, and even audio transcriptions using NLP and ML technology (Ferrari et al., 2023; Siddeshwar et al., 2024). It then pulls out useful information and turns it into structured requirements. AI reduces human bias, lowers the chance of missing important requirements, and speeds up the process of getting detailed input from stakeholders by automating this process (Ahmad et al., 2023; Habiba et al., 2024).

AI is very important for making sure that the requirements are of high quality once they have been gathered. It does this by finding inconsistencies, ambiguities, and contradictions in the requirements specifications. For instance, AI tools can check requirements against rules or standards that have already been set for the domain and automatically find conflicts or unclear terms (Liu et al., 2022; Taj et al., 2025). This makes sure that the final requirements are correct and consistent. So, AI helps get rid of mistakes that might be missed by manual inspection, which is important to make sure that the system meets its goals and requirements (Solomonides et al., 2022).

In this master thesis, a supervised machine learning approach is employed to automatically classify the stakeholder requirements into FR & NFR. Supervised learning involves training a model on a labelled dataset, where each requirement statement is paired with FR or NFR. A suitable dataset such as PURE dataset, which contains real world labelled software requirements, is used to build and evaluate the model (Siddeshwar et al., 2024). The flow begins with preprocessing steps including tokenization, cleaning and feature extraction using techniques such as TF-IDF or word embeddings (Kaur et al., 2024). For the classification of algorithms like Support Vector Machine (SVM) and fine-tuned BERT are considered. BERT is particularly effective in capturing contextual semantics of requirement sentences and outperforms traditional models in text classification tasks (Taj et al., 2025). The experiments with models like SVM and Logistic Regression (LR) as baselines and compares them against a fine-tuned BERT based classifier, which has been shown to be state of the art in many NLP tasks. The classifier is trained and evaluated using metrics such as accuracy, precision, recall and F1 -score ensuring its effectiveness in correctly identifying FR and NFRs (Liu et al., 2022; Wang et al., 2024).

The central AI technique used to power the interactive prototype is Retrieval Augmented Generation (RAG) an advanced architecture that combine the neural information with Large Language Model (LLM). Unlike traditional question-answering systems that rely on static, pre-trained knowledge. RAG dynamically pulls the relevant documents from the knowledge base at runtime to generate contextually grounded & up to date answers (Ferrari et al., 2023). This makes RAG ideal for application in RES where system knowledge must evolve based on user feedback and newly discovered requirements (Maalej et al., 2023; Zowghi et al., 2023).

## Overview of Requirements Engineering Practices for AI in Literature

This section provides a complete analysis of existing literature on Requirements Engineering Specification (RES) Practices for Artificial Intelligence (AI). The challenge of adapting RES methodology to address AI's unique characteristics is explored, dealing with data-based requirements and facilitating iterative development. The existing literature solution is reviewed, highlighting its effectiveness and its shortcomings in addressing these challenges.

The analysis found gaps in current practices, particularly in the ability to address specific needs of AI, such as ensuring equity, promoting transparency, and managing dynamic requirements. Key concepts of RES in the context of AI, such as the participation of stakeholders and non-deterministic behaviour of AI systems, are introduced and defined to establish the basis for further discussion.

The relevant references are provided to establish the framework for proposed frameworks and demonstrate the evolution of RES practice in response to the growing complexity of AI (Maalej et al., 2023; Zowghi et al., 2023). These references also underline the pressing need for innovative and adaptive RES approaches that align with the technical, ethical, and social demands of modern AI systems.

In (Maalej et al., 2023), the authors address the challenges of tailoring Requirements Engineering Specification (RES) for responsible AI systems. They propose adapting traditional RE methods to meet unique AI demands, focusing on ethical, societal, and stakeholder requirements. The paper highlights six key areas: quality requirements, data-centric RE, prototyping, responsible AI terminology, trade-off analysis, and continuous testing. For example, they emphasize the need to specify measurable quality attributes like fairness and explainability early in the RE process, linking these to data and system behaviour. The study discusses methods such as interviews, observational studies, and systematic trade-off analyses to ensure stakeholder needs are met and trust in AI systems is maintained.

The main advantage is the systematic integration of societal and ethical considerations into RE, providing a robust foundation for AI system development (Maalej et al., 2023). However, its applicability is limited by the complexity of operationalizing high-level concepts like fairness and transparency. Verification relied on literature reviews and case studies; no direct experiments were performed. The techniques include quality failure analysis and using structured prototypes that align data and user-centric approaches. Tools and datasets were not explicitly mentioned.

This paper contributes to this master thesis by identifying gaps in current RE practices for AI, particularly in managing ethical and data-related challenges, helping me focus on developing frameworks to align AI systems with human values. Summing up, based on the analysis, current RE techniques lack structured approaches to address ethical and societal needs in AI systems, such as fairness and explainability. Similarities exist in emphasizing stakeholder involvement, but differences lie in AI's demand for iterative, data-driven requirements specifications compared to traditional software. Effective trade-off analysis and prototyping are critical for balancing AI's technical capabilities with user expectations and societal norms.

In (Kaur & Kaur, 2024), the authors present a systematic literature review (SLR) investigating the use of AI techniques—specifically machine learning (ML), deep learning (DL), and transfer learning (TL) in software requirements classification. They identify 60 studies, emphasizing that transfer learning approaches consistently outperform ML and DL in terms of accuracy. The paper highlights common techniques like Support Vector Machines (SVM) and Convolutional Neural Networks (CNN), and evaluation metrics such as Precision and Recall. Key experiments involve text preprocessing, word embedding techniques (e.g., Word2Vec, GloVe, and BERT), and the application of ML/DL models to classify functional and non-functional requirements from datasets like app reviews and software specifications. Tools like Fast Text and pre-trained language models are also utilized. A notable limitation is the lack of exploration into real-world applicability, where complex settings may present challenges not addressed by the reviewed techniques. This research contributes to this Master Thesis by providing insights into cutting-edge AI-driven methods for automating requirement classification, informing the design of robust, scalable solutions. Summing up, based on the analysis performed, transfer learning approaches exhibit higher classification accuracy compared to traditional ML/DL but require significant computational resources. Similarities across studies lie in their focus on improving accuracy with modern embedding techniques, while dissimilarities are evident in dataset types and preprocessing strategies. Limitations include inadequate evaluation in real-world contexts and the need for more robust, domain-specific datasets.

