UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Computer Science Curriculum

Konstantin Tenman

AI-Powered MMSE: Enhancing Cognitive Assessment

Master's Thesis (30 ECTS)

Supervisor: Mohamad Gharib, PhD

Tartu 2024

AI-Powered MMSE: Enhancing Cognitive Assessment

Abstract:

The increasing incidence of dementia among the elderly highlights the critical demand for cognitive assessment tools that are both efficient and widely accessible. Traditional methods, such as the Mini-Mental State Examination (MMSE), are typically conducted in clinical settings using paper-based formats, limiting access due to resource constraints and needing trained professionals. This thesis addresses these challenges by converting the MMSE into an AI-powered, web-based application, allowing assessments to be completed at home with minimal non-professional assistance.

The digital implementation leverages sophisticated artificial intelligence (AI) models, particularly the Llama 3.1:70B, for automating the administration of the MMSE. This makes it more consistent and sensitive to small changes in cognitive function. By leveraging Machine Learning (ML) and Natural Language Processing (NLP), the system improves the consistency, accuracy, and accessibility of cognitive assessments through web-based administration.

Adopting the Design Science Research (DSR) framework, this study incorporates contemporary web technologies alongside a hybrid AI strategy, enhancing performance while safeguarding data privacy. In trials, the AI-powered MMSE achieved a 92.9% success rate in confirming response correctness compared to traditional methods and slightly higher user satisfaction despite longer administration times.

While this work significantly improves cognitive assessment accessibility and sensitivity, further studies are needed to validate its effectiveness across diverse clinical settings. Future research should optimize response times, expand language support, and address ethical considerations in AI-driven cognitive assessments.

Keywords: cognitive assessment, dementia, web-based MMSE, artificial intelligence, machine learning, natural language processing, accessibility, healthcare democratization

CERCS: P170 - Computer science, numerical analysis, systems, control

Tehisintellektil põhinev MMSE: Kognitiivse hindamise täiustamine

Lühikokkuvõte:

Eakate seas dementsuse esinemissageduse kasv rõhutab tõhusate ja laialdaselt kättesaadavate kognitiivse hindamise tööriistade vajadust. Traditsioonilised meetodid, nagu Mini-Mental State Examination (MMSE), viiakse tavaliselt läbi kliinilistes tingimustes, kasutades paberkandjal vorme, mis piirab nende kättesaadavust ressursside nappuse ja koolitatud spetsialistide vajaduse tõttu. Käesolev lõputöö tegeleb nende väljakutsetega, muutes MMSE tehisintellektipõhiseks veebirakenduseks, mis võimaldab hinnangu andmist kodustes tingimustes minimaalse mitteprofessionaalse abiga.

Magistritöö käigus loodud digitaalne prototüüp kasutab keerukaid tehisintellekti (AI) mudeleid, nagu Llama 3.1 70B, et MMSE läbiviimist automatiseerida. See muudab hindamise järjepidevamaks ja tundlikumaks väikeste kognitiivsete funktsioonide muutuste suhtes. Masinõppe (ML) ja loomuliku keele töötlemise (NLP) abil parandab süsteem kognitiivsete hinnangute järjepidevust, täpsust ja kättesaadavust veebipõhise läbiviimise kaudu.

Uurimistöös on kasutatud disainiteaduse uurimismeetodit (ingl. Design Science Research, DSR), mis ühendab kaasaegsed veebitehnoloogiad ja hübriidse AI-strateegia, et parandada jõudlust ja kaitsta andmete privaatsust. Katsetes saavutas AI-põhine MMSE 92.9% täpsuse vastuste õigsuse tuvastamisel võrreldes traditsiooniliste meetoditega. Kuigi läbiviimise aeg oli veidi pikem, oli kasutajate rahulolu siiski pisut suurem.

