Tampere University

Asmat Ullah

AN EMPIRICAL STUDY ON THE USE OF AI FOR REQUIREMENTS ENGINEERING IN HEALTHCARE PROJECTS

Faculty of Information Technology and Communication Sciences
M. Sc. Thesis
May 2025

ABSTRACT

An empirical study on the use of AI for requirements engineering in Healthcare Projects
M. Sc. Thesis
Tampere University
Master's Degree Programme in Computing Sciences
May 2025

This thesis addresses the challenge of managing complex and evolving requirements in healthcare software projects, a domain characterized by stringent regulatory demands, sensitive patient data, and diverse stakeholder perspectives. Traditional Requirements Engineering (RE) methods, reliant on manual analysis and subjective judgment, often struggle to capture, validate, and maintain high-quality requirements under these conditions. At the same time, advances in Artificial Intelligence (AI), particularly in Natural Language Processing (NLP), Machine Learning (ML), and Large Language Models (LLMs), offer promising tools for automating and enhancing RE tasks. Yet, the practical effectiveness and domain suitability of these AI techniques in healthcare remain underexplored.
To bridge this gap, the study employs a mixed-methods approach. A systematic literature review synthesizes research on AI applications in healthcare RE, identifying key benefits such as automated requirement extraction, ambiguity detection, and classification as well as limitations, including vocabulary mismatches and ethical concerns. Building on these findings, expert interviews and practitioner surveys were conducted with clinicians, administrators, developers, and regulatory specialists. These empirical investigations validated the literature insights, highlighting real-world barriers such as model interpretability, trust, and compliance risks, while revealing opportunities for domain-specific AI adaptation.
Based on this dual evidence base, a comprehensive framework was developed. It advocates for AI models trained on healthcare-specific corpora, embedded ethical and legal filters, interactive feedback loops for stakeholder validation, and interdisciplinary collaboration among AI engineers, clinicians, and legal experts. Pilot guidelines and training strategies are proposed to support successful tool adoption. Initial evaluations suggest that this framework can reduce elicitation time, improve requirement precision, and strengthen compliance confidence, without sacrificing human oversight.
Overall, this thesis advances both theory and practice by providing actionable strategies and a validated roadmap for integrating AI into Requirements Engineering in healthcare ultimately aiming to enhance the safety, quality, and regulatory adherence of health IT systems.

The originality of this thesis has been verified using the Turnitin Originality Check service.

USE OF AI IN THESIS

I have utilised AI tools in my thesis:

- ☐ No
- ☒ Yes

The AI tools utilised in my thesis and their purposes are described below:
Names and versions of AI tools: ChatGPT (OpenAI), DeepSeek, Mistral
Purpose of using AI tools: AI tools were used throughout the thesis process to enhance academic quality, maintain coherence across sections, and support language and structural refinement. ChatGPT was instrumental in developing research ideas, refining arguments, paraphrasing, and ensuring consistency in tone and style. DeepSeek and Mistral contributed by summarizing academic texts, suggesting improvements to structure and clarity, and enhancing the articulation of complex concepts. These tools served as writing assistants to improve clarity, cohesion, and academic rigor.
Sections where AI tools were used: AI tools were applied across almost all sections of the thesis, including:

- Title and abstract formulation
- Introduction and problem statement development
- Literature review synthesis and structuring
- Methodology clarification and description
- Data analysis and interpretation
- Discussion and conclusion writing
- Referencing style verification and proofreading

I acknowledge that I am fully responsible for the entire content of my thesis, including the parts generated by AI, and accept accountability for any violations of ethical standards in publications.

PREFACE

The integration of Artificial Intelligence (AI) into Requirements Engineering (RE) represents a transformative shift in how software systems are designed, developed, and validated, particularly in high-stakes domains such as healthcare. As healthcare systems grow increasingly complex, the need for precise, efficient, and stakeholder-aligned requirements has never been more critical. AI techniques, such as Natural Language Processing (NLP), Machine Learning (ML), and Large Language Models (LLMs), offer promising solutions to automate and enhance RE tasks, from requirements elicitation to validation. However, the unique challenges of the healthcare domain ranging from stringent regulatory compliance to the inherent ambiguity of clinical requirements pose significant barriers to the effective application of AI in this context.

This thesis emerges from a deep-seated interest in exploring the intersection of AI and healthcare RE, driven by both academic curiosity and practical experience. Having worked on healthcare projects in Finland and the United States, I witnessed first-hand the challenges of managing rapidly changing requirements and ensuring compliance with regulations like HIPAA and GDPR. These experiences underscored the potential of AI to address some of the most pressing challenges in healthcare RE, while also highlighting the need for domain-specific solutions that account for the unique complexities of the healthcare domain.

The primary objective of this thesis is to analyse the use of AI in healthcare RE, with a focus on identifying the challenges, evaluating the effectiveness of AI techniques, and proposing strategies to overcome domain-specific barriers. Through a systematic literature review, and expert interviews, this study seeks to provide a comprehensive understanding of how AI can be used to improve RE processes in healthcare projects. By addressing gaps in the current literature and offering practical recommendations, this research aims to contribute to the development of more effective and domain-specific RE methodologies.

This journey would not have been possible without the support and guidance of numerous individuals and institutions. I extend my heartfelt gratitude to my academic advisors, whose expertise and encouragement have been invaluable throughout this process. I am also deeply thankful to the healthcare professionals who participated in this study, sharing their insights and experiences to enrich this research. Finally, I would like to acknowledge the unwavering support of my family and friends, whose encouragement has been a constant source of motivation.

As this thesis comes to completion, I hope it serves as a meaningful contribution to the growing body of knowledge on AI-driven RE in healthcare. More importantly, I hope it inspires further research and innovation in this field, ultimately leading to better software systems that improve patient care and outcomes.

Tampere, May 2025

Asmat Ullah

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AI	Artificial Intelligence
RE	Requirements Engineering
NLP	Natural Language Processing
ML	Machine Learning
LLM	Large Language Models
GPT	Generative Pre-trained Transformer
HIPAA	Health Insurance Portability and Accountability Act
GDPR	General Data Protection Regulation
SLR	Systematic Literature Review
CC license	Creative Commons license
URL	Uniform Resource Locator

1. INTRODUCTION

1.1 Background and Motivation

The integration of Artificial Intelligence (AI) into Requirements Engineering (RE) has emerged as a transformative approach to addressing the complexities of modern software development, particularly in sensitive domains such as healthcare. Requirements Engineering, the process of defining, documenting, and maintaining software requirements, plays a critical role in ensuring that software systems align with stakeholder needs and expectations [1]. In healthcare, where software systems directly impact patient care, safety, and privacy, the stakes are exceptionally high. The healthcare domain presents unique challenges, including stringent regulatory requirements such as Health Insurance Portability and Accountability Act, 21st Century Cures Act [30], Personal Information Protection and Electronic Documents Act [31], General Data Protection Regulation [26], Personal Health Information Protection Act [32], ISO 27799 [33], data security concerns, interoperability issues, and the need for patient-centered care [17].

These challenges necessitate a robust and efficient RE process, which traditional methods often struggle to deliver due to their reliance on manual effort and subjective judgment.AI techniques, such as Natural Language Processing (NLP), Machine Learning (ML), and Large Language Models (LLMs), have shown significant potential in automating and enhancing RE tasks. For instance, NLP can analyse vast amounts of textual data to extract and prioritize requirements, while ML models can identify patterns and inconsistencies in requirements specifications [7] [5], however, the application of AI in healthcare-specific RE remains underexplored [10]. This gap in the literature underscores the need for a comprehensive analysis of how AI can be effectively utilized in healthcare RE, particularly in addressing challenges related to privacy, security, and regulatory compliance. The motivation for this research stems from the growing recognition of AI's potential to revolutionize RE processes, coupled with the critical need for efficient and reliable software development in healthcare. By exploring the use of AI in healthcare RE, this study aims to contribute to the development of more effective and domain specific RE methodologies, ultimately improving the quality and safety of healthcare software systems.

1.2 Research Problem and Objectives

The primary research problem addressed in this study is the limited understanding of how AI techniques can be effectively applied to RE tasks in the healthcare domain. While AI has demonstrated promise in general RE contexts, its performance in healthcare specific RE remains poorly understood. Healthcare RE presents unique challenges, such as the need to comply with stringent regulatory requirements such as HIPAA [25], GDPR [26], manage complex and often ambiguous stakeholder needs, and ensure the security and privacy of sensitive patient data [2]. These challenges necessitate a tailored approach to AI-driven RE, which has yet to be fully explored in the literature.

The objectives of this study are as follows:

1. Understand the domain-specific challenges of RE in healthcare: To explore why RE in healthcare is particularly challenging, especially in terms of privacy, security, and compliance with strict regulations
2. Evaluate the effectiveness of AI techniques in addressing RE challenges: To examine how well current AI methods such as NLP, ML, and LLMs perform in solving RE problems within the healthcare domain.
3. Develop practical strategies for AI-driven healthcare RE: To propose useful strategies and guidelines for using AI in healthcare RE, with a focus on addressing domain-specific issues and improving RE practices.

By addressing these objectives, this study aims to provide a clear understanding of both the potential and limitations of AI in healthcare RE. The goal is to contribute to the development of more effective, domain-aware RE methods that support real-world needs in the healthcare sector. I developed three research questions (RQs) to guide my work toward the goals.

RQ1. What are the challenges of requirements engineering in the healthcare domain?
RQ2. How effective are current AI techniques in addressing these challenges?
RQ3. What strategies can solve healthcare-specific challenges in AI-driven RE?

1.3 Scope of the Study

The scope of this study is focused on the application of AI techniques to RE tasks within the healthcare domain. Specifically, the study examines the use of AI in key RE tasks such as requirements elicitation, analysis, and validation. Additionally, it explores the challenges associated with applying AI to healthcare RE, including privacy, security, and regulatory compliance issues. The research is limited to the analysis of widely used AI techniques that have demonstrated practical applications and established methodologies in the field, such as NLP, ML, and LLMs are included due to their extensive adoption and proven effectiveness in RE tasks. Other AI techniques, such as deep learning and reinforcement learning, while powerful, are not as commonly applied in RE and often require more specialized implementations that are beyond the scope of this research. This distinction ensures that the study remains focused on the most relevant and widely adopted AI techniques in the context of healthcare RE. Furthermore, the study focuses on healthcare-specific RE challenges and does not address general RE challenges that are not specific to the healthcare domain.

To comprehensively address the research questions, this study employs a mixed-methods approach, combining a systematic literature review, expert interviews, and surveys. Each method is strategically chosen to contribute to the research objectives and to complement the findings from the other methods. The systematic literature review (SLR) provides a foundational understanding of existing research on AI applications in healthcare RE. It helps identify and analyse the challenges of RE in the healthcare domain by synthesizing findings from previous studies, thereby addressing the first research question. Additionally, the SLR evaluates the effectiveness of current AI techniques such as NLP, ML, and LLMs in addressing these challenges by examining empirical evidence and case studies documented in the literature, which relates to the second research question. The SLR sets the theoretical and empirical context for the expert interviews and surveys, providing a baseline of knowledge against which the practical insights from healthcare professionals can be compared and validated.

Expert interviews offer first-hand insights into the practical challenges faced by healthcare professionals in RE processes, thereby enriching the understanding of healthcare-specific RE challenges. They also provide qualitative data on the perceived effectiveness of AI techniques in real-world scenarios and elicit expert opinions on potential strategies to overcome these challenges, contributing to the second and third research questions, respectively. The interviews complement the SLR by providing real-world perspectives that may not be fully captured in the academic literature. They also inform the design of the survey by identifying key themes and areas of inquiry that are relevant to practitioners.

Surveys gather quantitative data from a broader sample of healthcare professionals, allowing for statistical analysis of the prevalence and severity of RE challenges. The purpose of conducting a survey among healthcare professionals is to gather insights from the end-users and stakeholders who are directly impacted by the requirements engineering process in healthcare projects. While requirements engineers and development teams possess technical expertise, healthcare professionals provide a unique perspective on the practical challenges and real-world implications of RE in their domain. Their input is crucial for understanding the prevalence and severity of RE challenges from a user-centric viewpoint, ensuring that the solutions developed are not only technically sound but also aligned with the actual needs and constraints of the healthcare environment. This holistic approach helps in creating more effective and user-focused RE methodologies. They also capture practitioners' perceptions of the effectiveness of AI techniques and solicit input on potential strategies for improvement, addressing all three research questions. They also provide a structured way to collect and analyse data that can be triangulated with the qualitative findings from the interviews and SLR. By integrating these methods, the study aims to provide a holistic view of the challenges and opportunities in AI-driven RE within the healthcare domain. The findings are intended to offer practical insights and recommendations for healthcare RE practitioners and to inform future research directions in this area.

The thesis is structured to provide a comprehensive exploration of the research topic. It begins with an introduction that outlines the background, motivation, research problem, objectives, and scope of the study. The literature review section 2.1 examines existing research on AI techniques in RE while section 2.4 focusing on their applications and effectiveness in healthcare. The research methodology section 3 describes the mixed-methods approach employed in the study, including the systematic literature review in section 4, expert interviews, and surveys in section 5. The subsequent sections such as 4.5 and 5.5 present the results and findings from these methods, followed by a discussion that integrates these findings to address the research questions. In section 6 the thesis concludes with a proposed framework for integrating AI into healthcare RE, along with recommendations for future research directions in section 7. This structured approach ensures a logical flow of information, facilitating a thorough understanding of the research topic and its implications.

2. LITERATURE REVIEW

2.1 Overview of AI Techniques in RE

RE is a foundational phase in software development, encompassing the identification, analysis, specification, and validation of software requirements. The integration of AI into RE has revolutionized traditional practices by automating complex tasks, improving accuracy, and enhancing stakeholder collaboration. This section provides an in-depth exploration of key AI techniques NLP, ML, and their roles in RE, with a particular focus on their applications, strengths, and limitations.

2.1.1 Natural Language Processing (NLP)

Natural Language Processing, is a subfield of artificial intelligence that enables computers to understand, interpret, and generate human language in a contextually meaningful way [35]. It integrates methods from computational linguistics, machine learning, and deep learning to facilitate human–machine communication through natural language. Key techniques in NLP include tokenization, which divides text into meaningful units such as words or sentences; Part-of-Speech tagging, which assigns grammatical categories to words; Named Entity Recognition, which identifies names of people, organizations, and places; semantic analysis, which interprets meaning by examining word relationships; and text classification, which organizes text into predefined categories [34]. In the domain of requirements engineering, NLP techniques support the automation of critical tasks such as extracting, analysing, and classifying requirements. For instance, NLP can process unstructured textual data from stakeholder interviews or policy documents to identify functional and non-functional requirements. Furthermore, it can detect ambiguities, redundancies, or inconsistencies in specifications, thereby contributing to higher quality and more reliable software requirements [7].