In (Solomonides et al., 2022), the AMIA Position Paper presents a framework that combines medical ethics with technical principles and guides the ethical management of the life cycle of medical AI systems. It highlights the benefits of using safe and equitable AI and acknowledges the challenges of the complexity of implementation. Techniques such as machine learning and tools such as adaptive decision support systems are a key focus, with verification and monitoring through case studies. The emphasis on the treatment of biases is consistent with the relevance of the document to the establishment of ethical and operational standards for AI systems in health care and contributes to the user's research on request engineering. In summary, based on analysis of ethical governance, principles, and bias mitigation, it is found that, despite increasing the trustworthiness of artificial intelligence and aligning it with the ethics of health care, the challenges in implementation, adaptation, and standardisation persist.

In (Nalchigar et al., 2021), the GR4ML framework is proposed for modelling machine learning (ML) requirements using a goal-oriented approach across three perspectives: business, analytics design, and data preparation. The main advantage of this approach is its ability to bridge communication gaps between business stakeholders and technical teams by visually modelling goals, decisions, and ML requirements. However, its disadvantages include being time-consuming, requiring skilled modelers, and lacking built-in prioritization methods. The framework utilizes goal decomposition, soft goal modelling, decision trees, and algorithm-performance matching as key techniques.

A case study was used to validate the framework in a real-world healthcare startup. The team applied the model to an existing ML-based analytics product that supports clinical decisions for primary care. Modelling sessions were conducted using diagramming tools, prompting questions, and structured interviews. Tools included goal modelling techniques, entity-relationship analysis, and catalogues of ML algorithms and data transformation patterns. Though no open datasets were used, actual company data (e.g., patient records, intervention histories) were applied in modelling.

This paper contributes to my master’s thesis by demonstrating a comprehensive requirement engineering methodology for ML healthcare systems, which aligns with my thesis focus on specifying AI-powered diagnosis requirements. Based on the analysis of (Nalchigar et al., 2021)it is found that goal-oriented modelling helps elicit comprehensive ML requirements but is resource-intensive and requires expert facilitation. All analysed approaches aim to align ML development with business goals; however, GR4ML stands out by offering a structured modelling methodology across multiple views. A common limitation among such frameworks is the lack of automated guidance for prioritizing goals and estimating business impact, which could enhance decision-making.

In (Ferrari, 2023), the author presents a keynote discussing the evolving role of Requirements Engineers (REs) in the age of Artificial Intelligence (AI). The main contribution is a vision of how large language models (LLMs) like ChatGPT are transforming requirements engineering by enabling prompt-based specification and automatic model generation. The keynote outlines several use cases, including explainable ML for agronomists, NLP for railway systems, and AI for river and road maintenance. The advantages of the proposed approach include increased speed, automation, and accessibility in RE tasks. A key disadvantage is the societal risk posed by AI’s rapid adoption, such as misinformation, inequality, and job displacement. While no formal experiments are reported, the insights are grounded in practical projects and observations. Tools discussed include LLMs and NLP-based systems, though no specific datasets or algorithms are detailed. This paper contributes to our master thesis by highlighting how emerging AI technologies impact RE practices and the need for responsible, human-centric AI system design.

Based on the analysis performed on, it is found that AI democratization enhances RE efficiency but brings ethical and social risks. While this approach leverages LLMs for automation, it differs from model-driven frameworks that rely on formal modelling and structured requirements. A shared limitation among such visionary approaches is the lack of empirical validation and concrete guidance for implementation in regulated domains like healthcare.

In (Ramachandran et al, 2024), an AI and blockchain framework for healthcare applications—called AI-BlockchainOps—is proposed to enhance security, privacy, and sustainability in medical systems. The paper introduces a systematic development process integrating AI models, blockchain architecture, smart contracts, and BPMN-based simulation for Electronic Health Records (EHRs). A key advantage is improved transparency and interoperability of patient data, while challenges include integration complexity, scalability, and regulatory compliance. The approach uses deep learning for AI model development, smart contracts for data control, and Business Process Model and Notation (BPMN) tools (e.g., Bizagi) for simulation. A simulation experiment involving 100 user requests demonstrated resource utilization and processing efficiency of blockchain-backed EHR workflows. The reference architecture emphasizes standardization, sustainability, and compliance by integrating AI, blockchain, IoT, and cybersecurity. Although no open datasets were explicitly used, the simulation operated on modelled patient workflows. This paper contributes to my master thesis by illustrating a structured requirements engineering and modelling process for secure, explainable, and scalable AI-powered diagnostic systems in healthcare.

Based on the analysis performed on, it is found that AI-BlockchainOps enhances system security and efficiency but presents integration and scalability challenges. Compared to traditional ML-based healthcare models, this framework offers a more comprehensive, modular, and standards-driven development process. A shared limitation among integrated AI-healthcare systems is the complexity in validation, regulatory alignment, and ensuring explainability across technologies.

In (Sporsem et.al, 2024), the research proposes a qualitative empirical approach to discover explainability requirements in ML-based software. The study emphasizes the integration of qualitative methods (e.g., interviews, participatory observation) and Continuous Software Engineering (CSE) principles to iteratively gather user feedback, aiming to bridge gaps in Requirements Engineering (RE) for ML systems. Key techniques include Yin’s case study methodology, Phenomenon-focused Problematization Framework, and qualitative sensemaking to analyse how developers elicit domain-specific explainability needs. Verification is conducted through a multi-phase case study strategy involving 40–60 interviews and 3–4 months of observational studies across Norwegian enterprises developing ML software. No specific tools or datasets are mentioned, though qualitative analysis tools (e.g., NVivo) are implied. This research contributes to a Master Thesis by highlighting the necessity of qualitative feedback in understanding user trust and domain-specific explanation requirements, offering practical insights for integrating explainability into ML development cycles.

Based on the analysis performed on the case study design and methodology, it is found that the research’s focus on Norwegian enterprises may limit generalizability to other regions or industries. While the iterative feedback approach aligns with Agile and Continuous Software Engineering practices, its reliance on resource-intensive qualitative methods (e.g., interviews, observations) poses scalability challenges. Additionally, prioritizing qualitative sensemaking over quantitative techniques highlights a divergence from traditional data science practices, emphasizing contextual depth at the expense of broader generalizability.

## Summary of research papers

In this section, we review the most recent research on AI-based approaches to improve the requirement engineering in Table 2.1. It consists of the following rows:

1. Reference: A study citation including the author and publication year.
2. Used approach: Methods of research used to solve problems, such as experiments, literature reviews, surveys, or conceptual analyses.
3. Fields studied/areas of application: Specific areas or areas of application focusing on, such as AI in healthcare.
4. Data sets used: Information about any dataset used in the study, including the data source and characteristics.
5. Attributes used for prediction: Key attributes and features analysed for prediction and classification, such as fairness, transparency, stakeholder needs, or system goals.
6. Evaluation of methods: Summary of methods and models evaluated, including performance metrics and case studies.
7. Results: Results of the study, including important conclusions, observations, or contributions.