See töö annab märkimisväärse panuse kognitiivse hindamise kättesaadavuse ja tundlikkuse parandamisse, kuid edasised uuringud on vajalikud, et kinnitada selle efektiivsust erinevates kliinilistes tingimustes. Tulevased uuringud peaksid keskenduma hindamiskiiruse optimeerimisele, keelelise toe laiendamisele ning AI-põhiste kognitiivsete mõõtmiste eetiliste küsimuste käsitlemisele.

Võtmesõnad: kognitiivne hindamine, dementsus, veebipõhine MMSE, tehisintellekt, masinõpe, loomuliku keele töötlus, kättesaadavus, tervishoiuteenuste demokratiseerimine

CERCS: P170 - Arvutiteadus, arvanalüüs, süsteemid, kontroll

Contents

1 Introduction 6
- 1.1 The Aging Population and Cognitive Health . . . 6
- 1.2 The Mini-Mental State Examination (MMSE) . . . 7
- 1.3 Research Objectives . . . 8
- 1.4 Thesis Structure . . . 9

2 Baseline and Literature Review 11
- 2.1 The MMSE . . . 11
  - 2.1.1 MMSE Selection Rationale . . . 11
  - 2.1.2 MMSE Scoring and Significance . . . 12
  - 2.1.3 Limitations and Criticisms . . . 12
- 2.2 Digital Cognitive Assessment . . . 13
  - 2.2.1 Advantages . . . 13
  - 2.2.2 Trends . . . 14
  - 2.2.3 Challenges . . . 15
- 2.3 AI in Assessment . . . 16
  - 2.3.1 AI Applications and Ethics . . . 16
  - 2.3.2 AI Techniques for Cognitive Assessment . . . 17
- 2.4 Ethical Considerations . . . 18
  - 2.4.1 Cultural Adaptation Methods . . . 19
  - 2.4.2 Web-Based Assessment Fairness . . . 19
- 2.5 Current Technology Gaps . . . 20
- 2.6 Summary of Current State . . . 21

3 Research Methodology 22
- 3.1 Problem Identification . . . 23
  - 3.1.1 MMSE Limitations . . . 24
  - 3.1.2 Needs Assessment . . . 24
- 3.2 Artifact Design and Development Process . . . 25
  - 3.2.1 Conceptualizing the MMSE and System Design . . . 25
  - 3.2.2 Challenges and Solutions in Automating MMSE Tasks . . . 25
- 3.3 Implementation and Development . . . 27
  - 3.3.1 Database and Performance . . . 31
  - 3.3.2 Assessment Process and AI Integration . . . 34
- 3.4 Evaluation Approach . . . 36
  - 3.4.1 Performance Indicators . . . 36
  - 3.4.2 Expert Evaluation Process . . . 36
  - 3.4.3 User Experience Assessment . . . 37
  - 3.4.4 Data Analysis . . . 37

4 Implementation 38
- 4.1 Architecture Diagram . . . 38
- 4.2 AI Integration . . . 39
  - 4.2.1 AI Models and Roles . . . 39

    - 4.2.2 Integration Process . . . 40
    - 4.2.3 Advantages of the Hybrid Approach . . . 41
    - 4.2.4 Future Extensibility . . . 41
- 4.3 Design Principles . . . 42
    - 4.3.1 Microservices and Python Services . . . 42
    - 4.3.2 Architectural Design of RESTful APIs . . . 42
- 4.4 Backend Implementation . . . 42
    - 4.4.1 Overall Structure . . . 43
    - 4.4.2 Domain Model and Services . . . 43
    - 4.4.3 Question Interface . . . 45
    - 4.4.4 REST API and AI Integration . . . 46
- 4.5 Testing Practices . . . 48
- 4.6 Frontend Implementation . . . 49
    - 4.6.1 Component Structure Overview . . . 49
    - 4.6.2 Interaction with Backend Logic . . . 49
    - 4.6.3 User Experience and Interface Design . . . 49
- 4.7 Data Management . . . 49