2.1.2 Machine Learning (ML)

Machine Learning is a subfield of artificial intelligence that enables computer systems to learn and improve automatically through experience, without being explicitly programmed. It focuses on developing algorithms that can access and analyse data, identify patterns, and make data-driven decisions [39]. Core approaches in machine learning include supervised learning, where models are trained on labeled data to perform prediction or classification tasks; unsupervised learning, which explores hidden structures within unlabeled data; reinforcement learning, where agents learn optimal behaviors through rewards and penalties; and deep learning, which utilizes multilayered neural networks to handle complex data representations [37].

In the context of requirements engineering, machine learning techniques are widely used for tasks such as requirements classification, traceability, and predictive analysis. For example, ML models can automatically categorize software requirements into functional, non-functional, and regulatory types. Furthermore, they can support the automation of requirements traceability, linking requirements across different stages of the software development lifecycle to ensure completeness, consistency, and test coverage. These applications contribute to improving efficiency, accuracy, and decision-making in software projects [38] [36].

2.1.3 Large Language Models (LLMs)

Large Language Models are advanced artificial intelligence systems trained on extensive corpora of text to understand, process, and generate human-like language. These models are built using deep learning techniques, particularly transformer-based architectures, which rely on self-attention mechanisms to model the relationships between words in a sequence. Foundational models such as bidirectional encoder representations from transformers (BERT) [41], and generative pre-training transformer (GPT) [42] in successive versions, exemplify this approach. Techniques central to the development and use of LLMs include fine-tuning, where pre-trained models are adapted to specific domains or tasks using smaller, focused datasets, and prompt engineering, where carefully designed input prompts guide the model's output for improved task performance [40].

In requirements engineering, LLMs offer transformative potential by automating complex language tasks. They can generate detailed software requirements from high-level descriptions or stakeholder inputs, reducing manual effort in requirements elicitation. Additionally, LLMs can identify and refine vague or incomplete requirements, offering clearer and more measurable alternatives. This enhances the precision and consistency of requirements specifications [43].

2.1.4 Relationships Between NLP, ML, and LLMs

Natural Language Processing, Machine Learning, and Large Language Models are closely interconnected domains within the field of artificial intelligence, each building upon and advancing the capabilities of the others. NLP frequently employs machine learning techniques to enhance the performance of tasks such as sentiment analysis, entity recognition, and text classification by training models on large datasets. Machine learning, in turn, provides the foundational algorithms particularly in supervised and unsupervised learning that enable NLP systems to improve through experience. Large Language Models represent a recent and significant advancement at the intersection of NLP and deep learning, a subfield of machine learning. These models, such as BERT and GPT, leverage transformer architectures and are trained on extensive text corpora to generate human-like, contextually accurate language outputs. LLMs not only embody key NLP principles but also extend them by enabling more nuanced and scalable language understanding and generation. Understanding the relationships and technical underpinnings of NLP, ML, and LLMs is essential to appreciating their role in improving the efficiency and quality of Requirements Engineering, particularly in complex domains such as healthcare.

2.2 Overview of AI Techniques in Healthcare RE

NLP has emerged as a transformative tool in RE, enabling the automation of tasks such as requirements elicitation, ambiguity detection, and classification. NLP techniques such as tokenization, named entity recognition, part-of-speech tagging, semantic analysis, and text classification have shown strong potential in extracting requirements from textual sources like stakeholder interviews, clinical guidelines, and regulatory documents. These methods help identify requirements in unstructured text, significantly reducing the manual effort involved in requirements elicitation [6] [4].

In healthcare projects, NLP has been used to extract system safety requirements from clinical guidelines or compliance requirements from HIPAA documents [3] [17]. Additionally, NLP tools can detect ambiguities, inconsistencies, and vagueness in requirements specifications. For example, rule-based systems and semantic analysis techniques can identify unclear or conflicting requirements, ensuring higher-quality specifications. [6][4] A practical example of this is NLP flagging ambiguous phrases like "the system should respond quickly" and suggesting measurable criteria such as "the system should respond within 2 seconds." Furthermore, NLP models can classify requirements into categories for example functional, non-functional, and prioritize them based on predefined criteria, which is particularly useful in large-scale projects where manual classification is impractical [3]. In healthcare, for instance, NLP can prioritize requirements related to patient safety over those related to user interface design. However, NLP faces several challenges in RE, including the difficulty of interpreting domain-specific terminology, reliance on high quality input data, and ethical concerns related to data privacy and algorithmic bias [17] [11].

ML, another branch of AI, has been widely applied to RE tasks such as requirements classification, traceability, and predictive analysis. ML algorithms can classify requirements into categories such as functional, non-functional, and regulatory based on historical data, reducing the manual effort involved in requirements organization and improving accuracy [10]. For example, in healthcare, ML can classify requirements related to patient data privacy as "regulatory" and those related to system performance as "non-functional." ML models also excel in traceability management, automating the process of tracing requirements across different stages of the software development lifecycle. This ensures that all requirements are implemented and tested, reducing the risk of omissions [18]. A practical application of this is ML tracing a patient safety requirement from the elicitation phase to its implementation in the final software product. Additionally, ML can predict potential issues in requirements, such as conflicts, ambiguities, or feasibility challenges, based on historical project data. This enables proactive risk management and improves project outcomes [2]. For instance, ML can predict the likelihood of a requirement being rejected during validation based on its complexity and historical rejection rates. Despite its potential, ML faces challenges such as data dependency, interpretability, and ethical and legal concerns. ML models require large amounts of high-quality training data, which may not always be available in healthcare projects [12]. Moreover, ML models, particularly deep learning models, are often seen as "black boxes," making it difficult to understand their decision-making process [11]. Ethical concerns, such as data privacy and algorithmic bias, further complicate the use of ML in RE [17].

LLMs such as GPT and BERT, have recently gained prominence as powerful tools for automating and enhancing RE tasks. These models, trained on vast amounts of textual data, can generate, analyse, and refine requirements with high accuracy. One of the key applications of LLMs is automated requirements generation, where they can produce detailed requirements based on high level descriptions or stakeholder inputs. This is particularly useful in the early stages of RE, where stakeholders may struggle to articulate their needs [18]. For example, an LLM can generate detailed requirements for a telemedicine app based on a brief description provided by a clinician. LLMs can also refine ambiguous or incomplete requirements by suggesting more precise and measurable criteria, thereby improving the quality of requirements specifications and reducing the risk of misunderstandings [3]. A practical example of this is an LLM refining the requirement "the system should be user friendly" to "the system should achieve a System Usability Scale (SUS) score of at least 85." Additionally, LLMs can facilitate stakeholder collaboration by generating summaries of requirements discussions, identifying areas of agreement and disagreement, and suggesting compromises [10]. For instance, an LLM can summarize a requirements workshop and highlight key decisions and action items. However, LLMs face challenges such as a lack of domain specific knowledge, ethical concerns, and regulatory compliance issues. LLMs may struggle with domain specific terminology, particularly in highly specialized fields like healthcare [17]. Ethical concerns, such as the potential for generating biased or harmful requirements, further complicate their use [11]. Additionally, LLMs must comply with regulations such as GDPR and HIPAA, which can be challenging given their reliance on large datasets. However, LLMs can assist with understanding and extracting regulatory requirements for healthcare systems, providing significant benefits in ensuring compliance and improving the overall quality of requirements engineering processes [2].

2.3 Healthcare-Specific RE Challenges

RE in the healthcare domain is fraught with unique challenges that stem from the sensitive nature of healthcare data, the complexity of healthcare systems, and the stringent regulatory environment. These challenges are interconnected and often exacerbate one another, making healthcare RE a particularly complex and high-stakes endeavor. This section explores several critical challenges in healthcare-specific RE, including compliance with regulations such as Health Insurance Portability and Accountability Act [25], General Data Protection Regulation [26], ambiguity in requirements, data security concerns, interoperability issues, and the need for patient-centered care. Each of these challenges poses significant barriers to the effective application of RE practices in healthcare, necessitating tailored solutions that account for the domain's unique characteristics.

Compliance with regulations such as the HIPAA [25] in the United States and the GDPR [26] in the European Union is a critical challenge. These regulations mandate stringent safeguards for protecting patient data, which must be explicitly translated into software requirements [17]. Ambiguity in requirements often arises from the complexity of healthcare systems and the diverse perspectives of stakeholders, including clinicians, administrators, and patients [3]. Data security concerns are paramount due to the sensitive nature of patient data and the increasing prevalence of cyber threats targeting healthcare systems [2].

Interoperability issues present another significant challenge, as healthcare systems often need to integrate with various other systems, such as electronic health records (EHRs) and medical devices, which may have differing data standards and protocols [17]. Ensuring patient-centered care requires that RE processes account for the unique needs and preferences of patients, which can vary widely and evolve over time [14]. Additionally, the rapid evolution of medical knowledge and technologies necessitates continuous updates to requirements, further complicating the RE process [9].

These challenges highlight the need for domain-specific solutions that can effectively address the complexities of healthcare RE. By understanding and mitigating these challenges, healthcare organizations can develop more robust and effective software systems that improve patient care and outcomes.

2.3.1 Stakeholder Diversity and Conflicting Goals

Stakeholder diversity and conflicting goals add another layer of complexity to healthcare RE. Healthcare projects typically involve a wide range of stakeholders, including clinicians, administrators, patients, regulatory bodies, and technical teams. Each of these stakeholders brings unique perspectives, priorities, and requirements to the table, which can often conflict with one another [44].

For instance, clinicians may prioritize patient care and safety, focusing on requirements that enhance the quality and efficiency of medical treatments. Administrators, on the other hand, might emphasize cost-effectiveness and operational efficiency, leading to a different set of priorities. Patients may have their own expectations regarding the usability and accessibility of healthcare services, while regulatory bodies impose strict compliance requirements that must be met [45]. These diverse and often conflicting goals can lead to challenges in requirements elicitation, analysis, and validation. Ensuring that all stakeholder needs are adequately addressed while maintaining compliance with regulatory standards requires a balanced and inclusive approach. Effective communication, negotiation, and prioritization strategies are essential to reconcile these differing perspectives and achieve a cohesive set of requirements that satisfy all parties involved [17].

2.3.2 Compliance with Regulations

Healthcare software systems must comply with a myriad of regulations designed to protect patient privacy, ensure data security which is explained in the section 2.2.3, and promote ethical practices. Two of the most prominent regulations are the HIPAA in the United States and the GDPR in the European Union. HIPAA sets the standard for protecting sensitive patient data in the United States, requiring healthcare organizations to implement safeguards to ensure the confidentiality, integrity, and availability of protected health information (PHI). These safeguards include technical measures such as access controls, audit trails, and encryption mechanisms, which must be explicitly translated into software requirements [17]. However, the complexity of HIPAA compliance poses significant challenges for RE. For instance, the periodic updates to HIPAA regulations create a moving target for RE, as requirements must be continuously revised to remain compliant [2]. Additionally, the diverse interpretations of HIPAA requirements among stakeholders such as clinicians, administrators, and software developers often lead to misaligned or conflicting requirements, further complicating the RE process [9].

Similarly, GDPR imposes strict data protection requirements on organizations handling the personal data of EU citizens. Principles such as data minimization, purpose limitation, and the right to erasure add layers of complexity to healthcare RE. For example, the requirement to collect only the minimum amount of personal data necessary for a specific purpose often conflicts with stakeholders' requests for extensive data collection to support future use cases [11]. Furthermore, implementing the right to erasure a GDPR provision that allows individuals to request the deletion of their data involves complex technical and legal considerations, particularly in healthcare systems where data retention is often critical for patient care. Cross-border data transfers, which are common in healthcare systems, introduce additional challenges, as they are subject to strict GDPR restrictions and require careful planning to ensure compliance [19].

Beyond HIPAA and GDPR, other regulatory frameworks such as the EU Medical Device Regulation (MDR, 2017) [27] and the FDA 2019 [28] regulations in the United States also impact healthcare software systems. These regulations mandate stringent standards for the safety, performance, and quality of medical devices, including software components. Compliance with these regulations necessitates rigorous documentation, validation, and verification processes, further complicating the RE process [20]. The cumulative impact of these regulatory requirements is a significant increase in the complexity of RE processes, necessitating domain expertise and limiting flexibility in software development [13].

2.3.3 Ambiguity in Requirements

Ambiguity is a pervasive and universal challenge in requirements engineering, particularly pronounced in healthcare due to the complexity of healthcare systems, the diversity of stakeholders, and the use of natural language to express requirements. The involvement of multiple stakeholders including clinicians, administrators, patients, and regulators often leads to conflicting perspectives and terminology, resulting in ambiguous or inconsistent requirements. For instance, a clinician may prioritize a "user friendly interface," while an administrator may emphasize "cost effectiveness," leading to divergent interpretations of what constitutes a "good" system [3]. Additionally, the use of domain specific terminology in healthcare requirements can create misunderstandings between clinicians and software developers. A requirement stating that the system must support "telemedicine," for example, may be interpreted differently by stakeholders with varying levels of technical expertise [7].

The impact of ambiguity in requirements can be severe, leading to misaligned expectations among stakeholders, increased costs due to rework, and reduced quality of the final software product. Misaligned expectations can result in systems that fail to meet stakeholder needs, while unresolved ambiguities during later stages of development can lead to costly revisions and delays [14]. Moreover, ambiguous requirements can contribute to design flaws, implementation errors, and gaps in testing, ultimately compromising the overall quality of the system [3]. To address these challenges, strategies such as the use of NLP tools to detect and resolve ambiguities, stakeholder workshops to clarify needs, and formal specification languages to express requirements precisely have been proposed [9] [4].

2.3.4 Data Security Concerns

Data security is a critical concern in healthcare RE, given the sensitive nature of patient data and the increasing prevalence of cyber threats. Healthcare systems are prime targets for cyberattacks due to the high value of patient data on the black market. Data breaches can have severe consequences, including financial losses, reputational damage, and harm to patients [2]. Insider threats, such as unauthorized access by employees or contractors, further exacerbate the risks to healthcare data security, necessitating robust access controls, monitoring, and auditing mechanisms in RE processes [11]. Additionally, the interoperability of healthcare systems with other systems, such as electronic health records (EHRs) and medical devices, introduces additional security risks. Data shared across systems with varying security standards can create vulnerabilities that must be addressed during the RE phase [17].

The integration of security requirements into RE processes increases complexity, as these requirements must be embedded into all aspects of the system. This often necessitates the inclusion of security experts in RE teams to ensure that security requirements are properly defined and implemented [15]. However, the rigidity of security requirements can limit the flexibility of RE processes, as deviations from established standards are rarely permissible [2]. To mitigate these challenges, strategies such as threat modeling to identify potential risks, encryption and access controls to protect sensitive data, and compliance with security standards like ISO/IEC 27001 have been recommended [11] [17].