The summary of the articles (Table 2.1) on AI-based requirements engineering provides an overview of recent advancements in leveraging artificial intelligence to address challenges such as dynamic, iterative refinement, and stakeholder validation, with a particular focus on the specification phase of requirements engineering.

Table 2.1 Summary of Articles

| **Reference** | **Used approach** | **Problem domain studied** | **Data set used** | **Attributes used for prediction** | **Evaluation of the approach** | **Main Result** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| (Solomonides et al., 2022) | Definition and rationale for ethical principles governing AI systems, based on established medical practices and engineering standards. | Ethical governance of AI systems in healthcare and related services. | Not specified; the focus is on principles and governance rather than empirical data. | Ethical principles such as beneficence, nonmaleficence, autonomy, justice, explainability, fairness, and accountability. | Emphasis on the need for transparency, accountability, and user education, particularly regarding vulnerable populations. | Framework of ethical principles for AI governance throughout its lifecycle, highlighting the importance of ongoing research and user education in biomedical applications. |
| (Sofian et al., 2022) | Systematic mapping study, examining trends in AI techniques across different SE phases. | Software Engineering. | Not specified in the provided text. | Not applicable; no prediction task is involved. | Analysis and mapping of literature on AI techniques in SE phases. | Identified gaps in the application of AI techniques across SE phases, suggesting areas for further research. |
| (Siddeshwar et al., 2024) | Systematic literature review (SLR) of 122 papers on AI in requirements elicitation. | AI applications in requirements elicitation tasks (15 identified tasks, e.g., requirement mining, classification, glossary extraction). | 12 public datasets (e.g., PROMISE, PURE, DePaul corpus) + app reviews, social media, legal documents. | Textual features (pre-processed via tokenization, stemming, etc.) for binary/multi-label classification (e.g., FR/NFR, requirement relevance). | Standard metrics: Precision, Recall, F1-score, Accuracy. BERT dominated performance (avg F1=0.91). | BERT outperforms traditional ML (SVM/NB) in most tasks. Key limitations: Low generalizability, data dependency, and suboptimal results despite sufficient data. Future: LLMs for automation, synthetic data generation, and explainable AI. |
| (Liu et al., 2022) | Machine Learning (ML), Natural Language Processing (NLP), Neural Networks, Classifiers (SVM, Naïve Bayes, RF, etc.) | Software Requirements Engineering (RE): elicitation, specification, validation | PROMISE dataset, App store reviews, Twitter data, user forums, software specs | Textual features (BoW, TF-IDF, POS tags, linguistic dependencies, sentiment, etc.) | Accuracy, F-score, precision, recall, comparative analysis across classifiers | AI models can achieve up to 98% recall (CNN), 91% accuracy (MNB), or 83% F-score (SVM-TFIDF-n-gram); NLP preprocessing significantly improves performance |
| (Maalej et al., 2023) | Analysis of specific RE methods for addressing AI-related challenges; tailored RE methods focused on responsible AI practices. | AI systems, specifically responsible for AI in domains requiring high standards for quality, fairness, and accountability. | The article does not specify a dataset but emphasizes the importance of data-centric approaches and prototyping in RE for AI projects. | Discusses attributes related to model quality such as accuracy, explainability, fairness, and adaptability. | Qualitative analysis based on case studies, literature review, and examples of AI system failures due to poor RE practices. | Suggests that tailored RE practices can help avoid AI project failures by addressing issues like transparency, data governance, and alignment with user expectations, which enhance AI system acceptance and effectiveness. |
| (Ahmad et al., 2023) | Development of a framework based on human-centred AI guidelines and user surveys; uses a catalogue for requirements elicitation and a conceptual model for visual presentation. | AI-based software, specifically in the context of enhancing 360° videos for virtual reality (VR) users | Not explicitly specified; uses insights gathered from user surveys to guide requirement collection. | Focuses on human-centred attributes, such as inclusivity, fairness, and user experience quality for VR. | Applied to a case study to test the framework’s effectiveness in capturing early-stage requirements. | The framework facilitated a comprehensive understanding of human-centred needs, ensuring critical requirements are identified early in the RE process for AI-based solutions. |
| (Ahmad et al., 2023) | Systematic mapping study, reviewing 43 primary studies to analyse methodologies, models, and tools used in RE4AI. | RE for AI systems, with a particular focus on real-world applications like autonomous vehicles. | No specific dataset: the study uses literature (43 primary studies) as its data source. | Not applicable, as this is a review study rather than a predictive analysis. | Analysis of literature to identify challenges, limitations, and gaps in RE4AI. | Identified gaps in current RE practices for AI, with a need for new RE techniques addressing ethics, trust, and explainability in AI systems. |
| (Kaur & Kaur, 2024) | Systematic literature review (SLR) to collect and analyse primary studies on AI techniques in requirements classification. | Requirements engineering, specifically in requirements identification and classification for software systems. | A collection of 60 studies from the literature on AI techniques in requirements classification. | Precision and recall metrics, as well as AI techniques like Support Vector Machine (SVM) and Convolutional Neural Network (CNN). | Performance of AI-based techniques is evaluated primarily through precision and recall metrics. | Transfer learning approaches show high accuracy in requirements classification, outperforming other ML and DL methods, though evidence in complex real-world settings remains limited. |
| (Zowghi et al., 2023) | Conceptual analysis highlighting key challenges posed by AI in SE, particularly focusing on model design, data curation, system autonomy, and ethical considerations. | Software engineering methodologies tailored to AI systems. | Not applicable, as the study is conceptual and discusses theoretical SE challenges for AI systems. | None specified, but mentions aspects like continuous learning, handling evolving datasets, and managing system autonomy. | Conceptual examination of SE challenges and risks associated with AI’s autonomous features and the need for ethical guidelines. | Identifies that AI systems require new SE approaches to address unique challenges such as continuous learning, ethical guidelines, and risk management, emphasizing the limitations of traditional SE practices for AI. |
| (Taj et al., 2025) | Fine-tuned BERT for ABSA with integration of Explainable AI (LIME). | Requirement elicitation from user reviews (software engineering). | AWARE (ABSA Warehouse of Apps Reviews) dataset: 11,321 annotated app reviews from social, productivity, and gaming domains. | Aspect categories and sentiment polarity from user review text. | Compared with SVM, CNN, and baseline BERT, metrics: accuracy, F1-score; 10-fold stratified cross-validation; LIME used for explanation. | Achieved F1-score of 0.83 for Aspect Category Detection (ACD) and 0.94 for Aspect Category Polarity (ACP); outperformed all baselines; improved interpretability using LIME |
| (Ferrari, 2023) | Exploratory keynote on LLM-driven Requirements Engineering | AI in engineering and society, especially RE transformation | NA (conceptual paper) | NA (focus on prompt-based RE, LLM impact) | Observations and examples from real-world AI applications | LLMs shift RE from static specification to prompt-based modelling; ethical and social risks highlighted |
| (Habiba et al., 2024) | Systematic Mapping Study (SMS) based on 126 primary studies from 2010–2023. | Requirements Engineering for AI-based Systems (RE4AI). | Not a single dataset; instead, 126 research papers were systematically selected using a rigorous protocol including database queries and snowballing. | Not applicable (qualitative thematic analysis, not predictive modelling). | Used thematic and SWEBOK-based classification; evaluated study maturity using Wieringa's taxonomy and empirical validation presence. | Identified that most research focuses on requirements analysis and elicitation; there are many new RE practices for AI systems, but tools and foundational processes are lacking; proposed 7 research directions for future work. |
| (Wang et al., 2024) | Federated Learning with Feature Fusion (semantic + statistical features) using ERNIE + TF-IDF in a ring architecture with knowledge distillation | Automated requirement classification in requirement engineering for complex product systems | Custom-built dataset of 2500 spacecraft-related requirements (500 each of functional, performance, interface, safety, and reliability) extracted from PDFs | Semantic features (ERNIE-based embeddings), Statistical features (TF-IDF scores) | Accuracy, Precision, Recall, F1-score compared across centralized learning, FedAvg, and the proposed method | Proposed model (ERINE+TFIDF with federated learning) achieved 88.63% accuracy and 88.71% F1-score, slightly lower than centralized learning (92.54%) but with strong data privacy. |
| (Maleki et al., 2022) | Intelligent Digital Twin (iDT) architecture, Neural networks, Computation Tree Logic (CTL), Python libraries (TensorFlow, NLTR) - Feedback-driven ML | Automated Requirements Engineering (RE) and Model-Based Systems Engineering (MBSE) for medical devices, addressing critical factors like human factors, security, safety, and reliability. | 350 critical words (curated library), 200 training requirements, 275 test requirements (ventilator case study) | Criticality levels of words, Component names/attributes, CTL-derived relationships, Feedback labels (corrected CTL models) | 100% critical requirement categorization,92% component identification accuracy,90% neural network accuracy (post-training), 31 CTL errors corrected via feedback | The software service automates critical requirements modelling and system architecture generation for medical devices, achieving high accuracy. Feedback integration improves adaptability, demonstrating iDT’s potential in healthcare RE/MBSE. |
| (Sporsem et al, 2024) | Qualitative empirical case studies, Yin’s and Runeson’s case study protocols, Phenomenon-focused Problematization Framework, Interviews (40–60) and participatory observation (3–4 months) | Eliciting explainability requirements in ML-based software across healthcare, automotive, energy, and banking sectors. | Qualitative data from interviews and observations (Norwegian enterprises developing ML software) | Not applicable (qualitative study) | Effectiveness assessed via iterative feedback loops, Comparative analysis of case studies, Validation through theoretical propositions | The study proposes a methodology for continuously eliciting explainability requirements via qualitative feedback (interviews, observations) and iterative development, bridging gaps between developers and domain experts to build trust in ML systems. |