5 Results and Evaluations 51
- 5.1 Data Collection Overview . . . 51
- 5.2 Evaluation of Different AI Models . . . 51
    - 5.2.1 Methodology . . . 51
    - 5.2.2 Results and Analysis . . . 52
    - 5.2.3 Model Selection . . . 52
- 5.3 Comparison: Traditional vs Web-based MMSE . . . 53
- 5.4 Novelty of the Study . . . 54
- 5.5 Evaluation and Implications . . . 55
    - 5.5.1 Analysis of Results . . . 55
    - 5.5.2 Research Contributions . . . 55
    - 5.5.3 Study Limitations . . . 56
- 5.6 Future Directions . . . 56

6 Conclusion 58

References 67

Appendix 68
- I. Repository . . . 68
- II. Licence . . . 68
- III. Traditional Mini-Mental State Examination (MMSE) . . . 69
- IV. System Architecture Overview . . . 70
- V. Participant Consent Form . . . 71
- VI. Writing Workflow . . . 72

1 Introduction

Europe is experiencing a demographic shift, with projections indicating that by 2050, 29.4% of the population will be 65 years or older, contributing to a rise in age-related conditions such as dementia [1, 2]. This shift places a significant burden on healthcare systems, requiring the development of innovative approaches to support the aging population [3]. Early detection and intervention in cognitive decline are crucial, as they can substantially affect disease progression and improve quality of life [4].

However, traditional cognitive assessment tools like the Mini-Mental State Examination (MMSE) face limitations due to cultural biases and the need for in-person administration. Recent advancements in Artificial Intelligence (AI) and tangible interaction technologies offer promising solutions to these challenges, potentially enhancing the accessibility and effectiveness of cognitive assessments [2].

Given these trends, it is essential to address the limitations of traditional methods such as the MMSE, which is widely used for detecting mild cognitive impairments [5]. The current study proposes developing a web-based version of the MMSE to enhance accessibility, efficiency, and reach, especially in underserved regions. The digital platform could reduce biases, improve consistency in administration, and allow remote assessments, offering potential benefits in broader population screening and long-term monitoring of cognitive health [6–8].

1.1 The Aging Population and Cognitive Health

Europe's demographic landscape is undergoing a significant transformation due to its rapidly aging population, a change that correlates with an increase in age-related cognitive issues, most notably:

1. **Mild Cognitive Impairment (MCI):**
    - *Definition:* A condition involving slight but noticeable decline in cognitive abilities.
    - *Prevalence:* Affects 6–12% of adults aged 60 and older [9].
2. **Dementia:**
    - *Definition:* A complex neurological condition characterized by a substantial deterioration in cognitive abilities, impacting various mental functions including memory, reasoning, and communication skills [10].

These trends highlight the growing importance of cognitive health in the aging population of Europe, emphasizing the need for effective screening and intervention strategies.

The increasing prevalence of cognitive impairment poses significant challenges to healthcare systems, families, and societies. However, early detection and targeted interventions can significantly alter the course of decline. Prompt identification allows healthcare providers to implement strategies that may slow progression, improve quality of life, and alleviate overall stress on healthcare infrastructure [11]. This approach creates an opportunity for timely diagnosis, empowering patients and their families to map out future care strategies proactively. It enables them to make informed decisions about the diverse treatment options available through various channels [12].

Considering these views, an emerging need demands efficient, accessible, and trustworthy cognitive assessment tools, enabling mass-scale screening and monitoring of the aging population based on cognitive function. This research necessitates the development of an improved web-based Mini-Mental State Examination application to enhance accessibility and efficiency in cognitive assessment within clinical and research settings.

1.2 The Mini-Mental State Examination (MMSE)

Cognitive assessment is crucial in detecting and managing cognitive impairments, especially in the elderly. These kinds of tests can check for different kinds of cognitive disorders and give an idea of how bad certain impairments are, so that doctors can see how the disease is getting worse and see what treatments are working. These tools make early detection possible and initiate timely intervention, stopping deterioration and improving quality of life, in addition to providing essential information for studies on cognitive aging [9, 13, 14]. Even though it is relevant, the administration of cognitive assessment has faced a variety of challenges: variability across settings, time-consuming paper-and-pencil methods, cultural and linguistic biases, and remote assessment-related complications [15–20].