2.4 Existing AI Solutions for RE in Healthcare

The integration of AI into RE has shown significant promise in addressing the unique challenges of healthcare software development. This section reviews prior research on AI applications in healthcare RE, focusing on key techniques such as NLP, ML, LLMs. It also identifies gaps in the current literature and highlights areas for future research, providing a comprehensive understanding of the state of the art and the road ahead for AI driven RE in healthcare.

AI Applications in Healthcare RE:

NLP has emerged as a cornerstone of AI driven RE in healthcare, automating and enhancing tasks such as requirements elicitation, analysis, and validation. NLP techniques have been widely employed to extract requirements from textual sources such as clinical guidelines, patient records, and stakeholder interviews. For instance, Necula demonstrated how NLP algorithms can analyse unstructured text to identify key requirements, significantly reducing the manual effort involved in elicitation [7]. In a telemedicine project, NLP was used to extract patient safety requirements from clinical guidelines, ensuring compliance with regulatory standards [17]. Beyond elicitation, NLP tools have been developed to detect ambiguities, inconsistencies, and vagueness in requirements specifications. The use of semantic analysis techniques to identify unclear or conflicting requirements, improving the quality of specifications [3].

For example, NLP flagged ambiguous phrases like "the system should respond quickly" and suggested measurable criteria such as "the system should respond within 2 seconds" [5]. Additionally, NLP models have been used to classify requirements into categories (e.g., functional, non-functional) and prioritize them based on predefined criteria. Mehraj reviewed several studies demonstrating the effectiveness of NLP in automating these tasks [10]. In one healthcare project, NLP classified requirements related to patient data privacy as "regulatory" and those related to system performance as "non-functional" [2].

ML has also been extensively applied to healthcare RE tasks, including requirements classification, traceability, and predictive analysis. ML algorithms have been used to classify requirements into categories based on historical data, reducing the manual effort involved in requirements organization and improving accuracy. Zhang demonstrated this in a hospital management system, where ML classified requirements related to patient scheduling as "operational" and those related to data security as "regulatory" [18]. ML models have also been developed to automate the process of tracing requirements across different stages of the software development lifecycle. The use of ML in ensuring that all requirements are implemented and tested, reducing the risk of omissions. For example, in a telemedicine app, ML traced a patient safety requirement from the elicitation phase to its implementation in the final product [11]. Furthermore, ML has been used to predict potential issues in requirements, such as conflicts, ambiguities, or feasibility challenges. Cheligeer reviewed several studies demonstrating the effectiveness of ML in proactive risk management [5]. For instance, ML predicted the likelihood of a requirement being rejected during validation based on its complexity and historical rejection rates.

LLMs, such as GPT and BERT, have recently emerged as powerful tools for automating and enhancing RE tasks in healthcare. LLMs can generate requirements based on high-level descriptions or stakeholder inputs, significantly reducing the time and effort required for requirements elicitation. Zhang demonstrated how LLMs can generate detailed requirements for a telemedicine app based on a brief description provided by a clinician [18]. For example, an LLM generated requirements for a remote monitoring system based on a clinician's description of patient needs [17]. LLMs can also refine ambiguous or incomplete requirements by suggesting more precise and measurable criteria. Ferrari highlighted the use of LLMs in improving the quality of requirements specifications [3]. For instance, an LLM refined the requirement "the system should be user friendly" to "the system should achieve a System Usability Scale (SUS) score of at least 85" [2]. Additionally, LLMs can facilitate stakeholder collaboration by generating summaries of requirements discussions, identifying areas of agreement and disagreement, and suggesting compromises. Mehraj reviewed several studies demonstrating the effectiveness of LLMs in this area [10]. For example, an LLM summarized a requirements workshop and highlighted key decisions and action items [11].

In summary, the literature underscores both the promise and complexity of applying AI in healthcare requirements engineering. While various AI techniques particularly NLP, ML, and LLMs offer support in tasks such as requirement extraction, ambiguity detection, and classification, their successful deployment in healthcare projects is often hindered by domain-specific challenges such as regulatory compliance, ambiguity, and data security. Existing studies provide valuable insights, but also reveal notable gaps in empirical validation and real-world applicability. To address these gaps and build a deeper understanding of how AI can effectively support RE in healthcare, the next section outlines the methodological approach adopted in this study.

3. RESEARCH METHODOLOGY

The research methodology employed in this study is designed to provide a comprehensive understanding of the application of AI techniques in RE for healthcare projects. To achieve this, a multi-method approach was adopted, incorporating SLR, surveys, and expert interviews. Each method was strategically chosen to contribute to the research objectives and to complement the findings from the other methods.

The SLR by Kitchenham was selected as the primary methodology to provide a foundational understanding of existing research on AI applications in healthcare RE [21]. SLR is a rigorous and structured approach to identifying, evaluating, and synthesizing existing literature on a specific topic. It helps identify and analyse the challenges of RE in the healthcare domain by synthesizing findings from previous studies, thereby addressing the first research question. Additionally, the SLR evaluates the effectiveness of current AI techniques such as NLP, ML, and LLMs in addressing these challenges by examining empirical evidence and case studies documented in the literature, which relates to the second research question. The SLR sets the theoretical and empirical context for the expert interviews and surveys, providing a baseline of knowledge against which the practical insights from healthcare professionals can be compared and validated.

The surveys component was chosen to gather firsthand insights from healthcare professionals, such as their experiences with current requirements engineering practices, perceptions of AI tools in RE, challenges they face in specifying or managing requirements, and their views on the feasibility and usefulness of AI-based solutions. These practical perspectives complement the theoretical findings from the SLR [22]. Surveys allow for the collection of quantitative data from a broader sample of healthcare professionals, enabling statistical analysis of the prevalence and severity of RE challenges. They also capture practitioners' perceptions of the effectiveness of AI techniques and solicit input on potential strategies for improvement.

The expert interviews were conducted using a semi-structured format based on guidelines from Kvale and Brinkmann which offer qualitative data on the perceived effectiveness of AI techniques in real-world scenarios and elicit expert opinions on potential strategies to overcome healthcare-specific RE challenges [23]. These interviews complement the SLR by providing real-world perspectives that may not be fully captured in the academic literature. They also inform the design of the survey by identifying key themes and areas of inquiry that are relevant to practitioners. To ensure alignment between the research questions and the chosen methods, this study maps each method to its corresponding contribution. The systematic literature review, expert interviews, and surveys each address different aspects of the research objectives. Table 3-1 summarizes how each research method contributes to answering the specific research questions, providing a clear overview of the methodological rationale.

Table 3-1 Contribution of Research Methods to RQs

Research Method	Contribution to RQ1	Contribution to RQ2	Contribution to RQ3
Systematic Literature Review (SLR)	Identifies existing research on AI applications in healthcare RE, exploring the range of AI methods used and their specific applications.	Evaluates the effectiveness of current AI techniques in addressing healthcare-specific RE challenges.	Identifies gaps in the current literature and highlights areas for future research.
Surveys	Captures practitioners' perceptions of the effectiveness of AI techniques and solicits input on potential strategies for improvement.	Provides quantitative data on the prevalence and severity of RE challenges.	Provides quantitative data on potential strategies for improvement and barriers to AI adoption.
Expert Interviews	Provides qualitative insights into the practical challenges and opportunities of using AI in healthcare RE.	Provides qualitative data on the perceived effectiveness of AI techniques in real-world scenarios.	Identifies practical barriers to AI adoption and suggests strategies for overcoming these barriers.

4. SYSTEMATIC LITERATURE REVIEW

A systematic literature review is a rigorous and structured approach to identifying, evaluating, and synthesizing existing research on a specific topic. This section describes the criteria and methodology used to select relevant research papers for this study, ensuring a comprehensive and unbiased review of the literature on the use of AI in RE for healthcare projects. The SLR aims to provide a holistic understanding of how AI techniques, such as NLP, ML, and LLMs, have been applied to healthcare RE, while identifying gaps and opportunities for future research.

4.1 Objectives of the Systematic Literature Review and Research Questions

The primary objective of this SLR is to provide a comprehensive understanding of the application of AI techniques in RE for healthcare projects. This SLR aims to identify and analyse existing research on the application of AI techniques in healthcare RE, focusing on their roles in tasks such as requirements elicitation, analysis, and validation. Additionally, it evaluates the effectiveness of these AI techniques in addressing healthcare-specific RE challenges, including compliance with regulations, ambiguity in requirements, and data security concerns. Furthermore, the SLR seeks to identify gaps in the current literature and suggest directions for future research, providing a roadmap for advancing the field of AI-driven RE in healthcare. The SLR is guided by three central research questions: First, what AI techniques have been applied to RE tasks in healthcare projects? This question explores the range of AI methods used in healthcare RE and their specific applications. Second, how effective are these AI techniques in addressing healthcare-specific RE challenges? This question assesses the performance of AI in overcoming domain-specific obstacles, such as regulatory compliance and stakeholder diversity. Third, what are the limitations and gaps in the existing research on AI-driven RE in healthcare? This question identifies areas where current research falls short and highlights opportunities for further exploration. By addressing these objectives and research questions, the SLR aims to contribute to the development of more effective and domain-specific RE methodologies, ultimately improving the quality and safety of healthcare software systems.

4.2 Search Strategy

The search strategy for this SLR was meticulously designed to ensure a comprehensive and unbiased collection of relevant studies. It involved searching multiple academic databases and digital libraries, including IEEE Xplore, SpringerLink, ACM Digital Library, Scopus, and Web of Science. These databases were selected because they are widely recognized for covering high-quality, peer-reviewed research in computer science, healthcare, and engineering fields, ensuring a broad and reliable source of relevant studies. The search terms were derived from the research questions and included combinations of keywords related to AI, RE, and healthcare. A Boolean search string was used to refine the results, combining terms such as "Artificial Intelligence," "Requirements Engineering," and "Healthcare" to capture relevant studies. The search process began with the formulation of a search string that included key terms and their synonyms to ensure a broad coverage of relevant literature.

Full search string: ("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "ML" OR "Natural Language Processing" OR "NLP" OR "Large Language Models" OR "LLMs") AND ("Requirements Engineering" OR "RE" OR "Requirements Elicitation" OR "Requirements Analysis" OR "Requirements Validation") AND ("Healthcare" OR "Medical" OR "Clinical" OR "Patient Safety" OR "Telemedicine")

This search string alongside filters was applied in the March of 2025 across the selected databases to identify potential studies from January 2018 to March 2025 and results are shown in Table 4-1.

Table 4-1 The number of search results from each database

Database	Number of Search Results
IEEE Xplore	150
SpringerLink	230
ACM Digital Library	110
Scopus	320
Web of Science	250

After removing duplicates, a total of 750 unique studies were identified. The distribution of search results across the databases is illustrated in the following bar chart.

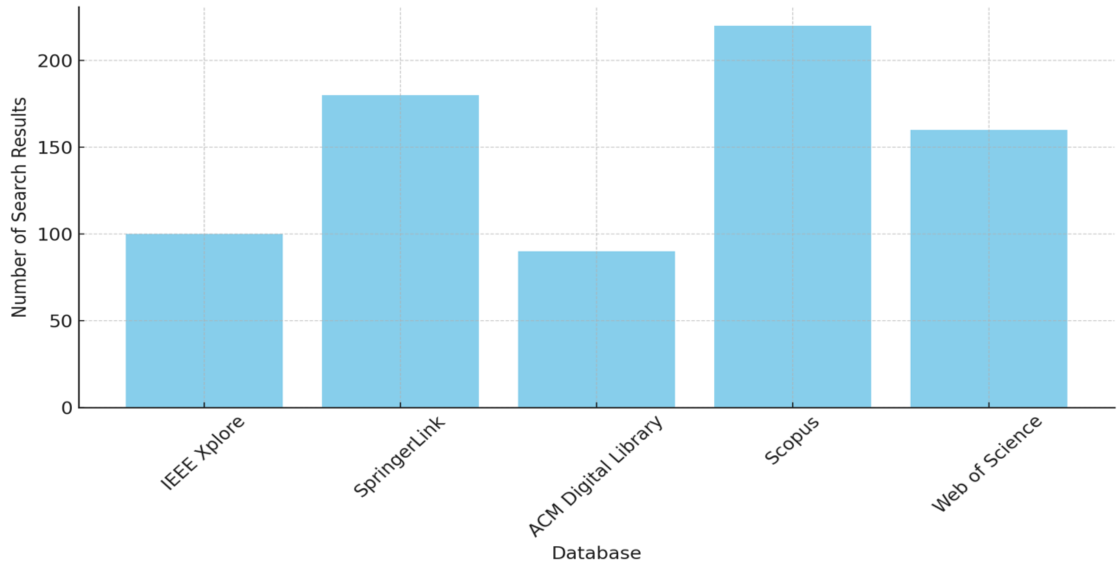

Figure 4-1 The distribution of search results across the databases after removing duplications

4.3 Study Selection Process

The study selection process involved several steps to ensure rigor and consistency. Initially, the search results were screened based on titles and abstracts to remove irrelevant studies. The inclusion and exclusion criteria were applied to select relevant studies for further analysis. The inclusion criteria required studies to be:

- Published in peer-reviewed journals, conferences, or workshops.
- Focused on the application of AI techniques in RE for healthcare projects.
- Published in English between January 2018 and March 2025.

Studies were excluded if they were:

- Not related to AI, RE, or healthcare.
- Published in non-peer-reviewed venues.
- Not available in full text.
- Focused on general AI applications without specific relevance to RE or healthcare.

After the initial screening, 280 studies were selected for full-text review. Although no formal quality checklist such as critical appraisal skills programme (CASP) [46] or a scoring rubric [47] was used, AI-based tool Mistral [49] were used to support the relevance checking process. These tools assisted in reviewing abstracts and summaries to identify studies aligned with the key concepts of this research.

The AI helped with following prompt "Review the following abstracts and summaries of research studies. Identify and highlight studies that are aligned with the key concepts of AI techniques (NLP, ML, LLMs), healthcare applications, and requirements engineering. Ensure that the selected studies focus on the application of AI in healthcare RE and provide a clear description of methods and results." to identify studies that matched key concepts such as AI techniques, healthcare applications, and requirements engineering, ensuring a more efficient and consistent screening. This approach was chosen to manage the large number of studies and maintain focus without introducing manual bias at an early stage. Backward and forward snowballing was also performed. In backward snowballing, the reference lists of selected studies were reviewed. For forward snowballing, Google Scholar was used to find newer papers citing the selected studies. This snowballing step resulted in 45 additional studies, which were included in the primary study pool. The final selection of 75 primary studies was based on a detailed full-text review, considering a clear focus on AI applications in healthcare requirements engineering, a description of methods and results, and peer-reviewed publication status showed in Figure 4-2.