After gathering the scientific papers, the information can be summarized as follows:

**Reference (1):** This column highlights important studies in this field, including authors and year of publication. It provides a comprehensive view of basic and recent research on the use of artificial intelligence in requirements specifications. These references guide discussions by showing various applications and advances.

**Used Approach (2):** The methodology described in this column includes various AI-based technologies, such as natural language processing (NLP) to extract requirements, machine learning (ML) (eg, BERT, SVM) to classify and predict requirements, and concept frameworks to structure requirements. These approaches show the flexibility and innovation of artificial intelligence adaptation to improve the process of requirements engineering specification.

**Problem Domain Studied (3):** This column shows specific application areas for which artificial intelligence has been applied in requirements specifications, such as healthcare, autonomous systems, software development, and financial services.

**Dataset Used (4):** The studies used a wide range of datasets (eg, PROMISE, AWARE), app reviews, social media data, technical documents, and domain-specific case studies.

**Attributes Used for Specification (5):** The column focuses on functional and non-functional requirements, stakeholders' inputs, ethical considerations, and system objectives considered in the requirements specification process. These attributes play a key role in ensuring accurate and comprehensive specifications.

**Evaluation of Methods (6):** The effectiveness of the proposed approaches is measured through both quantitative and qualitative methods. Metrics such as accuracy, precision, recall, and F1-score are commonly used, while others assess stakeholder satisfaction, alignment with ethical standards, or compliance with domain-specific guidelines. These evaluations validate the practical utility of AI-enhanced requirements engineering methods.

**Main Result (7):** Overall, the studies report improvements in the accuracy, traceability, explainability, and ethical soundness of the requirements engineering specification process.

## Summary of the 2nd section

This research marks an important step forward in adapting RES practices to the unique challenges of AI systems. The research addresses key gaps in traditional RES methods by proposing a new framework that is adapted to AI-based complexities such as data-driven requirements and iterative development processes.

1. Literature analysis revealed that there were no robust RES approaches designed to adapt to the dynamic and non-deterministic nature of AI-based development. Ethical considerations such as equity, transparency, and accountability, as well as technical challenges such as changes in requirements and diversity of stakeholders, remain underexplored in existing methods. These conclusions laid the foundations for the development of the proposed approach.
2. The development and implementation of an artificial intelligence-based specification approach validate its ability to improve accuracy, efficiency, and consistency in the specification documentation process.
3. The proposed RES approach has been tested through practical implementation and experiments and has shown its ability to improve transparency, accountability, and adaptability in AI-based projects. It successfully managed the evolving requirements and aligned them with the expectations of the stakeholders.

The key findings and contributions of this study include the transparency and responsibility of the customized RES approach for AI-based development. The approach addresses key issues such as data-based requirements and iterative development processes to ensure that AI-based development is consistent with social values and technical requirements. It offers, Improved transparency, accountability, and responsibility in AI-based development.

## AI Algorithm Comparison Table for NLP-based RES

To effectively support requirement classification, ambiguity detection, and semantic understanding in NLP-based Requirements Engineering Specification, various AI algorithms can be leveraged based on their unique strengths, limitations, and suitability as summarized in Table 2.2 below.