Folstein et al. [5] developed the MMSE in 1975, becoming a standard cognitive screening tool in clinical practice and research. Test items cover various cognitive domains, such as orientation, attention, memory, language, and visuospatial skills of a subject [5, 21, 22]. In most cases, clinicians perform the test face-to-face with individuals asking them to complete tasks ranging from a subject's cognitive orientation to visual construction. The advantages of MMSE are its ease of administration and its high sensitivity toward moderate to severe levels of cognitive impairments. Nevertheless, it is known to be relatively insensitive to mild impairments, to have cultural biases, and is limited in measuring executive functions [5].

The need to overcome traditional assessment limitations and address challenges like the COVID-19 pandemic, which highlighted the need for remote healthcare solutions, drives the move towards a web-based MMSE. A web-based MMSE can increase accessibility, reduce administration time, and enable remote patient monitoring, enhancing the ability to track changes over time [6–8, 19, 23–25]. Modern cognitive evaluations have embraced technologies, such as AI, in personalization and innovating test procedures that capture subtle changes in cognitive function. However, challenges accompany these digital versions of MMSE, including the still-existing digital divide among older adults, concerns about data security, and validation against traditional evaluation methods [20, 26]. Discussing these challenges remains open to building a robust and widely accepted web-based cognitive assessment tool.

The Novelty of the Study: This research introduces a novel approach to cognitive assessments by employing state-of-the-art AI techniques with Large Language Models (LLMs). Specifically, the study utilizes the Llama 3.1:70B model, executed on the Ollama platform, in conjunction with a vision model for tasks requiring visual processing. Additionally, fallback mechanisms from ChatGPT 4o are integrated to ensure robustness across diverse assessment scenarios. This advanced suite of technologies enhances cognitive assessments' adaptability, accuracy, and interpretability, setting a new standard in the field. Adding these advanced AI technologies to the web-based MMSE is meant to improve traditional ways of testing cognitive abilities. This will push the limits of automation and ensure the test works well in a wide range of cognitive and system conditions.

1.3 Research Objectives

This thesis aims to design and critically evaluate a web-based adaptation of the Mini-Mental State Examination (MMSE), addressing the limitations of traditional cognitive assessment methods by leveraging the benefits offered by modern digital technologies. The research focuses on enhancing accessibility, efficiency, and accuracy in cognitive assessments, particularly for remote or underserved populations.

The main research objective is to develop and assess a web-based adaptation of the MMSE, validating its properties against traditional cognitive assessment methods while harnessing the advantages of digital technology.

To achieve this, the web-based MMSE will incorporate the following key features and innovations:

- **Adaptive Testing:** Utilize item response theory [27] to dynamically adjust question difficulty based on user responses, enhancing sensitivity to subtle cognitive changes.
- **Multimodal Input:** Implement various input methods including touchscreen, voice recognition, and microphone interactions to accommodate diverse user capabilities.
- **Automated Scoring and Analysis:** Employ machine learning algorithms [28] to automate scoring processes and provide detailed cognitive performance analyses.

- **Remote Monitoring:** Develop capabilities for longitudinal tracking and remote monitoring to detect changes in cognitive function over time.

These features and innovations will guide the development and evaluation of the web-based MMSE, aiming to enhance the accessibility, efficiency, and effectiveness of cognitive assessments through innovative technology.

1.4 Thesis Structure

The structure takes the reader through the research process, the findings related to the web-based MMSE system, and what this tool implies. It has six main chapters related to some points about the research and what it adds regarding cognitive assessment.

Chapter 2: Baseline and Literature Review. This chapter comprehensively discusses and reviews existing literature on cognitive assessments, mainly focusing on the MMSE and its adaptations. It also explores the evolution of computerized cognitive testing and discusses the role of artificial intelligence in enhancing neuropsychological assessment.