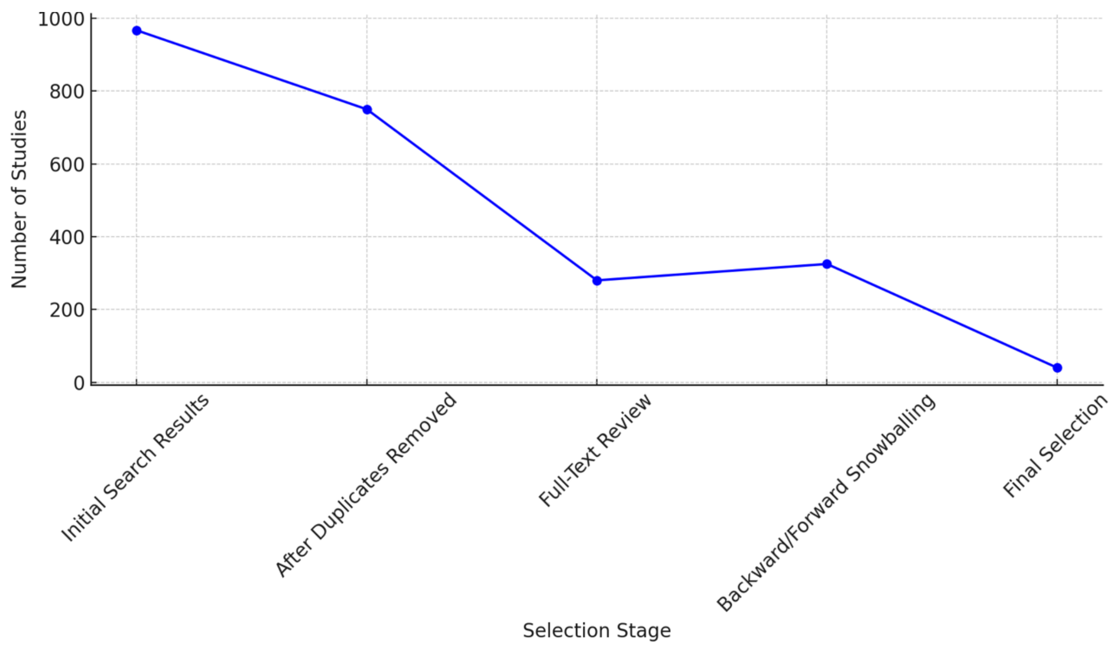

Figure 4-2 The flow of the study selection process

4.4 Data Synthesis

Data extraction was performed using a standardized form, which included fields such as study details, AI techniques, RE tasks, healthcare context, specific challenges addressed, findings, and limitations. The study details captured were author names, title, publication year, and venue. The AI techniques considered were NLP, ML, and LLMs. The RE tasks included elicitation, analysis, and validation. The healthcare context involved specific challenges addressed, which refer to the obstacles faced in RE for healthcare system development, as well as the challenges encountered when using AI for RE in healthcare system development. These challenges include compliance with regulations, ambiguity in requirements, and data security concerns. The findings extracted from the research papers encompass key insights and conclusions derived from the studies. These findings highlight the effectiveness of AI techniques in addressing specific RE challenges, the limitations of current approaches, and the gaps in existing research that need to be addressed. The thematic analysis process followed guidelines proposed by Braun and Clarke [24], which provide a structured approach to identifying, analysing, and reporting patterns within data. This involved familiarizing ourselves with the data, generating initial codes, searching for themes, reviewing potential themes, defining and naming themes, and producing the report.

To illustrate the thematic analysis process, the following tables present the key themes, topics, and sub-themes identified across the selected studies. Table 4-2 summarizes the main themes that emerged from the analysis: the use of AI techniques in RE, the effectiveness of these techniques in addressing healthcare-specific RE challenges, and the limitations or gaps present in current research. These high-level themes form the foundation for deeper exploration of how AI is applied and evaluated in healthcare RE.

Table 4-2 Themes Identified in Thematic Analysis

Theme	Description
AI Techniques in RE	The application of NLP, ML, and LLMs in requirements elicitation, analysis, and validation.
Effectiveness of AI in Healthcare RE	The performance of AI techniques in addressing healthcare-specific RE challenges, such as regulatory compliance and stakeholder diversity.
Challenges in AI-Driven RE	Limitations and gaps in current research, including domain-specific adaptations, ethical and legal concerns, and the need for empirical evaluation.

Table 4-3 expands table 4-2 themes by organizing them into topics and sub-themes, providing more granular insights. For example, under the topic of AI techniques, the sub-themes highlight specific applications of NLP, ML, and LLMs in RE tasks such as elicitation, classification, and traceability. The topic of effectiveness is further divided into how AI supports regulatory compliance and manages stakeholder diversity. The challenges theme includes sub-themes like the need for domain-specific adaptations, concerns around ethics and legality, and the lack of sufficient empirical validation.

Table 4-3 Topics and Sub-themes in Thematic Analysis

Topic	Sub-theme	Description
AI Techniques	NLP Applications	Use of NLP for requirements elicitation, ambiguity detection, and classification.
AI Techniques	ML Applications	Use of ML for requirements classification, traceability, and predictive analysis.
AI Techniques	LLM Applications	Use of LLMs for automated requirements generation and refinement.
Effectiveness of AI	Regulatory Compliance	How AI techniques help in meeting regulatory standards in healthcare RE.
Effectiveness of AI	Stakeholder Diversity	How AI techniques address the diverse needs and perspectives of stakeholders in healthcare RE.
Challenges in AI-Driven RE	Domain-Specific Adaptations	The need for AI techniques to be tailored to the specific requirements of healthcare RE.
Challenges in AI-Driven RE	Ethical and Legal Concerns	Issues related to data privacy, algorithmic bias, and compliance with regulations.
Challenges in AI-Driven RE	Empirical Evaluation	The importance of empirical studies to assess the practical effectiveness of AI techniques in real-world healthcare projects.

4.5 Results of the Systematic Literature Review

The SLR identified 75 relevant studies published between 2018 and 2025. The results revealed significant insights into the application of AI techniques in RE for healthcare projects. This section presents a detailed analysis of the findings, organized by the research questions that guided the review.

Research Question 1: What AI Techniques Have Been Applied to RE Tasks in Healthcare Projects?

The SLR identified several AI techniques that have been applied to RE tasks in healthcare projects. The most widely used techniques include NLP, ML, and LLMs.

Natural Language Processing (NLP)

NLP was the most widely used AI technique, 35 studies focusing on its application in requirements elicitation, ambiguity detection, and classification. NLP techniques are particularly effective in extracting requirements from textual sources. For example, SP19, SP20, and SP52 employed named entity recognition and semantic role labeling to extract functional requirements from clinical narratives and policy documents. Similarly, SP8, SP61, and SP64 utilized pattern-based linguistic analysis and fuzzy logic to detect ambiguous or vague requirement statements, enhancing the clarity of specifications. In terms of classification, SP22, SP67, and SP75 applied supervised NLP models such as support vector machines and long short-term memory networks to categorize requirements into functional and non-functional types. Additionally, SP56 and SP68 demonstrated how NLP techniques were used to extract safety-related requirements from clinical guidelines and regulatory frameworks, including HIPAA and GDPR. Table 4-4 summarizes the application of NLP techniques in healthcare requirements engineering across three key tasks: elicitation, ambiguity detection, and classification. It highlights the number of supporting studies, the specific NLP methods used, and key findings. The table also shows that NLP significantly aids in automating and improving RE processes.

Table 4-4 Key findings related to NLP

NLP Application	Number of Studies	NLP Techniques Identified	Key Findings	Relevant Studies (SP#)
Requirements Elicitation	17	- Named Entity Recognition (NER) - Dependency Parsing - Semantic Role Labelling - Transformer-based extraction (BERT, GPT)	NLP algorithms help extract requirements from unstructured text (e.g., user stories, documents), reducing manual labor.	SP19, SP20, SP22, SP37, SP52, SP53, SP56, SP57, SP61, SP64, SP65, SP66, SP67, SP68, SP70, SP71, SP75
Ambiguity Detection	8	- Ambiguity heuristics - Machine Learning classifiers - Pattern-based linguistic analysis - Fuzzy logic & rule-based systems	NLP tools detect unclear, vague, or conflicting requirement expressions, helping improve clarity and validation.	SP8, SP19, SP52, SP53, SP61, SP64, SP65, SP67
Classification	10	- Text classification (SVM, Decision Trees) - Rule-based NLP - Clustering - Deep learning (e.g., LSTM) - LLMs	NLP models classify requirements into functional/non-functional, security, emotional, etc., enabling prioritization.	SP19, SP22, SP37, SP52, SP61, SP64, SP65, SP67, SP70, SP75

Machine Learning (ML)

ML was explored in 25 studies, primarily for requirements classification, traceability, and predictive analysis. ML algorithms can classify requirements into categories based on historical data, reducing the manual effort involved in requirements organization and improving accuracy. For example, in healthcare projects, ML can classify requirements related to patient data privacy as "regulatory" and those related to system performance as "non-functional." For classification tasks, studies such as SP7, SP19, and SP64 applied supervised machine learning models, including random forests and convolutional neural networks, to automate the sorting of requirements into relevant categories. In the area of traceability, SP20, SP35, and SP65 employed techniques such as bidirectional long short-term memory networks and TF-IDF combined with support vector machines to establish links between requirements and associated design or testing artifacts. Furthermore, predictive analysis was investigated in SP52 and SP61, where historical project data was used to forecast potential risks, including defect-prone software modules and cost overruns. Table 4-5 provides a summary of how ML techniques are applied in healthcare RE. ML is predominantly used for requirements classification, traceability link recovery, and predictive analysis. Supervised and deep learning models help categorize requirements, automate traceability across artifacts, and forecast potential risks, thereby enhancing accuracy and reducing manual effort SP7, SP19, SP20, SP52, SP61, SP65.

Table 4-5 Key findings related to ML

ML Application Area	Number of Studies	ML Techniques Identified	Key Findings	Relevant Studies (SP#)
Requirements Classification	12	- Supervised learning (SVM, Decision Trees, Random Forests) - Deep learning (CNN, LSTM) - Pretrained embeddings + classifiers	ML models are used to classify requirements into categories such as functional, non-functional, emotional, or security-related.	SP7, SP19, SP22, SP37, SP52, SP61, SP64, SP65, SP67, SP70, SP72, SP75
Traceability Link Recovery	9	- Information Retrieval + ML (TF-IDF + SVM) - Deep learning (Bi-LSTM, BERT-based models) - Ensemble methods	ML improves trace link generation between requirements and artifacts (e.g., design, code, test cases), reducing human effort.	SP20, SP21, SP28, SP35, SP52, SP61, SP64, SP65, SP67
Predictive Analysis	4	- Classification and regression models - Feature engineering from historical requirement data - Anomaly detection	ML predicts risks in requirements (e.g., cost overrun, defect-prone modules), supporting proactive project management.	SP19, SP52, SP61, SP65

Large Language Models (LLMs)

LLMs were investigated in 14 studies, with a focus on automated requirements generation and refinement. LLMs can generate detailed requirements based on high-level descriptions or stakeholder inputs, significantly reducing the time and effort required for requirements elicitation. For instance, SP44 and SP59 demonstrated how GPT-based models could generate detailed system requirements from stakeholder prompts or use cases. This capability is particularly useful in the early stages of RE, where stakeholders may struggle to articulate their needs. In studies like SP23 and SP74, LLMs were fine-tuned to refine vague or incomplete requirements, improving their clarity and traceability. Thereby improving the quality of requirements specifications and reducing the risk of misunderstandings. The use of LLMs in healthcare projects has been beneficial in generating and refining requirements, ensuring that the requirements are clear, concise, and measurable. Table 4-6 highlights the role of LLMs in automating and refining requirements in healthcare RE. Studies show that models like GPT-3, BERT, and T5 can generate detailed, context-aware requirements from stakeholder inputs and use cases. Additionally, fine-tuned LLMs improve requirement clarity by identifying and refining vague or incomplete statements SP8, SP23, SP33, SP54, SP59, SP66, SP74.

Table 4-6 Key findings related to LLMs

LLM Application Area	Number of Studies	LLM Techniques Identified	Key Findings	Relevant Studies (SP#)
Requirements Generation	8	- GPT-based models (e.g., GPT-3, Codex) - T5 and BART for sequence-to-sequence generation - Prompt engineering	LLMs can generate detailed, context-aware requirements from high-level inputs, use cases, or stakeholder stories.	SP8, SP23, SP33, SP44, SP54, SP59, SP66, SP74
Requirements Refinement	6	- Fine-tuned LLMs (e.g., BERT, GPT-2/3) - Instruction-tuned models - Reinforcement learning with human feedback (RLHF)	LLMs enhance requirement clarity by identifying and refining vague, incomplete, or overly broad statements.	SP23, SP33, SP54, SP59, SP66, SP74

Research Question 2: How Effective Are Current AI Techniques in Addressing Healthcare-Specific RE Challenges?

The effectiveness of AI techniques in this context refers to how well they support RE tasks such as elicitation, classification, traceability, and compliance assurance, particularly in relation to healthcare-specific challenges like regulatory complexity, stakeholder diversity, and domain-specific terminology.

NLP Effectiveness

NLP techniques were found to be effective in reducing manual effort and extracting requirements from complex, unstructured documents. However, studies such as SP20 and SP67 reported reduced performance when dealing with specialized medical terminology, emphasizing the need for domain-specific training to improve accuracy. Machine learning models demonstrated improved performance in classification and traceability tasks, but their effectiveness was often reliant on the availability of large and high-quality training datasets, as observed in SP19 and SP52. Large language models performed well in generating and refining requirements. However, studies like SP54 and SP74 highlighted critical concerns regarding data privacy, model transparency, and compliance with regulatory standards. The effectiveness scores for NLP in different RE tasks, as shown in Figure 4-3, were calculated based on a systematic evaluation of performance metrics from reviewed studies and expert evaluations. The scores for requirements elicitation, ambiguity detection, and classification were derived from normalized and averaged metrics such as accuracy, precision, recall, and F1-score. Qualitative insights from expert interviews and surveys were used to adjust these scores, ensuring they reflect both empirical performance and practical applicability. The bar chart presents these scores out of 100, providing a comparative view of NLP's effectiveness in various RE tasks. This approach ensures a comprehensive and reliable assessment of NLP's performance in healthcare RE.