Table 2.2 Review of AI Algorithm

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Suitability for RES Tasks** | **Dataset Used** | **Evaluation Metrics** | **Tools** | **Key Features** | **Reference** |
| **1** | **2** | **3** | **4** | **6** | **7** | **8** |
| BERT (Transformer-based) | Excellent for requirement classification, extraction | PROMISE, PURE, DePaul, app reviews | Precision, Recall, F1-score, Accuracy | Hugging Face Transformers, PyTorch | Bidirectional encoding, pre-trained on a large corpus, fine-tuneable for downstream tasks | (Siddeshwar et al., 2024) |
| Fine-Tuned BERT + LIME | Suitable for explainable sentiment-based requirement elicitation | AWARE (11,321 annotated app reviews) | Accuracy, F1-score | BERT, LIME, Python (scikit-learn, TensorFlow | Deep contextual learning + model interpretability via LIME | (Taj et al., 2025) |
| SVM, NB, CNN | Good for basic requirement classification, useful for requirement relevance classification, and suitable for binary classification of textual requirements | PROMISE, Twitter, app reviews | Accuracy, Precision, Recall, F1 | Scikit-learn, TensorFlow, Keras, Scikit-learn | Margin-based classification works with sparse, high-dimensional data. Probabilistic text classification assuming word independence. Margin-based classification works with sparse, high-dimensional data | (Liu et al., 2022) |
| ERNIE + TF-IDF (Federated Learning) | Ideal for requirement classification in sensitive domains | Custom Spacecraft dataset (2500 samples) | Accuracy, F1-score | PaddlePaddle, FedAvg, Custom FL system | Combines deep semantic embeddings (ERNIE) with lexical frequency scores (TF-IDF) under the FL setup | (Wang et al., 2024) |
| Transfer Learning | Best suited for requirements classification in data settings | 60 primary studies (SLR-based analysis) | Precision, Recall | Hugging Face, TensorFlow, PyTorch | Leverages pretrained models on large datasets, fine-tuned on task-specific data | (Kaur et al, 2024) |

## Open Dataset Assessment & Review

A well-structured and effective process for determining the system requirements engineering specification (RES) is essential. RES processes are based on various structured inputs, such as, needs and expectations of stakeholders, commercial requirements, technical constraints, regulatory requirements, etc. These structured inputs guide the accurate formulation of system requirements to ensure that the system meets the needs of stakeholders and can be able to develop.

These inputs typically include stakeholder goals, user expectations, functional & non-functional requirements, and domain-specific constraints (Maalej et al., 2023). The quality of those inputs are directly influences the completeness, consistency, and correctness of software requirements documentation. Moreover, AI-based approaches for requirement extraction increasingly leverage unstructured data sources such as QA, interviews, and documents as supplementary inputs to automate the parts of the RES process (Ferrari et al., 2023; Siddeshwar et al., 2024).

The table 2.3 below reviews the various Kaggle datasets used as input for RES:

Table 2.3 Review of the Open Datasets of RES

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Use Case** | **Features** | **Size** | **Pros** | **Cons** | **Common Tools/ Methods used** | **Data Source** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| PROMISE | Software Requirements Classification | Functional & non-functional labels, textual requirements | ~15K requirements | Well-labelled, popular in RE tasks | Older dataset, limited real-world context | SVM, RF, NB, TF-IDF, BoW | https://www.kaggle.com/datasets/iamsouvik/promise |
| MedQuAD | QA in the medical domain, information retrieval | Doctor-verified QA pairs, medical terms | 47K QA pairs across 12 medical domains | High-quality, multi-domain, curated by NIH | Focused on healthcare, not RE-specific | BERT, GPT, NER, Semantic Similarity | https://www.kaggle.com/datasets/pythonafroz/medquad-medical-question-answer-for-ai-research |
| HealthCareMagic QA | Health-related question answering and dialogue modelling | Patient input + doctor response | ~100K QA pairs | Real-world conversational format, rich context | Informal, noisy language, anonymization challenges | LSTM, BERT, GPT, Token Classification | https://www.kaggle.com/datasets/kunyuzhou/healthcaremagic-100k-json |
| Banking Chatbot | Training an intent-based chatbot for banking customer service. | Dialogue data is structured as user questions paired with bot responses. | 146 QA pairs | Ideal for prototyping a banking-focused chatbot quickly. | Synthetic nature may limit diversity; real-world phrasing could be missing. | spaCy, HuggingFace Transformers, Rasa.  Tokenization, lemmatization, intent/entity labelling. | https://www.kaggle.com/datasets/manojajj/banking-chatbot |

The datasets comparison of PROMISE, MedQuAD, and HealthCareMagic QA shows the different strengths and limitations for use in AI-based RES. PROMISE is an classic dataset containing software requirements, each labelled as FR & non-FR. Making it deal for training and evaluating classification models like SVM or NB. MedQuAD is an high quality QA pairs curated from trusted medical websites making it suitable for medical information retrieval and requirement elicitation in healthcare context.

## Summary of AI algorithms and methods

The comprehensive breakdown of the AI algorithms and methods in the proposed approach, along with the detailed reasoning for their selection.

As an core architecture, Retrieval-augmented generation (RAG) is using. RAG is an innovative approach in the field of NLP that combines the strengths of retieval-based and gneration-based models to enchance the quality of generated text. The reason for using RAG, standard LLMs can be outdated and might be hallucinate the incorrect information. RAG can make the LLM‘s response in a dynamic, organization specific knowledge base and ensuring the accuracy and contextual relevance.

For the requirement classification, we using sentence-BERT. BERT (Bidirectional Encoder Representations from Transformers) leverages a transformer based neural network to understand and gnerate the human-like language. Reason for use of BERT , the literature review (Taj et al., 2025), (Siddeshwar et al., 2024) clearly mention that BERT – outperforms traditional ML model in text classification tasks due to deep contextual understanding of language.

The use of TF-IDF, a statistical measure to evaluate the advantages of a word in a document relative to a collection. It will identify the key terms. NER is an NLP technique is to identify and classify the key entities in text into predefined categories. The pattern recognition using the rules like regular expression to identify the specific structures. Those techniques are usually used to process systematically the unstructured text and extract the structured information to build a valued knowledge base for the RAG system.

The LLM are advanced AI systems based on the deep neural networks, designed to process and generate text that resembles human language. They form the backbone of modern NLP applications, powering tools like ChatGPT, Google Gemini. LLM have a primary role is to transform retrieved data fragments into well-written, formal and unambiguous requirement statements that adhere to organizational templates. It can be consolidating information from multiple retrieved documents into a single. It can rephrase requirements from various sources into standardized and consistent format.