Chapter 3: Research Methodology. The methodology described in this chapter encompasses the design and implementation of the web-based MMSE application. This includes the technical architecture, the development of cognitive tasks, the integration of artificial intelligence components, and the design of the user interface.

Chapter 4: Implementation. This chapter presents the technical details of the application's practical implementation. It covers the system architecture, AI integration process, backend and frontend development, data management strategies, and security measures. The chapter emphasizes the thorough development process of the web-based MMSE system with AI enhancements, concentrating on the crucial design choices and technical difficulties encountered during development.

Chapter 5: Results and Evaluation. This chapter outlines the implementation results and the findings from the evaluation study. It details the system's performance indicators, outcomes of various testing processes, and the comparison with conventional MMSE approaches. The chapter also analyzes the application's usability, discusses metrics for performance-related detection and monitoring of mild cognitive impairment, and examines how well the implementation met the initial objectives. Finally, it reviews the strengths and limitations of the web-based application, underlining its innovative aspects and discussing the implications for cognitive assessment practices in light of existing literature.

Chapter 6: Conclusion. This final chapter summarizes the essential findings and considers the general implications of the research work. It also suggests future research directions and potential enhancements to the web-based MMSE application.

Each chapter builds upon the previous one to thoroughly understand the web-based cognitive assessment application's development, evaluation, and implications. This thesis aims to contribute significantly to the field by improving the accessibility, efficiency, and effectiveness of cognitive assessments through innovative technology.

2 Baseline and Literature Review

Cognitive assessment is crucial in diagnosing and managing various neurological and psychiatric conditions. Technological advancements have driven a significant evolution in cognitive assessment methods and tools, progressing from traditional paper-and-pencil tests to sophisticated computerized, web-based, and AI-assisted approaches. This review examines this evolution, providing a comprehensive overview of traditional and modern cognitive assessment techniques. It analyzes the strengths and limitations of various methods, explores how recent innovations have improved accessibility and utility, and addresses the ethical considerations surrounding integrating new technologies in cognitive testing.

2.1 The MMSE

The Mini-Mental State Examination (MMSE), developed by Folstein et al. in 1975, has become a cornerstone of cognitive assessment [5]. This widely used screening tool provides a quick and standardized method for assessing cognitive function across various domains, including:

- awareness of both temporal and spatial contexts;
- registration and recall;
- attention and calculation;
- language, including naming, repetition, and comprehension;
- visuospatial abilities.

The MMSE's strength lies in its ability to quickly screen for global cognitive impairment, typically taking 5-10 minutes to administer. This efficiency has contributed to its widespread adoption in clinical settings, particularly in primary care and geriatric medicine [29]. However, recent research emphasizes the growing role of AI in enhancing traditional cognitive assessment methods like the MMSE, potentially increasing their accuracy and scope [30].

2.1.1 MMSE Selection Rationale

Several factors influence the decision to develop a web-based AI solution based on the MMSE:

- **Widespread adoption and standardization:** The MMSE provides a standardized framework familiar to many clinicians and researchers [5, 29].

- **Efficiency and ease of use:** Its quick administration makes it suitable for integration into a web-based platform [21].
- **Comprehensive initial screening:** The MMSE covers a wide range of cognitive domains, making it effective for initial screening [21].
- **Established clinical utility:** Numerous studies have validated the MMSE's effectiveness in diagnosing cognitive impairments and tracking changes over time [31, 32]. Furthermore, recent advances have shown that AI-enhanced MMSE can offer even greater diagnostic precision, especially in the early detection of conditions like mild cognitive impairment (MCI) and Alzheimer's disease (AD) [33].
- **Adaptability to digital platforms:** The MMSE's structured format makes it adaptable to digital platforms and AI automation [7].