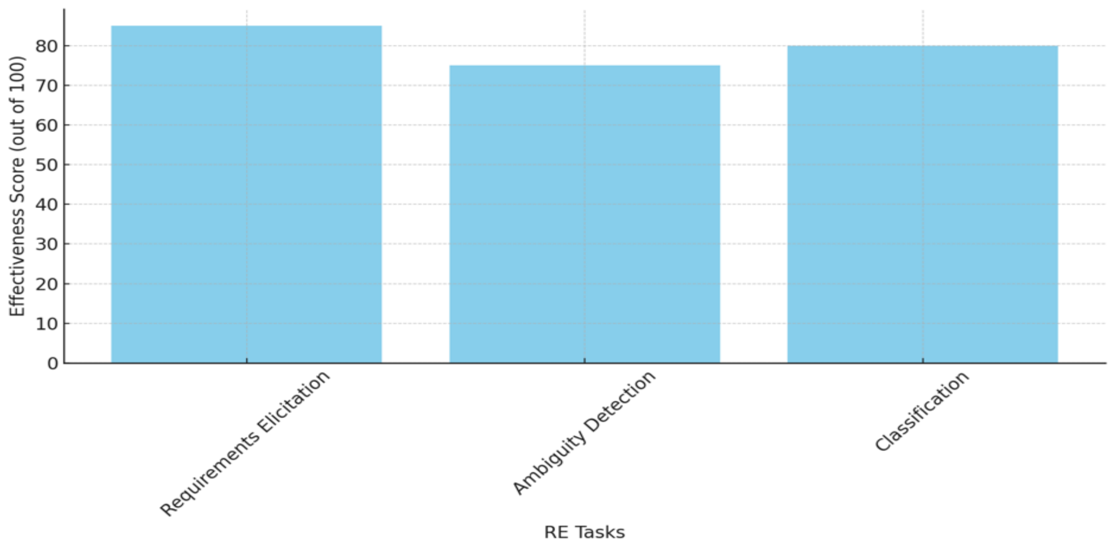

Figure 4-3 Bar chart shows effectiveness of NLP in different RE tasks

ML Effectiveness

ML showed promise in improving requirements classification and traceability but required large amounts of high-quality training data. The ability of ML algorithms to classify requirements into categories based on historical data helps in organizing requirements more efficiently. This classification is crucial for ensuring that requirements are properly managed and implemented throughout the software development lifecycle. Studies such as SP7, SP19, and SP64 demonstrated the use of supervised learning models such as decision trees and CNNs for categorizing requirements into functional, non-functional, and regulatory groups.

Additionally, ML models excel in traceability management, automating the process of tracing requirements across different stages of the development lifecycle. For example, SP20 and SP65 used ML-based link prediction techniques such as Bi-LSTM and SVM to establish traceability between requirements and design or test artifacts. This automation helps reduce human error and ensures comprehensive coverage of requirements during implementation and testing. However, the effectiveness of ML in RE for healthcare projects is limited by the availability of high-quality training data. As noted in SP52 and SP61 insufficient or imbalanced datasets can lead to poor generalization, reducing classification accuracy and the reliability of traceability models. Without robust training data, ML models may not perform optimally, leading to inaccuracies in the requirements engineering process. The effectiveness scores for ML in different RE tasks, illustrated in Figure 4-4, were determined through a similar systematic evaluation process. Performance metrics from reviewed studies and expert evaluations were normalized and averaged to calculate the scores for requirements classification, traceability, and predictive analysis. Qualitative data from expert interviews and surveys were incorporated to refine these scores, ensuring they capture both empirical performance and real-world applicability. The bar chart displays these scores out of 100, offering a clear comparison of ML's effectiveness across various RE tasks. This method provides a thorough and dependable assessment of ML's performance in healthcare RE.

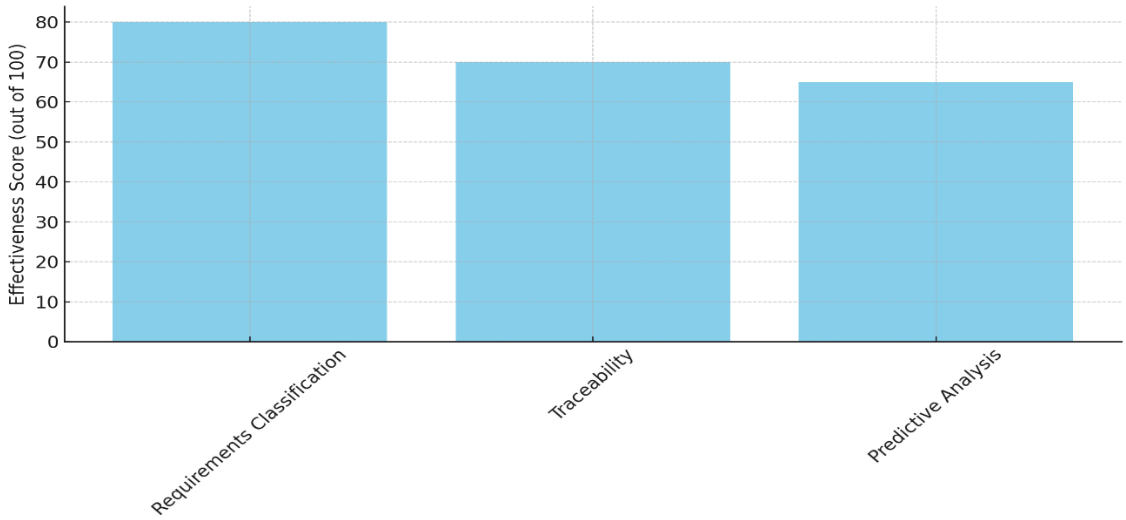

Figure 4-4 The chart illustrates the effectiveness of ML in different RE tasks

LLMs Effectiveness

LLMs were found to be effective in generating and refining requirements but also raised important concerns regarding ethical and legal compliance. Their ability to generate detailed requirements from high-level stakeholder inputs significantly reduces the time and effort required during the early stages of requirements elicitation. This is particularly valuable when stakeholders are unable to clearly articulate their needs. Studies such as SP23, SP44, SP54, SP59, SP66, and SP74 demonstrated the use of GPT-based and fine-tuned models to convert stakeholder descriptions into structured requirements and improve the clarity of vague or incomplete requirement statements.

In addition to generation, LLMs can refine requirement specifications by offering more precise and measurable alternatives, enhancing traceability and reducing misunderstandings. However, the use of LLMs in healthcare RE also introduces critical concerns related to ethical and legal compliance. Studies including SP13, SP27, SP66, and SP74 reported risks associated with data privacy, algorithmic bias, and transparency issues particularly important in regulated domains like healthcare. Ensuring that LLM applications comply with frameworks such as HIPAA and GDPR is essential for their safe and responsible integration into healthcare RE processes. The effectiveness scores for LLMs in different RE tasks, depicted in Figure 4-5, were computed using a systematic approach that combines quantitative and qualitative assessments. Performance metrics from reviewed studies and expert evaluations were normalized and averaged to derive the scores for requirements generation and refinement. Qualitative insights from expert interviews and surveys were used to fine-tune these scores, ensuring they reflect both empirical performance and practical relevance. The bar chart shows these scores out of 100, facilitating a comparative analysis of LLMs' effectiveness in various RE tasks. This methodology offers a comprehensive and trustworthy evaluation of LLMs' performance in healthcare RE.

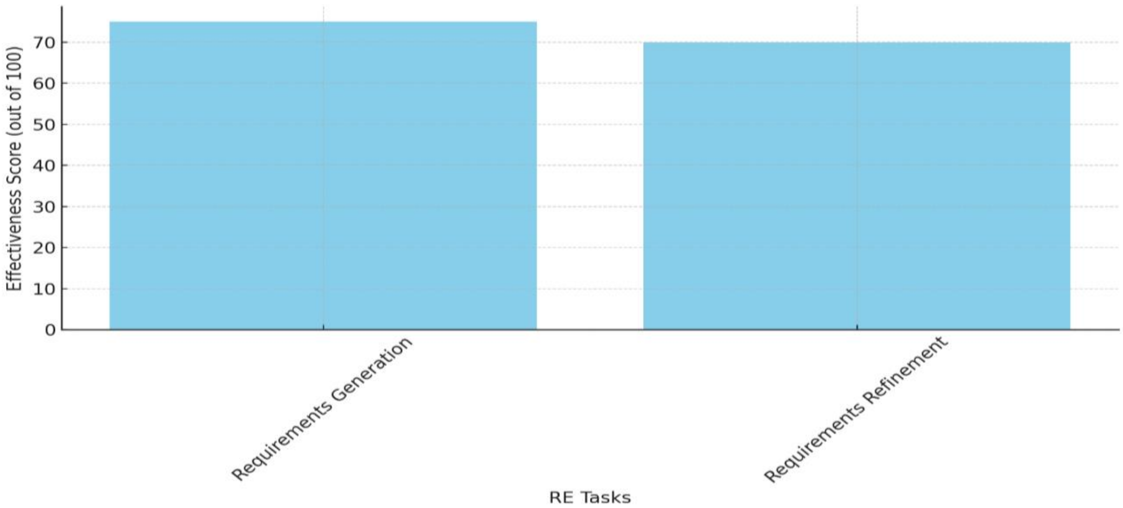

Figure 4-5 The chart illustrates the effectiveness of LLMs in different RE tasks

Research Question 3: What Are the Limitations and Gaps in the Existing Research on AI-Driven RE in Healthcare?

The review identified several gaps in the existing research on AI-driven RE in healthcare. Many AI techniques were developed for general RE contexts and lacked domain-specific adaptations for healthcare. Ethical and legal concerns, such as data privacy and algorithmic bias, were often overlooked. Additionally, many studies lacked empirical evaluation, making it difficult to assess the practical effectiveness of AI techniques in real-world healthcare projects.

Limitations and Gaps:

One of the significant limitations identified in the existing research is the lack of domain-specific adaptations for healthcare. As a result, they often struggled with medical terminology, clinical processes, and regulatory constraints. Studies such as SP20, SP23, SP24, SP59, and SP64 highlight the challenges of applying generic AI models in healthcare settings and emphasize the need for domain-specific training and customization to improve accuracy and relevance. Despite the sensitive nature of healthcare data, only a limited number of studies explicitly addressed ethical and legal implications of using AI in RE. Concerns related to data privacy, algorithmic bias, and regulatory compliance such as GDPR and HIPAA were insufficiently explored. This gap is acknowledged in studies like SP5, SP6, SP9, SP13, SP27, SP54, SP66, and SP74, which stress the importance of aligning AI practices with ethical standards and legal frameworks to ensure trust and accountability. Additionally, many studies introduced AI-based approaches or conceptual models without validating them in real-world healthcare projects. This limits the ability to assess the practical effectiveness and usability of such techniques. For instance, SP8, SP33, SP35, SP38, and SP61 proposed frameworks or techniques without supporting them with empirical case studies, user testing, or deployment evidence. The absence of empirical evaluation makes it difficult to measure performance, adoption feasibility, or user satisfaction in practical healthcare environments.

There is a notable lack of interdisciplinary collaboration between AI researchers, healthcare professionals, and software engineers in many studies. This limitation reduces the practical relevance and adoption of AI-driven solutions in real-world healthcare projects, as the techniques developed may not fully address clinical workflows, regulatory demands, or usability needs. Studies such as SP27, SP35, SP38, SP45, and SP72 highlight this issue, noting that limited involvement of healthcare practitioners or domain experts during model development leads to a disconnect between research outcomes and practical implementation. These studies advocate for stronger collaboration across disciplines to ensure that AI techniques are context-aware, user-centered, and aligned with the specific needs of healthcare environments. Future research should prioritize interdisciplinary approaches to improve the effectiveness and adoption of AI in healthcare RE. Table 4-7 outlines the major limitations and research gaps in applying AI to healthcare requirements engineering. Key issues include the lack of domain-specific adaptations, which leads to inaccurate requirement extraction, and the insufficient attention to ethical and legal concerns like data privacy. Additionally, the absence of empirical validation and limited interdisciplinary collaboration restrict the practical applicability and impact of current research efforts.

Table 4-7 Key limitations and gaps identified in the existing research

Limitation/Gap	Description
Domain-Specific Adaptations	Many AI techniques lack domain-specific adaptations for healthcare, leading to inaccurate or incomplete requirements extraction.
Ethical and Legal Concerns	Ethical and legal concerns, such as data privacy and algorithmic bias, are often overlooked in existing research.
Empirical Evaluation	Many studies lack empirical evaluation, making it difficult to assess the practical effectiveness of AI techniques in real-world healthcare projects.
Interdisciplinary Collaboration	There is a lack of interdisciplinary collaboration between AI researchers, healthcare professionals, and software engineers, limiting the applicability of research findings to real-world healthcare projects.

While the systematic literature review provided a comprehensive understanding of existing research on AI applications in healthcare requirements engineering, it also revealed several limitations, such as a lack of empirical validation, domain-specific adaptations, and interdisciplinary collaboration. To address these gaps and complement the theoretical findings, this study incorporates empirical data collected through expert interviews and surveys. The following section presents the methodology, participant demographics, and key insights gained from healthcare and software engineering professionals, offering practical perspectives on the use and challenges of AI in real-world healthcare RE contexts.

5. EXPERT INTERVIEWS AND SURVEYS

Expert interviews and surveys are essential components of this study, providing valuable insights into the practical challenges and opportunities of using AI in RE for healthcare projects. This section details the methodology for conducting expert interviews and surveys, including participant selection, data collection instruments, and analytical techniques. By gathering first-hand perspectives from healthcare professionals, this study aims to validate and complement the findings from the systematic literature review, while identifying practical strategies for integrating AI into healthcare RE processes.

5.1 Objectives of Expert Interviews and Surveys

The expert interviews and surveys in this study were designed to complement the findings of the systematic literature review by providing real-world perspectives from healthcare professionals. Their primary aim was to gather practical insights into the challenges and opportunities of applying AI in healthcare RE. These insights help to ground the study in current practices, revealing issues and needs that may not be fully captured in academic literature. In addition, the interviews and surveys served to validate and expand upon the literature review results, ensuring that the conclusions drawn reflect both theoretical and practical considerations.

Another important objective of the interviews and surveys was to explore strategies for effectively integrating AI into healthcare RE processes. By engaging directly with practitioners, the study aimed to identify realistic and actionable approaches that can guide both researchers and professionals in this evolving field. The data collection process was structured around three central research questions: identifying the key challenges healthcare professionals face in RE, understanding their perceptions of how AI techniques can address these challenges, and investigating the practical barriers to AI adoption along with potential ways to overcome them. Together, these objectives and questions provided a comprehensive framework for connecting theoretical knowledge with practical experience.

5.2 Participant Selection

The target population for this study includes healthcare professionals involved in RE processes, such as doctors, nurses, other healthcare providers, hospital administrators, IT managers, software engineers, developers, and regulatory experts such as professionals with expertise in healthcare regulations like HIPAA and GDPR. A purposive sampling strategy was used to select participants with relevant expertise and experience. Inclusion criteria required participants to have at least five years of experience in healthcare or software development, direct involvement in RE processes for healthcare projects, and a willingness to participate in interviews or surveys. Participants without direct experience in healthcare RE or those unable to provide informed consent were excluded.

A total of 21 participants were selected, including 11 clinicians, 4 administrators, 4 developers, and 2 regulatory experts. The raw data from these interviews and surveys is accessible, with responses provided in reference [29]. This sample size was deemed sufficient to achieve data saturation. The diverse composition of the sample ensured a comprehensive understanding of the challenges and opportunities of using AI in healthcare RE from multiple perspectives.