Essentially, the suggested method combines cutting-edge deep learning (BERT, LLMs), traditional natural language processing (NLP) methods (TF-IDF, NER), and a new feedback-driven architecture (RAG + continuous learning) to build a reliable, flexible, and automated system that is suited to the difficulties of contemporary requirements engineering.

# Proposed Approach

The proposed chatbot system is designed to automate and improve requirements engineering specification processes, especially in complex, data-driven fields such as healthcare and finance. The system begins by submitting requests to chatbot, allowing the users to interact in real time. These questions are received and processed by an AI model using the Retrieval-Augmented Generation (RAG) method to search for relevant information from structured knowledge bases. Then the content extracted is used to generate a coherent and contextually accurate response that is returned to the user. After receiving the output, the user evaluates its usefulness. If the response is satisfactory, the interaction ends, and if not, the system records a detailed feedback and timetable for analysis. This feedback is collected and analysed by the system trainer module and identifies gaps and gaps in knowledge. Based on this analysis, new requirements are generated and classified into FR & NFR. These refined insight is used to update and enrich knowledge base, and chatbots are trained or finely tuned accordingly. As a result, each feedback loop leads to an improved version of thr chatbot, making it more accurate and context-sensitive over time. This continuous learning and updating cycle ensures that chatbots adapt to the needs of dynamic users and capture valuable requirements in an automated and efficient way.

A diagram of a flowchart

AI-generated content may be incorrect.

Figure 3.1 General Schema of Proposed approach

## Describing the Process using BPMN Diagrams

In this section useses the BPMN model to represent the methodological workflows begining with the identification of data sources and culminating with the validation of detailed requirements. The model represents the strategic integration of decision points, the potentional for parallel processing mecahnism throughout the process flow.

### Main Process of Workflow

The Figure 3.2 illustrated the systematic workflow of automated extraction of FR & NFR from the requirements sets. This process is applying the integration of RAG and LLM methods to improve the consistency. Once we get the input of documents, methodology is accepting the gaps. Possisble inputs include the SRS documents, interview notes, compliance documents from different regulations.

Text Normalization & cleaning, setup of knowledge base, LLM- Based Extraction steps are discussed detail in the following section. Once those steps are finished it will structured into documentation formats and will stored in the repositories. After the approval of requiremnets sets, it will goes through the metrics based assessements. The methods will give the actual metrics that can assess using RAG. The RAG based metrics like Data & results metrics, which give Dataset coverage or Quality. Next Quality / Errors identify the number of defects, complaints. If the threshold is not reached, the evaluation results are recorded and returned to the quick refinement loop.

A diagram of a software flowchart

AI-generated content may be incorrect.

Figure 3.2 Main Process of Workflow

### Subprocess of Text Normalization & Cleaning

After providing the document as an input, the workflow processes to document normalization task that multiple steps. These steps are including the documents distributing, data cleaning, content refinement, classification of file. Performing these process operations ensuring the next stages of the workflow carried out smoothly and the making the document structure consistent.

In the file classification, it will classify the requiremnts to FR & NFR. After the classification we need to check the quality. The documents doesn‘t meet the criteria, then it will goes to manual check. In this manual check, it can be identify the details where it missed etc, or again do the automation process workflow. If the process running sucessfully, next step is to construct the knowledge base. Any documents that are missing any valuable information or details it will sent to check manually . After the quality checking , then only we can proceed to construct the Knowledge base.

A diagram of a flowchart

AI-generated content may be incorrect.

Figure 3.3 Subprocess of Text Normalisation & Cleaning

### Subprocess of Knowledge Base setup

The knowledge base setup is an important role in this approach. If we need a perfect knowledge base we should validate the knowledge. The backbone of Knowledge base setup is RAG and LLM. Generally knowledge base is a centralised repoistory of information. An essential part of knowledge management systems are knowledge based. They are employed to maximise the gathering, organising and retrieval of information. A knowledge base can give customers access to information that would otherwise require contact with an organization‘s staff. This capability simplifies interaction for both the customer and the organization.

The workflow starts with preprocessed document and parallely doing the input systematically.

The TF-IDF (Term frequency-Inverse Document Frequency) method is used to extract the vocabulary, terms. TF-IDF is a statistical measure used in natural language processing and information retrieval to evaluate the importance of a word in a document relative to a collection of documents. Unlike simple word frequency, TF-IDF balances common and rare words to highlight the most meaningful terms. Concurrently, entity and pattern recognition techniques are used to extract the entities and relationship. Entity recognition is the process of identifying and classifying named entities in a given text. This process involves the identification of the entity type, such as person, organization or location, and is used in various natural language processing tasks such as text classification, information retrieval, and sentiment analysis. Pattern recognition in NER typically involves using regular expressions to identify specific patterns of text that correspond to named entities.

The compliance rules incorporates relevent organisational policies and legal documents that govern the RES process. The continuous knowledge integration enchanes the knowledge base to update the past requirements specification and make it to strong. Once it‘s updated the knowledge base it will check the quality of knowledge base because we need to confirm its valid or not . If the knowledge base is not valid, it will checks the automated flow, retry extraction, LLM fallback and request to cancel the invalid information and log the error.

A diagram of a flowchart

AI-generated content may be incorrect.

Figure 3.4 Subprocess of Knowledge Base Setup

### Subprocess of LLM – based Extraction

The LLM processing is the core functionality of the process. LLM process will provide structured requirement sets. Most documents, especially in business environments, contain unstructured data. Traditional rule-based extraction methods struggle with this complexity. LLMs, excel in recognizing patterns and context in unstructured text, making them ideal for extracting relevant information.

The flow begins with proceeding the prompt data. In the requirement augmentation task, an requirement sets is identified and extracted free text. In the format normalization task check the organisational template compliance with the document‘s struture and formatting. After the formatting, it will check for the consistency and save the learned data for the future use. Furthermore, we will check the output satisfactory. If the ouput is not satisfies, it will re-prompt, cross verify and manually review and save the error logs. If the output is satisfies, it will goes for validation.

Quality validation will check based on the design and implementation of the system assessement process aimed to ensure all generated requirements are appropriated and expected one. Requirement pass validation undergo output formatting tasks and if there need any improvenent it will revert to prior processing stages based on the specific quality concerns flagged.

A diagram of a company

AI-generated content may be incorrect.