5.3 Data Collection Instruments

The data collection process for this study combined semi-structured interviews and surveys to obtain both qualitative and quantitative insights from healthcare professionals. These activities were conducted individually rather than in groups, and on separate days based on each participant's availability. An interview guide was used to structure the sessions, covering topics such as participant background, challenges in healthcare RE, perceptions of AI techniques, barriers to AI adoption, and potential strategies for addressing these issues. To accommodate participants' availability and locations, interviews were conducted via video conferencing platforms such as Zoom and Microsoft Teams, with each session lasting approximately 25 to 35 minutes [29].

During these 21 video meetings, the survey component was also administered through screen sharing. Participants responded in real time, and their answers were recorded locally. A key innovation in the data collection process was the integration of AI tools specifically GPT-4-turbo [48] and Mistral Le Chat [49] to enhance the precision and clarity of participants' responses. As participants provided their answers verbally, these were immediately input into the AI models to help interpret and reformulate the responses into clearer and more structured statements. The interaction with these AI models was guided by structured prompts designed to ensure clarity and faithfulness to the participants' original intent. A typical prompt used was “Rephrase the following answer into a clear, formal, and structured requirement statement while preserving the speaker's intent”. In cases where ambiguity was suspected, an additional prompt was: “Identify and remove ambiguity from the following input while keeping the intended meaning intact.”

For example, when a clinician stated “I want the system to alert me if a patient’s condition gets worse at night,” this was entered into GPT-4-turbo using the rephrasing prompt. The model output “The system shall provide real-time alerts to clinicians when a patient’s vital signs indicate deterioration during nighttime hours.” This refined output was then shown to the participant during the session for confirmation. The participant approved the AI-enhanced version, stating that it accurately conveyed their intended meaning. This iterative and validated process not only improved the clarity and reliability of the data but also demonstrated how real-time integration of AI tools can support interpretation and refinement during empirical research involving complex, domain-specific input. By combining traditional data collection techniques with AI-assisted refinement, the study achieved a deeper and more accurate understanding of practitioner perspectives in the context of AI integration into healthcare RE [29].

5.4 Data Analysis

The interviews were analysed using thematic analysis. This involved familiarization with the data through repeated reading of the responses, generating initial codes to capture key concepts and patterns, grouping codes into themes based on their relevance to the research questions, reviewing and refining the themes to ensure coherence and consistency, and summarizing the findings in the context of the research questions. To enhance the trustworthiness of the thematic analysis, member validation was used. Refined themes and interpreted statements were shared with participants during or after the sessions for confirmation. This participant feedback ensured that the themes accurately reflected their intended meanings.

The survey data were analysed using descriptive and inferential statistics. Descriptive statistics, such as frequencies, means, and standard deviations, were calculated for Likert-scale questions to summarize the data. Inferential statistics, such as chi-square tests, were used to identify significant differences between participant groups. Open-ended responses were analysed using thematic analysis to identify common themes and patterns.

5.5 Results of Expert Interviews/Surveys

The expert interviews and surveys revealed several important findings that provide practical insights into the use of AI in healthcare requirements engineering. One recurring theme was the challenge of ensuring compliance with data protection regulations such as HIPAA and GDPR. Participants noted that aligning system requirements with these regulations is complex, particularly in settings where legal enforcement is inconsistent or unclear. For instance, one neurosurgeon explained that in Pakistan, data privacy laws are not strictly enforced, and patient data can sometimes be accessed by unauthorized parties, including law enforcement, without proper consent. Another frequently reported issue was ambiguity in requirements, especially when dealing with domain-specific medical terminology or when stakeholders struggle to articulate precise needs. This ambiguity often led to misunderstandings between development teams and clinical staff.

Data security concerns were also widely mentioned. Participants emphasized the critical importance of including clear and robust security requirements to prevent unauthorized access, breaches, or data misuse. In terms of AI applications, participants generally agreed that techniques such as NLP, ML, and LLMs hold promise for automating RE tasks and improving efficiency. Clinicians, for example, described how ML could assist in analysing medical images or predicting surgical outcomes, while LLMs could help in refining vague requirements into more actionable specifications.

However, several limitations were also highlighted. Participants noted that AI systems often struggle with domain-specific nuances and terminology, which can reduce their effectiveness. Trust was another concern where participants expressed skepticism about the reliability, transparency, and interpretability of AI-generated requirements, especially in safety-critical healthcare contexts. Ethical and legal risks, particularly around data privacy and algorithmic bias, were identified as additional barriers to adopting AI in healthcare RE. Participants were especially cautious about using AI tools that function as black boxes, where the reasoning behind decisions is not easily explainable to medical staff or patients. Further barriers included the lack of domain-specific AI models, the high cost of acquiring and maintaining AI tools, and the need for specialized training among healthcare professionals. To address these issues, participants suggested several strategies: developing and validating domain-specific AI models, offering training and capacity building for clinicians and developers, and encouraging collaboration between AI researchers and healthcare experts to ensure that tools are both technically sound and clinically relevant.

5.6 Discussion

The findings from the expert interviews and surveys complement and reinforce the insights gained from the systematic literature review. While AI techniques such as NLP, ML, and LLMs show clear potential to automate and enhance RE tasks particularly in the early stages of requirements elicitation and analysis their current limitations restrict widespread adoption in healthcare settings. Trust, explainability, and legal compliance emerged as major concerns that must be addressed for AI to be accepted by healthcare professionals. Participants repeatedly stressed the importance of developing AI models that are tailored to the specific needs, language, and regulatory context of healthcare. These findings highlight the pressing need for robust ethical frameworks that account for data privacy, bias, and patient consent, and for AI tools that can provide transparent and explainable reasoning. Moreover, the gap in interdisciplinary collaboration was a central theme. Many participants felt that current AI development processes lack sufficient involvement from medical professionals, which can result in tools that are technically impressive but practically limited. Addressing this gap requires building stronger, ongoing collaboration between software engineers, AI researchers, and clinical practitioners. Only through such collaboration can AI tools be properly aligned with the real-world requirements and constraints of healthcare practice. Ultimately, while the integration of AI into RE processes in healthcare holds promise, realizing its full potential will depend on careful attention to domain-specific challenges, user trust, legal compliance, and cross-disciplinary cooperation.

6. PROPOSED FRAMEWORK

Healthcare software systems are characterized by strict regulatory oversight, sensitive data management requirements, complex stakeholder ecosystems, and continuous innovation. While AI offers tremendous potential in streamlining RE, its application in healthcare RE introduces novel challenges that require domain-specific solutions. Based on insights derived from the SLR, expert interviews, and surveys, this section outlines a framework consisting of six interrelated strategies aimed at addressing the core challenges in applying AI to healthcare RE. The framework components are closely aligned with the primary results of the study and aim to provide both immediate practical solutions and a long-term vision for AI-supported RE in healthcare.

6.1 Develop Domain-Specific AI Models

One of the most consistent themes across expert responses and the SLR is the inadequacy of generic AI models in handling the specialized language and contexts of healthcare. For instance, NLP models trained on general-purpose corpora fail to capture the nuanced vocabulary of medical diagnostics, treatment protocols, and patient care standards. A practical solution is the development of AI models trained on healthcare-specific data including annotated clinical guidelines, historical requirements document from healthcare IT projects, and regulatory texts such as HIPAA and GDPR. By embedding healthcare ontology and domain knowledge into model training, AI tools can become significantly more accurate in tasks such as requirement elicitation, ambiguity detection, and classification. Domain-specific models also reduce the risk of misinterpretation, a concern repeatedly flagged in the empirical portion of this research.

6.2 Embed Ethical and Legal Constraints in AI Design

Healthcare is a highly regulated domain where missteps can have legal consequences and patient safety implications. AI systems used in RE must be designed with built-in ethical safeguards that align with regulatory expectations. For instance, requirements extracted from clinical documents using LLMs should be evaluated for compliance with privacy frameworks such as HIPAA in the US, GDPR in the EU. A strategy proposed during interviews includes incorporating automated regulatory compliance checklists within AI tools. These checklists can scan generated requirements for prohibited data-sharing practices or improper access protocols. Tools can also employ explainable AI

(XAI) techniques to provide traceability of decisions showing how a requirement was derived, what data it was based on, and what ethical filters were applied.

6.3 Promote Interdisciplinary Collaboration

AI specialists alone cannot capture the richness and constraints of healthcare systems. The framework emphasizes the formation of multidisciplinary RE teams that include clinicians, AI engineers, healthcare IT professionals, legal experts, and end-users. These collaborations are vital for ensuring AI-generated requirements are realistic, compliant, and clinically valid. Interviewees consistently emphasized that AI solutions in RE should not operate as black boxes. Instead, development should involve domain co-design, where healthcare experts review and provide feedback on model outputs during the training and deployment phases.

6.4 Establish Continuous Feedback Loops

Another critical strategy is the integration of real-time feedback mechanisms within AI tools. As observed in this study, tools like ChatGPT [48] or DeepSeek [50] were used interactively during interviews to clarify participant responses. This approach can be translated to practice by allowing stakeholders to annotate, edit, or approve AI-suggested requirements iteratively. For example, when an AI tool drafts a requirement such as "The system must ensure secure patient data access," stakeholders should be able to refine it into: "The system must implement role-based access controls compliant with ISO 27799, limiting access to authorized personnel only." This iterative process not only refines requirement quality but also builds stakeholder trust in AI-driven RE processes.

6.5 Facilitate Training and Capacity Building

The successful deployment of AI in RE hinges on practitioner competence. As revealed in survey responses, many healthcare professionals and RE specialists lack confidence in using AI tools effectively. A proactive approach involves formal training programs that educate stakeholders on both the potentials and limitations of AI in RE. Workshops, certifications, and simulation-based training can help bridge this gap. Furthermore, educational resources should be tailored to specific roles for instance, training modules for developers may focus on model integration, whereas those for clinicians may emphasize interaction with NLP powered elicitation tools.

6.6 Conduct Real-World Pilot Studies

Finally, the framework encourages empirical validation through pilot deployments. Much of the literature to date is theoretical or lab-based. Conducting pilot studies in actual healthcare environments such as hospitals or telemedicine platforms would provide critical insights into system usability, error rates, and overall effectiveness. These studies should be designed to assess not only technical metrics (precision, recall) but also human-centric outcomes such as stakeholder satisfaction, efficiency gains, and compliance accuracy. Results from such pilots can guide further AI tool refinement and inform healthcare organization's RE practices.

6.7 Framework Mapping and Practical Impact

To make the framework more actionable, Table 6-1 maps each of the six proposed strategies to the corresponding findings from the SLR and expert interviews, showing their practical use in AI-supported RE and outlining both short-term applications and long-term goals.

Table 6-1 Long term goals of proposed framework

Framework Strategy	Aligned Challenge from Findings	Practical Use in AI-Supported RE	Long-Term Vision
Develop Domain-Specific AI Models	Generic models fail with healthcare language and data	Improve elicitation and classification accuracy	Specialized, adaptive AI trained on real-time hospital data
Embed Ethical and Legal Constraints	Legal non-compliance, lack of transparency	Compliance checklists, explainable AI (XAI)	AI RE tools auto-updated with evolving regulations
Promote Interdisciplinary Collaboration	Disconnect between developers and clinicians	Co-design sessions and stakeholder reviews	Cross-functional AI-RE healthcare teams
Establish Continuous Feedback Loops	Misaligned and unclear requirements	Real-time stakeholder validation of AI outputs	Interactive AI assistants embedded in RE tools
Facilitate Training and Capacity Building	Lack of AI tool confidence and competence	Role-specific training modules, workshops	AI literacy embedded in medical and software education
Conduct Real-World Pilot Studies	Lack of empirical evaluation	Evaluate AI tool performance in real projects	Standardized benchmarks for AI-driven RE in healthcare

This framework provides a structured and actionable path forward for applying AI in healthcare RE. It connects directly with the practical challenges and insights reported in Chapters 4 and 5 and offers a foundation for both incremental improvements and strategic transformation.

6.8 Threats to Validity

Several limitations affect the validity and generalizability of this study. First, the systematic literature review may have missed relevant studies due to publication bias or limitations in the selected databases. To mitigate this, multiple databases and backward snowballing were used. However, some included works are preprints from *arXiv* such as SP3, SP24, SP25, and SP46 which have not undergone peer review. While these sources provide timely insights into fast-evolving areas like AI in requirements engineering, their findings should be interpreted with caution. Second, the expert interviews and surveys had a limited sample size (21 participants), which may not fully capture the diversity of perspectives in global healthcare settings. However, participants were drawn from diverse roles including clinicians, administrators, developers, and regulatory experts to ensure a broad viewpoint. Third, the use of AI tools like ChatGPT and Mistral Le Chat in real-time during data collection may have introduced interpretive bias. This was addressed through participant validation, where refined responses were reviewed and approved during sessions. Fourth, researcher bias in thematic analysis was reduced through independent code reviews and triangulation. Fifth, the framework itself is based on synthesized insights and has not yet been validated in practice. To address this, the final strategy recommends pilot testing in real healthcare settings. Lastly, although the six strategies were partially guided by literature and expert opinion, AI tools were also used to synthesize and refine the strategy descriptions. Their role was limited to support and not to independently generate content without human oversight.

7. FUTURE RESEARCH DIRECTIONS

Building on the foundations laid by this study, several avenues for future research emerge. First, empirical validation of the proposed framework in live healthcare projects is essential. Longitudinal case studies that measure key performance indicators such as requirement accuracy, elicitation time, stakeholder satisfaction, and compliance adherence will provide robust evidence of AI's impact and reveal unforeseen challenges in deployment. Second, the creation and open publication of annotated healthcare RE datasets would catalyze further innovation. Shared benchmarks would enable researchers to fine-tune domain-specific AI models and compare the performance of various approaches under standardized conditions. Third, research into hybrid human-AI collaboration models should be deepened, exploring how best to allocate tasks between human experts and automated tools, and how to design user interfaces that maximize transparency, trust, and efficiency. Fourth, the ethical and legal dimensions warrant deeper investigation, particularly in developing formal verification techniques that can guarantee AI-generated requirements comply with evolving regulations and ethical norms. This includes studying how explainable AI methods can be integrated into RE tools to provide clear audit trails for requirement provenance and decision rationale. Finally, interdisciplinary work is needed to address the rapid evolution of medical knowledge and technology. Future research could examine how continual learning paradigms where AI systems adapt to new guidelines, treatments, and device standards can be incorporated into RE pipelines without compromising system stability or regulatory compliance. By pursuing these directions, the field can move from exploratory pilot studies toward mature, field-tested AI solutions that reliably support the demanding requirements engineering needs of modern healthcare.

8. CONCLUSION

This study provides a multifaceted exploration of how Artificial Intelligence can be harnessed to enhance Requirements Engineering within the complex and highly regulated context of healthcare. Through a systematic literature review, expert interviews, and practitioner surveys, the research has illuminated both the promise and the pitfalls of deploying AI techniques particularly Natural Language Processing, Machine Learning, and Large Language Models in eliciting, analysing, and validating requirements for healthcare systems. One key contribution lies in the identification and categorization of domain-specific challenges, ranging from stringent compliance mandates under HIPAA and GDPR to the inherent ambiguity of clinical terminology and data security concerns. By mapping these challenges to AI capabilities and limitations, the thesis establishes a clear linkage between theoretical AI methods and the real-world needs of healthcare practitioners. Furthermore, the proposed framework offers concrete strategies for overcoming these challenges, including the development of specialized AI models trained on healthcare data, the embedding of ethical and legal constraints directly into AI workflows, interdisciplinary collaboration among clinicians, legal experts, and AI engineers. This framework not only bridges a critical gap in academic literature where most prior work has been predominantly conceptual or narrowly focused but also provides a roadmap for healthcare organizations and software developers to implement AI-driven RE processes that are both effective and compliant. The empirical insights gathered from real-world practitioners add significant practical value, validating the proposed strategies and highlighting the importance of iterative feedback loops and capacity-building initiatives to ensure successful AI adoption. Collectively, these contributions advance the discourse on AI in software engineering, demonstrating how intelligent automation can be tailored to serve the stringent demands of healthcare, ultimately improving the quality, safety, and compliance of health IT systems.

REFERENCES

[1] Zowghi, Didar, and Nur Nurmuliani. "A study of the impact of requirements volatility on software project performance." *Ninth Asia-Pacific Software Engineering Conference, 2002.*

[2] Bertl, Markus, Yngve Lamo, Martin Leucker, Tiziana Margaria, Esfandiar Mohammadi, Suresh Kumar Mukhiya, Ludwig Pechmann, Gunnar Piho, and Fazle Rabbi. "Challenges for AI in healthcare systems." In *International Conference on Bridging the Gap between AI and Reality*, pp. 165-186. Cham: Springer Nature Switzerland, 2023.

[3] Ferrari, Alessio, Felice Dell'Orletta, Andrea Esuli, Vincenzo Gervasi, and Stefania Gnesi. "Natural language requirements processing: a 4D vision." *IEEE SOFTWARE* 34, no. 6 (2017): 28-35.

[4] Ferrari, Alessio, Paola Spoletini, and Stefania Gnesi. "Ambiguity and tacit knowledge in requirements elicitation interviews." *Requirements Engineering* 21.3 (2016)1-20.

[5] Cheligeer C, Huang J, Wu G, Bhuiyan N, Xu Y, Zeng Y. Machine learning in requirements elicitation: a literature review. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*. 2022;36:e32. doi:10.1017/S0890060422000166

[6] Janssens, D.T.P.C.A.M. *Natural Language Processing in Requirements Elicitation and Requirements Analysis: A Systematic Literature Review*. Master's thesis, Department of Information and Computing Sciences, Utrecht University, 2019. https://studenttheses.uu.nl/handle/20.500.12932/32670.

[7] Necula, Sabina-Cristiana, Florin Dumitriu, and Valerică Greavu-Şerban. "A Systematic Literature Review on Using Natural Language Processing in Software Requirements Engineering." *Electronics* 13.11 (2024): 2055.

[8] Guest, Greg, Arwen Bunce, and Laura Johnson. "How many interviews are enough? An experiment with data saturation and variability." *Field Methods* 18.1 (2006): 59-82.

[9] Marques, Johnny, and Adilson Marques da Cunha. "Ares: An agile requirements specification process for regulated environments." *International Journal of Software Engineering and Knowledge Engineering* 29.10 (2019): 1403-1438.

[10] Mehraj, Ali, Zheying Zhang, and Kari Systä. "A tertiary study on AI for requirements engineering." *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Cham: Springer Nature Switzerland, 2024.

[11] Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. "From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices." *Science and Engineering Ethics* 26, no. 4 (2020): 2141-2168.

[12] Rajpurkar, Pranav, Emma Chen, Oishi Banerjee, and Eric J. Topol. "AI in health and medicine." *Nature medicine* 28, no. 1 (2022): 31-38.

[13] Faridoon, Amen, and M. Tahar Kechadi. "Healthcare data governance, privacy, and security-a conceptual framework." *EAI International Conference on Body Area Networks*. Cham: Springer Nature Switzerland, 2024.

[14] Lamsweerde, A. van. *Requirements engineering: from system goals to UML models to software specifications*. John Wiley & Sons, Ltd, 2009.

[15] Ramachandran, Muthu. "Software security requirements engineering: State of the art." *International Conference on Global Security, Safety, and Sustainability*. Cham: Springer International Publishing, 2015.

[16] Sharma, Malvika, Carl Savage, Monika Nair, Ingrid Larsson, Petra Svedberg, and Jens M. Nygren. "Artificial intelligence applications in health care practice: scoping review." *Journal of Medical Internet Research* 24, no. 10 (2022): e40238.

[17] Topol, Eric. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK, 2019.

[18] Zhang, Zheying, Maruf Rayhan, Tomas Herda, Manuel Goisauf, and Pekka Abrahamsson. "Llm-based agents for automating the enhancement of user story quality: An early report." In *International Conference on Agile Software Development*, pp. 117-126. Cham: Springer Nature Switzerland, 2024.

[19] Ruan, Kehan. "The Territorial Scope of GDPR-Conditions, Extraterritorial Application and Implications." (2024).

[20] Toivakka, Henrik. *Integration of EU medical device regulatory requirements into a CI/CD pipeline*. Diss. Master's thesis. Tampere University. URL: https://trepo. tuni. fi/bitstream/handle/10024/134314/ToivakkaHenrik. pdf, 2021.

[21] Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. EBSE Technical Report. Version 2.3.

[22] Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.). Wiley.

[23] Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the Craft of Qualitative Research Interviewing* (2nd ed.). Sage Publications.

[24] Braun, Virginia, and Victoria Clarke. "Using thematic analysis in psychology." *Qualitative Research in Psychology* 3.2 (2006): 77-101.

[25] U.S. Department of Health and Human Services. *The Standards for Privacy of Individually Identifiable Health Information*. Accessed March 10, 2025. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[26] European Union. *General Data Protection Regulation (GDPR)*. Accessed March 10, 2025. https://gdpr-info.eu/.

[27] European Parliament and Council of the European Union. *Regulation (EU) 2017/745 of 5 April 2017 on Medical Devices*. Accessed March 29, 2025. https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng.

[28] U.S. Food and Drug Administration. *Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products*. Accessed March 17, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/demonstrating-substantial-evidence-effectiveness-human-drug-and-biological-products.

[29] Ullah, Asmat. Expert interviews and surveys https://drive.google.com/drive/folders/15r99kAJkV6akLiGoooeY2n2jlNOiiVas?usp=sharing Conducted March 1, 2025.

[30] U.S. Congress. *Public Law 114–255, 114th Congress*. Accessed March 28, 2025. https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf.

[31] Government of Canada. *Personal Information Protection and Electronic Documents Act (S.C. 2000, c. 5)*. Accessed March 17, 2025. https://laws-lois.justice.gc.ca/eng/acts/p-8.6/.

[32] Government of Ontario. *Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched.* Accessed March 17, 2025. https://www.ontario.ca/laws/statute/04p03.

[33] International Organization for Standardization. *Health Informatics — Information Security Management in Health Using ISO/IEC 27002*. Accessed March 20, 2025. https://www.iso.org/standard/62777.html.

[34] E. Cambria and B. White, "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]," in *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, pp. 48-57, May 2014, doi: 10.1109/MCI.2014.2307227.

[35] Jurafsky, D., and Martin, J. H. 2023. *Speech and Language Processing*. Third Edition, Draft. Stanford University.

[36] Abualhaija, Sallam, Chetan Arora, Mehrdad Sabetzadeh, Lionel C. Briand, and Michael Traynor. "Automated demarcation of requirements in textual specifications: a machine learning-based approach." *Empirical Software Engineering* 25 (2020): 5454-5497.

[37] Heaton, Jeff. "Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618." *Genetic programming and evolvable machines* 19, no. 1 (2018): 305-307.

[38] Harmain, H. M., and Robert Gaizauskas. "Cm-builder: A natural language-based case tool for object-oriented analysis." *Automated Software Engineering* 10 (2003): 157-181.

[39] Mitchell, Tom M., and Tom M. Mitchell. *Machine learning*. Vol. 1. No. 9. New York: McGraw-hill, 1997.

[40] Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.

[41] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171-4186. 2019.

[42] Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. *Improving Language Understanding by Generative Pre-Training*. OpenAI, 2018. https://openai.com/research/language-unsupervised.

[43] Hou, Xinyi, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John Grundy, and Haoyu Wang. "Large Language Models for Software Engineering: A Systematic Literature Review." *ACM Transactions on Software Engineering and Methodology* 33, no. 8 (2024): 1–79. https://dl.acm.org/doi/10.1145/3695988.

[44] Haring, Malte, Juliane Schiller, Martin Gersch, and Volker Amelung. "Why should stakeholders consider the effect of tensions in collaborative innovation in healthcare—lessons learned from surveying integrated care projects in Germany." *BMC Health Services Research* 23, no. 1 (2023): 1292.

[45] Hamza, Muhammad, and Muhammad Azeem Akbar. "Smart Healthcare System Implementation Challenges: A stakeholder perspective." *arXiv preprint arXiv:2208.12641* (2022).

[46] Critical Appraisal Skills Programme (CASP). *CASP Checklists*. Accessed March 15, 2025. https://casp-uk.net/casp-tools-checklists/.

[47] Northern Illinois University, Center for Innovative Teaching and Learning. *Rubrics for Assessment*. Accessed March 10, 2025. https://www.niu.edu/citl/resources/guides/instructional-guide/rubrics-for-assessment.shtml.

[48] OpenAI. *ChatGPT: Optimizing Language Models for Dialogue*. 2023. https://openai.com/chatgpt. Accessed March 1, 2025.

[49] Mistral AI. *Mistral AI: Open-Weight Language Models*. https://mistral.ai. Accessed March 2, 2025.

[50] DeepSeek. *DeepSeek AI: Advancing Open-Source Large Language Models*. 2024. https://www.deepseek.com. Accessed March 1, 2025.

APPENDIX A: LIST OF PRIMARY STUDIES

SP1. Banerjee, Sri, Pat Dunn, Scott Conard, and Roger Ng. "Large language modeling and classical AI methods for the future of healthcare." Journal of Medicine, Surgery, and Public Health 1 (2023): 100026.

SP2. Awasthi, Raghav, Sai Prasad Ramachandran, Shreya Mishra, Dwarikanath Mahapatra, Hajra Arshad, Aarit Atreja, Anirban Bhattacharyya. "Artificial Intelligence in Healthcare: 2024 Year in Review." medRxiv (2025): 2025-02.

SP3. Wang, Junda, Zhichao Yang, Zonghai Yao, and Hong Yu. "JMLR: Joint medical LLM and retrieval training for enhancing reasoning and professional question answering capability." arXiv preprint arXiv:2402.17887 (2024).

SP4. Sathe, Neha, Vaibhav Deodhe, Yash Sharma, and Anand Shinde. "A comprehensive review of AI in healthcare: Exploring neural networks in medical imaging, LLM-based interactive response systems, NLP-based EHR systems, ethics, and beyond." In 2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech), pp. 633-640. IEEE, 2023.

SP5. Negri-Ribalta, Claudia, Marius Lombard-Platet, and Camille Salinesi. "Understanding the GDPR from a requirements engineering perspective a systematic mapping study on regulatory data protection requirements." Requirements Engineering 29, no. 4 (2024): 523-549.

SP6. Aberkane, Abdel-Jaouad, Geert Poels, and Seppe Vanden Broucke. "Exploring automated GDPR-compliance in requirements engineering: A systematic mapping study." IEEE Access 9 (2021): 66542-66559.

SP7. Elkourdi, Farah, Chenhao Wei, Lu Xiao, Zhongyuan Yu, and Onur Asan. "Exploring Current Practices and Challenges of HIPAA Compliance in Software Engineering: Scoping Review." IEEE Open Journal of Systems Engineering (2024).

SP8. Alsaadi, Mohmood, Alexei Lisitsa, and Malik Qasaimeh. "Minimizing the ambiguities in medical devices regulations based on software requirement engineering techniques." In Proceedings of the second international conference on data science, e-learning and information systems, pp. 1-5. 2019.

SP9. Riad, ABM Kamrul Islam, Md Abdul Barek, Md Mostafizur Rahman, Mst Shapna Akter, Tahia Islam, Md Abdur Rahman, Md Raihan Mia, Hossain Shahriar, Fan Wu, and Sheikh Iqbal Ahamed. "Enhancing HIPAA Compliance in AI-driven mHealth Devices

Security and Privacy." In 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 2430-2435. IEEE, 2024.

SP10. Ermakova, Tatiana, Benjamin Fabian, Marta Kornacka, Scott Thiebes, and Ali Sunyaev. "Security and privacy requirements for cloud computing in healthcare: Elicitation and prioritization from a patient perspective." ACM Transactions on Management Information Systems (TMIS) 11, no. 2 (2020): 1-29.

SP11. Davenport, Thomas, and Ravi Kalakota. "The potential for artificial intelligence in healthcare." Future healthcare journal 6, no. 2 (2019): 94-98.

SP12. Lu, Zhiyong, Yifan Peng, Trevor Cohen, Marzyeh Ghassemi, Chunhua Weng, and Shubo Tian. "Large language models in biomedicine and health: current research landscape and future directions." Journal of the American Medical Informatics Association 31, no. 9 (2024): 1801-1811.

SP13. Ong, Jasmine Chiat Ling, Shelley Yin-Hsi Chang, Wasswa William, Atul J. Butte, Nigam H. Shah, Lita Sui Tjien Chew, Nan Liu. "Medical ethics of large language models in medicine." NEJM AI 1, no. 7 (2024): AIra2400038.

SP14. Liu, Qiming, Ruirong Yang, Qin Gao, Tengxiao Liang, Xiuyuan Wang, Shiju Li, Bingyin Lei, and Kaiye Gao. "A Review of Applying Large Language Models in Healthcare." IEEE Access (2024).

SP15. Bojesomo, Alabi, Mohamed Seghier, Leontios Hadjileontiadis, and Aamna AlShehhi. "Revolutionizing Disease Diagnosis with Large Language Models: A Systematic Review." *Research Square*, December 27, 2024. https://doi.org/10.21203/rs.3.rs-5704278/v1.

SP16. Ananthanarayanan, Aniruth. "Generating Medical Diagnostic Scenarios with LLM-Based Reinforcement Learning Feedback: Dataset Release and Methodology." In IEEE Integrated STEM Education Conference. 2025.