Figure 3.5 Subprocess of LLM - Based Extraction

# Result Measure Metrics

A multifaceted assessment technique is needed to experimentally evaluate the usefulness and efficiency of the proposed AI- enabled Requirements Engineering Specification (RES) approach. The Quantitative Performance Metrics will be used to evaluate the core AI components of the framework, primarily the classification and extraction modules, against the ground-truth data.

For the RAG-based Extraction & Generation, Retrieval Precision/Recallmeasures the effectiveness of the knowledge base retrieval step. What percentage of retrieved document chunks are truly relevant to the user's query?

In Hallucination Rate, the percentage of generated requirements that contain factually incorrect or unsupported information not present in the retrieved context. This metric is critical for ensuring reliability.

In the **standard classification metrics**, the purpose is to quantitatively evaluate the performance of the core BERT classifier for distinguishing FR and NFR, providing a baseline for the system’s accuracy. It is evaluated by the classifier will be tested on a held-out, labelled test set. Metrics used is precision, Recall, F1- score and Accuracy will be reported for FR & NFR classification.

In **fairness**, prevents the hallucinations and ensures the traceability to original source documents by assessing if the created is truly based on the context that was acquired from the knowledge base. Every produced requirement (FR/NFR) will be broken down into separate clauses. Each clause's ability to be inferred from or directly supported by the recovered document chunks utilised during the RAG process will be evaluated by a verification function (such as a heuristic rule-based checker or an improved NLI model).

In **appropriate answer**, evaluates the semantic alignment between the produced demand and the original user prompt. It guarantees that the result is not only a generic or off-topic text, but rather a direct and appropriate answer. The evaluation method will be cosine similarity between the vector representations of the user's input question, and the AI-generated requirement will be computed using embedding models (such as Sentence-BERT). High relevance is indicated by a high similarity score.

# Initial Experiment & Analysis

## Domain Selection & Justification

To evaluate the effectiveness of integrating Requirements Engineering Specification (RES) processes into dynamic, data-driven AI systems, this study selects the domain of Bank using Banking Chatbot datasets. Specifically, the case study utilizes the 146 question-answer pairs. The banking domain because it combines complexity, criticality, structured rules, and unstructured user queries, making it an ideal test bed to justify the effectiveness of AI-enabled RES process generation.

The domain is selected based on several key considerations:

**Complexity of Requirements:** The banking systems deal with transactions, security, compliance and user interactions, making them rich in diverse requirement types (FR & NFR). This ensures that the method is tested against the real-world and high-stake requirements.

**Structured Contextual Data:** The banking processes are well-defined, but users often express needs in unstructured natural language like queries, chats. This mismatch is perfect for testing retrieval + LLM generation Pipelines.

**Industry Relevance:** The financial institutions are increasingly adopting AI-enabled requirements engineering tools to improve the automation, compliance and efficiency. This research directly aligns with a high demand industry use case.

## Technical Feasibility Analysis Based on the Dataset

The features, organisation, and appropriateness of the dataset utilised to generate requirements have a significant impact on the technical viability of this study. The dataset was examined to see whether it was sufficient to support the suggested retrieval-augmented generation (RAG) pipeline. The dataset comprises user queries and the system replies that correspond to them in the banking sector. This dataset contains synthetically generated question answers for building an interactive banking chatbot.

A white background with red and blue text

AI-generated content may be incorrect.

Figure 5.1 Banking Chatbot Dataset Sample

This banking chatbot dataset sample is publicly available dataset, comprising approximately 145 question-answer (QA) pairs.  This dataset does not contain any sensitive/personal/classified information about any individual or any bank. This dataset is purely for educational and practice purpose of generative AI only. In this study, only 12 QA pairs were taken.

The suggested strategy uses a hybrid technique that blends generative requirement formulation with retrieval-based context selection. To ensure that the created demand is based on pertinent domain knowledge, Sentence-BERT and FAISS are used in conjunction to construct an embedding-based retrieval process that finds contextually related replies from the dataset. Then, FLAN-T5 is used to apply a prompt-based large language model (LLM) generation approach, which uses the user query and recovered context to provide a clear and simple demand. Lastly, the structured prefix "The system shall," which implicitly categorises all outputs as functional requirements, is enforced to direct the creation process using templates. Contextual correctness and formal consistency in the generated requirements are guaranteed by this combination of retrieval, generative modelling, and structured prompting.

## Initial Experiment Results

Retrieval-Augmented Generation (RAG), which combines prompt-driven generation and semantic retrieval, is the foundation of the suggested technique. SentenceTransformers (all-MiniLM-L6-v2) are used in the first stage to encode the dataset's existing replies into vector embeddings, which are then placed in an FAISS index to facilitate quick and effective similarity searches. In order to allow precise requirement creation, the system obtains the top k most pertinent answers whenever a new query is entered.

A close up of a text

AI-generated content may be incorrect.

Figure 5.2 Extracted Functional Requirement

In the second step, the user question and the context that was collected are combined to create a structured prompt that directs the language model to generate a demand. The prompt ensures uniformity in output by enforcing a predetermined format that starts with "The system shall." Functional requirements are then produced in a clear, simple, and testable manner using a text-to-text generation model (FLAN-T5). Crucially, the categorisation component is implicit: the structured prompting approach guarantees that all produced outputs are immediately defined as functional needs, as opposed to explicitly classifying requirements as functional or non-functional. The system is able to convert unstructured user queries into well-structured and standardised requirement statements by integrating retrieval, creation, and template assistance.

# General Conclusion

The crucial task of modifying conventional Requirements Engineering Specification (RES) procedures to satisfy the requirements of Artificial Intelligence (AI) systems was the focus of this thesis. Because of their data-dependency, non-deterministic behaviour, and iterative growth, AI systems make traditional RE techniques inadequate and frequently result in ambiguity, a lack of transparency and accountability, and misalignment with stakeholder expectations.   
 Through a thorough literature study and the creation of a unique AI-enabled framework, the main research topic, "How can requirement engineering (RE) Specification processes be improved using AI-based development?" was investigated. This work's main contribution is a chatbot system based on Retrieval-Augmented Generation (RAG) that automatically extracts, categorises, and formally specifies needs from unstructured textual inputs.

The suggested strategy effectively illustrates a way to:

Automate and Structure Elicitation: The system lessens the manual, labour-intensive aspect of traditional RES by processing a variety of inputs, such as interview notes, regulatory papers, and user questions.

Boost Accuracy and Consistency: The framework uses cutting-edge natural language processing (NLP) models such as Sentence-BERT and FLAN-T5 to categorise requirements and provide precise, unambiguous, and testable specifications that are preceded by "The system shall."