SP17. Nassiri, Khalid, and Moulay A. Akhloufi. "Recent advances in large language models for healthcare." BioMedInformatics 4, no. 2 (2024): 1097-1143.

SP18. Sarker, Abeed, Rui Zhang, Yanshan Wang, Yunyu Xiao, Sudeshna Das, Dalton Schutte, David Oniani, Qianqian Xie, and Hua Xu. "Natural Language Processing for Digital Health in the Era of Large Language Models." Yearbook of Medical Informatics 33, no. 01 (2024): 229-240.

SP19. Kaur, Kamaljit, and Parminder Kaur. "The application of AI techniques in requirements classification: a systematic mapping." Artificial Intelligence Review 57, no. 3 (2024): 57.

SP20. Luschi, Alessio. "Designing and Developing a Dedicated Natural Language Processing Framework for Healthcare Information Technology Management and Assessment." (2024). https://hdl.handle.net/2158/1353284

SP21. Nazi, Zabir Al, and Wei Peng. "Large language models in healthcare and medical domain: A review." Informatics, vol. 11, no. 3, p. 57. MDPI, 2024.

SP22. Khurshid, Iqra, Salma Imtiaz, Wadii Boulila, Zahid Khan, Almas Abbasi, Abdul Rehman Javed, and Zunera Jalil. "Classification of non-functional requirements from iot oriented healthcare requirement document." Frontiers in Public Health 10 (2022): 860536.

SP23. Matheny, Michael E., Jie Yang, Joshua C. Smith, Colin G. Walsh, Mohammed A. Al-Garadi, Sharon E. Davis, Keith A. Marsolo. "Enhancing postmarketing surveillance of medical products with large language models." JAMA Network Open 7, no. 8 (2024): e2428276-e2428276.

SP24. Wang, Zifeng, Hanyin Wang, Benjamin Danek, Ying Li, Christina Mack, Hoifung Poon, Yajuan Wang, Pranav Rajpurkar, and Jimeng Sun. "A perspective for adapting generalist AI to specialized medical ai applications and their challenges." arXiv preprint arXiv:2411.00024 (2024).

SP25. Garg, Muskan, Shaina Raza, Shebuti Rayana, Xingyi Liu, and Sunghwan Sohn. "The Rise of Small Language Models in Healthcare: A Comprehensive Survey." arXiv preprint arXiv:2504.17119 (2025).

SP26. Mersha, Melkamu, Khang Lam, Joseph Wood, Ali AlShami, and Jugal Kalita. "Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction." Neurocomputing (2024): 128111.

SP27. Oniani, David, Jordan Hilsman, Yifan Peng, Ronald K. Poropatich, Jeremy C. Pamplin, Gary L. Legault, and Yanshan Wang. "Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare." NPJ Digital Medicine 6, no. 1 (2023): 225.

SP29. Shen, Yiqiu, Laura Heacock, Jonathan Elias, Keith D. Hentel, Beatriu Reig, George Shih, and Linda Moy. "ChatGPT and other large language models are double-edged swords." Radiology 307, no. 2 (2023): e230163.

SP30. Friha, Othmane, Mohamed Amine Ferrag, Burak Kantarci, Burak Cakmak, Arda Ozgun, and Nassira Ghoualmi-Zine. "Llm-based edge intelligence: A comprehensive survey on architectures, applications, security and trustworthiness." IEEE Open Journal of the Communications Society (2024).

SP31. Jelassi, Mariem, Khouloud Matteli, Houssem Ben Khalfallah, and Jacques Demongeot. 2024. "Enhancing Personalized Mental Health Support Through Artificial Intelligence: Advances in Speech and Text Analysis Within Online Therapy Platforms" Information 15, no. 12: 813. https://doi.org/10.3390/info15120813.

SP32. Pesqueira, Antonio, Diana Gameiro Lopes, Mark Schwendinger, and Shipra Gupta. "Synergistic AI: Big Data Insights for Enhanced Biomedical Discovery." In Navigating Privacy, Innovation, and Patient Empowerment Through Ethical Healthcare Technology, pp. 139-170. IGI Global Scientific Publishing, 2025.

SP33. Birhane, Abeba. "The impossibility of automating ambiguity." Artificial Life 27, no. 1 (2021): 44-61.

SP34. Sohail, Shahab Saquib, Faiza Farhat, Yassine Himeur, Mohammad Nadeem, Dag Øivind Madsen, Yashbir Singh, Shadi Atalla, and Wathiq Mansoor. "The future of GPT: A taxonomy of existing ChatGPT research, current challenges, and possible future directions." Current Challenges, and Possible Future Directions (April 8, 2023) (2023).

SP35. Stawicki, Stanislaw & Barrera, Rafael & Christopherson, Nathan & Narayan, Mayur & Izurieta, Ricardo & Galwankar, Sagar & Moreno-Walton, Lisa & Wojda, Thomas & Patel, Vihas. (2024). The Ninth Annual Academic International Medicine Virtual Congress and Scientific Forum, June 20–22, 2024: Meeting Report and Scientific Abstracts. International Journal of Academic Medicine and Pharmacy. 10. 375-391. 10.4103/ijam.ijam_183_24.

SP36. Ennis, Edel, Patrick McAllister, and Colin Gorman, eds. *DMHW Conference Proceedings 2024*. Ulster University, October 2024. https://doi.org/10.21251/dee52a42-262c-4bb6-b81a-e6d8ddcdbcd5.

SP37. Necula, Sabina-Cristiana, Florin Dumitriu, and Valerică Greavu-Şerban. "A Systematic Literature Review on Using Natural Language Processing in Software Requirements Engineering." Electronics 13.11 (2024): 2055.

SP38. Grundy, John, Anuradha Madugalla, Jennifer McIntosh, and Truyen Tran. "Vision: Requirements engineering for software development in aged care." In 2023 IEEE 31st International Requirements Engineering Conference Workshops (REW), pp. 440-445. IEEE, 2023.

SP39. Srivastava, Nidhi, Sneha Verma, Anupama Singh, Pranki Shukla, Yashvardhan Singh, Ankit D. Oza, Tanvir Kaur, Sohini Chowdhury, Monit Kapoor, and Ajar Nath Yadav. "Advances in artificial intelligence-based technologies for increasing the quality of medical products." DARU Journal of Pharmaceutical Sciences 33, no. 1 (2024): 1.

SP40. Amann, Julia, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I. Madai, and Precise4Q Consortium. "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective." BMC medical informatics and decision making 20 (2020): 1-9.

SP41. Sarkar, Nilanjana, and Sumit Goel. "Artificial intelligence and machine learning in healthcare." In Biomedical Research Developments for Improved Healthcare, pp. 12-46. IGI Global, 2024.

SP42. Chen, Xieling, Haoran Xie, Xiaohui Tao, Fu Lee Wang, Mingming Leng, and Baiying Lei. "Artificial intelligence and multimodal data fusion for smart healthcare: topic modeling and bibliometrics." Artificial Intelligence Review 57, no. 4 (2024): 91.

SP43. Al-Remawi, Mayyas, Ahmed SA Ali Agha, Faisal Al-Akayleh, Faisal Aburub, and Rami A. Abdel-Rahem. "Artificial intelligence and machine learning techniques for suicide prediction: Integrating dietary patterns and environmental contaminants." Heliyon (2024).

SP44. Lin, Han, Jiatong Han, Pingping Wu, Jiangyan Wang, Juan Tu, Hao Tang, and Liuning Zhu. "Machine learning and human-machine trust in healthcare: A systematic survey." CAAI Transactions on Intelligence Technology 9, no. 2 (2024): 286-302.

SP45. Zidaru, Teodor, Elizabeth M. Morrow, and Rich Stockley. "Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice." Health Expectations 24, no. 4 (2021): 1072-1124.

SP46. Atf, Zahra, Seyed Amir Ahmad Safavi-Naini, Peter R. Lewis, Aref Mahjoubfar, Nariman Naderi, Thomas R. Savage, and Ali Soroush. "The challenge of uncertainty quantification of large language models in medicine." arXiv preprint arXiv:2504.05278 (2025).

SP47. Karatas, Mumtaz, Levent Eriskin, Muhammet Deveci, Dragan Pamucar, and Harish Garg. "Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives." Expert Systems with Applications 200 (2022): 116912.

SP48. Topol, Eric. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. First edition. New York: Basic Books, 2019. https://books.google.fi/books?id=_EFIDwAAQBAJ.

SP49. Verma, Ashish, Akanksha Sharma, Aditi Singh, Rishabha Malviya, and Shivkanya Fuloria. "Digital Assistant in the Pharmaceutical Field for Advancing Healthcare Systems." In Pharmaceutical industry 4.0: Future, Challenges & Application, pp. 213-254. River Publishers, 2023.

SP50. Mejía-Granda, Carlos M., José L. Fernández-Alemán, Juan M. Carrillo-de-Gea, and José A. García-Berná. "Security vulnerabilities in healthcare: an analysis of medical devices and software." Medical & Biological Engineering & Computing 62, no. 1 (2024): 257-273.

SP52. Hassan, Shoaib, Qianmu Li, Ayed Alwadain, Affan Yasin, and Javed Ali Khan. "Integrating Emotional Requirements in Use-Case Point Model for Effort Estimation of Healthcare System." IEEE Access (2024).

SP53. Izhar, Rahat, Shahid N. Bhatti, and Sultan A. Alharthi. "Bridging Precision and Complexity: A Novel Machine Learning Approach for Ambiguity Detection in Software Requirements." IEEE Access (2025).

SP54. Xu, Lina, Monika Simjanoska, Bojana Koteska, Vladimir Trajkovikj, Ana Madevska Bogdanova, Kristina Drusany Starič, and Fedor Lehocki. "What clinics are expecting from data scientists? A review on data oriented studies through qualitative and quantitative approaches." IEEE Access 7 (2018): 641-654.

SP55. Ashok, Kanneboina, and S. Gopikrishnan. "Statistical analysis of remote health monitoring based IoT security models & deployments from a pragmatic perspective." IEEE Access 11 (2023): 2621-2651.

SP56. Cai, Carrie J., Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. "" Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making." Proceedings of the ACM on Human-Computer Interaction 3, no. CSCW (2019): 1-24.

SP57. Prakash, Ashish Viswanath, and Saini Das. "Intelligent conversational agents in mental healthcare services: A thematic analysis of user perceptions." Pacific Asia Journal of the Association for Information Systems 12.2 (2020): 1.

SP58. Antioch, Kathryn Margaret, Angeline Kuek, John Pilla, and Rick Marshall. "A multi-criteria decision analysis framework that prioritizes economic policy to enhance value based health care during the COVID-19 pandemic." Discover Health Systems 3, no. 1 (2024): 12.

SP59. Wu, Jian, Guangyin Zhang, Yumei Xing, Yujia Liu, Zhen Zhang, Yucheng Dong, and Enrique Herrera-Viedma. "A sentiment analysis driven method based on public and personal preferences with correlated attributes to select online doctors." Applied Intelligence 53, no. 16 (2023): 19093-19114.

SP60. Raman, Raghu, Laurie Hughes, Santanu Mandal, Payel Das, and Prema Nedungadi. "Mapping metaverse research to the sustainable development goal of good health and well-being." IEEE Access (2024).

SP61. Ahmad, Khlood, Mohamed Abdelrazek, Chetan Arora, Muneera Bano, and John Grundy. "Requirements engineering for artificial intelligence systems: A systematic mapping study." Information and Software Technology 158 (2023): 107176.

SP62. Saraswat, Deepti, Pronaya Bhattacharya, Ashwin Verma, Vivek Kumar Prasad, Sudeep Tanwar, Gulshan Sharma, Pitshou N. Bokoro, and Ravi Sharma. "Explainable AI for healthcare 5.0: opportunities and challenges." IEEE Access 10 (2022): 84486-84517.

SP63. Ansari, Md Tarique Jamal, Fahad Ahmed Al-Zahrani, Dhirendra Pandey, and Alka Agrawal. "A fuzzy TOPSIS based analysis toward selection of effective security requirements engineering approach for trustworthy healthcare software development." BMC Medical Informatics and Decision Making 20 (2020): 1-13.

SP64. Nalchigar, Soroosh, Eric Yu, and Karim Keshavjee. "Modeling machine learning requirements from three perspectives: a case report from the healthcare domain." Requirements Engineering 26 (2021): 237-254.

SP65. Heyn, Hans-Martin, Eric Knauss, Amna Pir Muhammad, Olof Eriksson, Jennifer Linder, Padmini Subbiah, Shameer Kumar Pradhan, and Sagar Tungal. "Requirement engineering challenges for ai-intense systems development." In 2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN), pp. 89-96. IEEE, 2021.

SP66. Belani, Hrvoje, Marin Vukovic, and Željka Car. "Requirements engineering challenges in building AI-based complex systems." 2019 IEEE 27th international requirements engineering conference workshops (REW). IEEE, 2019.

SP67. Vogelsang, Andreas, and Markus Borg. "Requirements engineering for machine learning: Perspectives from data scientists." 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW). IEEE, 2019.

SP68. Khan, Z. Faizal, and Sultan Refa Alotaibi. "Applications of artificial intelligence and big data analytics in m-health: A healthcare system perspective." Journal of Healthcare Engineering 2020.1 (2020): 8894694.

SP69. Oami, Takehiko, Yohei Okada, and Taka-aki Nakada. "Performance of a Large Language Model in screening citations." JAMA network open 7.7 (2024): e2420496-e2420496.

SP70. Siddeshwar, Vaishali, Sanaa Alwidian, and Masoud Makrehchi. "A Systematic Review of AI-Enabled Frameworks in Requirements Elicitation." IEEE Access (2024).

SP71. Wadhwa, Shruti, and Karuna Babber. "Artificial intelligence in health care: predictive analysis on diabetes using machine learning algorithms." Computational Science and Its Applications–ICCSA 2020: 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part II 20. Springer International Publishing, 2020.

SP72. Alowais, Shuroug A., Sahar S. Alghamdi, Nada Alsuhebany, Tariq Alqahtani, Abdulrahman I. Alshaya, Sumaya N. Almohareb, Atheer Aldairem. "Revolutionizing healthcare: the role of artificial intelligence in clinical practice." BMC Medical Education 23, no. 1 (2023): 689.

SP73. Petersson, Lena, Ingrid Larsson, Jens M. Nygren, Per Nilsen, Margit Neher, Julie E. Reed, Daniel Tyskbo, and Petra Svedberg. "Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden." BMC Health Services Research 22, no. 1 (2022): 850.

SP74. Pesapane, Filippo, Caterina Volonté, Marina Codari, and Francesco Sardanelli. "Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States." Insights into imaging 9 (2018): 745-753.

SP75. Lunarejo, María Isabel Limaylla. "Requirements prioritization based on multiple criteria using Artificial Intelligence techniques." 2021 IEEE 29th International Requirements Engineering Conference (RE). IEEE, 2021.