Encourage Continuous Learning: An integrated feedback loop enables the system to gain information from user interactions, enhancing its knowledge base and performance over time. This is essential for handling changing AI needs.

Guarantee Contextual Grounding: By guaranteeing that created requirements are based on a validated knowledge base, the RAG architecture lessens hallucinations and enhances the output's traceability and factual accuracy.

The initial experiment conducted on a banking chatbot dataset validated the technical feasibility of the approach. The system effectively retrieved relevant context from a knowledge base and generated coherent functional requirements from user queries, demonstrating the practical application of the proposed methodology in a high-stakes, complex domain.

In conclusion, this research provides a significant step towards bridging the gap between traditional RE practices and the dynamic needs of AI development. The proposed AI-enabled RES framework offers a scalable, adaptive, and efficient solution to improve clarity, consistency, and stakeholder alignment in the specification of AI systems, ultimately contributing to the development of more reliable, transparent, and responsible AI.

# Future Work Plan

Building on the foundation laid in this thesis, the next semester’s work will focus on enhancing and extending the proposed AI-enabled Requirements Engineering Specification (RES) Process.

Developing a web based or application based chatbot system for the users with methods. Further development will focus on enhancing the AI components to explicitly handle non-functional requirements (NFRs) and integrate with common software engineering tools for practical adoption.

# References

Ahmad, K., Abdelrazek, M., Arora, C., Agrahari Baniya, A., Bano, M., & Grundy, J. (2023). Requirements engineering framework for human-centered artificial intelligence software systems. *Applied Soft Computing*, *143*. https://doi.org/10.1016/j.asoc.2023.110455

Ahmad, K., Abdelrazek, M., Arora, C., Bano, M., & Grundy, J. (2023a). Requirements engineering for artificial intelligence systems: A systematic mapping study. *Information and Software Technology*, *158*. https://doi.org/10.1016/j.infsof.2023.107176

Ahmad, K., Abdelrazek, M., Arora, C., Bano, M., & Grundy, J. (2023b). Requirements practices and gaps when engineering human-centered Artificial Intelligence systems. *Applied Soft Computing*, *143*. https://doi.org/10.1016/j.asoc.2023.110421

Ferrari, A. (2023). Artificial Intelligence in Engineering and Society: Blue Skies, Black Holes, and the Job of Requirements Engineers (Keynote). In K. Schneider, F. Dalpiaz, & J. Horkoff (Eds.), *2023 IEEE 31ST INTERNATIONAL REQUIREMENTS ENGINEERING CONFERENCE WORKSHOPS, REW* (Issue 31st IEEE International Requirements Engineering Conference (RE), pp. 67–67). https://doi.org/10.1109/rew57809.2023.00018

Habiba, U. E., Haug, M., Bogner, J., & Wagner, S. (2024). How mature is requirements engineering for AI-based systems? A systematic mapping study on practices, challenges, and future research directions. *Requirements Engineering*, *29*(4), 567–600. https://doi.org/10.1007/s00766-024-00432-3

Kaur, K., & Kaur, P. (2024). The application of AI techniques in requirements classification: a systematic mapping. *Artificial Intelligence Review*, *57*(3). https://doi.org/10.1007/s10462-023-10667-1

Liu, K., Reddivari, S., & Reddivari, K. (2022). Artificial Intelligence in Software Requirements Engineering: State-of-the-Art. In *Proceedings - 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science, IRI 2022* (Issues 23rd IEEE International Conference on Information Reuse and Integration for Data Science (IEEE IRI), pp. 106–111). https://doi.org/10.1109/IRI54793.2022.00034

Maalej, W., Pham, Y. D., & Chazette, L. (2023). Tailoring Requirements Engineering for Responsible AI. *Computer*, *56*(4), 18–27. https://doi.org/10.1109/MC.2023.3243182

Maleki, S., Jazdi, N., & Ashtari, B. (2022). Intelligent Digital Twin in Health Sector: Realization of a Software-Service for Requirements- and Model-based-Systems-Engineering. In *IFAC-PapersOnLine* (Vol. 55, Issue 19, pp. 79–84). https://doi.org/10.1016/j.ifacol.2022.09.187

Nalchigar, S., Yu, E., & Keshavjee, K. (2021). Modeling machine learning requirements from three perspectives: a case report from the healthcare domain. *Requirements Engineering*, *26*(2), 237–254. https://doi.org/10.1007/s00766-020-00343-z

Ramachandran, M. (2024). Ai and Blockchain Framework for Healthcare Applications. *Facta Universitatis, Series: Electronics and Energetics*, *37*(1), 169–193. https://doi.org/10.2298/FUEE2401169R

Siddeshwar, V., Alwidian, S., & Makrehchi, M. (2024). A Systematic Review of AI-Enabled Frameworks in Requirements Elicitation. *IEEE Access*, *12*, 154310–154336. https://doi.org/10.1109/ACCESS.2024.3475293

Sofian, H., Yunus, N. A. M., & Ahmad, R. (2022). Systematic Mapping: Artificial Intelligence Techniques in Software Engineering. *IEEE Access*, *10*, 51021–51040. https://doi.org/10.1109/ACCESS.2022.3174115

Solomonides, A. E., Koski, E., Atabaki, S. M., Weinberg, S., Mcgreevey, J. D., Kannry, J. L., Petersen, C., & Lehmann, C. U. (2022). Defining AMIA’s artificial intelligence principles. *Journal of the American Medical Informatics Association*, *29*(4), 585–591. https://doi.org/10.1093/jamia/ocac006

Sporsem, T. (2024). Discovering Explainability Requirements in ML-Based Software. In *Proceedings - International Conference on Software Engineering* (Issue 44th International Conference on Software Engineering: Companion Proceedings, pp. 170–172). https://doi.org/10.1145/3639478.3639807

Taj, S., Daudpota, S. M., Imran, A. S., & Kastrati, Z. (2025). Aspect-based sentiment analysis for software requirements elicitation using fine-tuned Bidirectional Encoder Representations from Transformers and Explainable Artificial Intelligence. *Engineering Applications of Artificial Intelligence*, *151*. https://doi.org/10.1016/j.engappai.2025.110632

Wang, R., Liu, J., Zhang, Q., Fu, C., & hou, Y. (2024). Federated learning for feature-fusion based requirement classification. *Cluster Computing*, *27*(3), 3397–3416. https://doi.org/10.1007/s10586-023-04147-y

Zowghi, D., Bano, M., & Borg, M. (2023). What’s Missing in Requirements Engineering for Responsible AI? *IEEE Software*, *40*(6), 11–15. https://doi.org/10.1109/MS.2023.3302934