While other cognitive assessments, such as the Montreal Cognitive Assessment (MoCA), are also widely used, the MMSE was chosen for this web-based AI solution due to its widespread adoption, ease of use, and established clinical utility. The MMSE's structure is particularly well-suited for digital adaptation, allowing for a seamless transition to a web-based platform [34]. This adaptability is increasingly relevant as AI integration in cognitive assessments grows, allowing for more personalized and context-aware applications.

2.1.2 MMSE Scoring and Significance

The MMSE is scored on a scale of 30 points, with higher scores indicating better cognitive function. Generally, scores are interpreted as follows:

- 24 points and above: typically associated with normal cognitive abilities;
- 18 to 23 points: may suggest the presence of mild cognitive difficulties;
- below 18 points: often indicative of more significant cognitive concerns.

These cutoff points should be interpreted cautiously, considering factors such as age, education, and cultural background [35].

2.1.3 Limitations and Criticisms

Despite its utility, the MMSE has several limitations:

- ceiling effect in high-functioning individuals;
- limited sensitivity to mild cognitive impairment;

- influence of education and cultural factors on performance;
- inadequate assessment of executive function;
- practice effects with repeated administration.

The ceiling effect and cultural bias can lead to potential missed diagnoses or misclassifications [36, 37]. The limited assessment of executive function is particularly problematic for early detection of certain types of cognitive impairment [38]. Although the MoCA addresses some of these limitations by including more challenging tasks, the MMSE was selected due to its established clinical utility and easier integration into digital platforms. Furthermore, AI-enhanced tools, such as the Digital Clock Drawing Test (dCDT), demonstrate the potential to overcome some of these limitations by providing more detailed analyses of cognitive functions [39].

2.2 Digital Cognitive Assessment

The transition from conventional paper-and-pencil tests to computerized platforms and, more recently, web-based applications has been a defining feature of cognitive assessment tools evolution. This transition has significantly improved the administration, accuracy, and accessibility of cognitive assessments [6, 7].

2.2.1 Advantages

Computerized and web-based cognitive assessments have evolved from simple digitized versions of paper-and-pencil tests to sophisticated, adaptive platforms. Key advantages include:

- standardized administration and scoring, minimizing variability;
- enhanced precision in measuring response times and processing speed;
- ability to generate extensive normative databases;
- reduced time and cost of administration;
- automated data collection and analysis;
- adaptive testing, dynamically adjusting based on performance;
- improved detection of subtle cognitive changes;
- reduced human error in administration and scoring.

These advantages have significantly boosted the adoption of digital assessments in clinical and research settings [40]. The precise measurement of reaction times and processing speed has been particularly valuable for detecting subtle cognitive changes, which may not be apparent through traditional methods [7]. Additionally, recent AI advancements have led to the development of highly sensitive and specific cognitive assessments capable of detecting early cognitive decline with unprecedented accuracy [41].

Web-based assessments further enhance these benefits by offering:

- greater accessibility for remote or underserved populations;
- convenience of conducting assessments from various devices and locations;
- potential for more frequent testing and longitudinal tracking;
- increased engagement through gamification and interactive interfaces [42].

Recent research demonstrates that platforms like MyCognition and BrainTest have made cognitive tests more accessible and user-friendly by employing cloud-based services and AI-driven analytics, which provide real-time feedback and adapt based on user performance [26, 43].

2.2.2 Trends

Key trends in digital cognitive assessment include:

- mobile-first design: optimizing tools for smartphones and tablets [7];
- gamification: incorporating game elements to boost engagement and motivation [42];
- adaptive testing: adjusting task difficulty dynamically based on performance [44];
- passive data collection: leveraging smartphone sensors and usage patterns for continuous cognitive monitoring [45];
- virtual reality-based assessments: creating immersive, ecologically valid testing environments [46].

These innovations enhance the engagement, efficiency, and individual tailoring of cognitive tests. For instance, recent advancements in gamification have significantly improved user engagement and data quality by transforming assessments into interactive experiences [47].

Web-based assessments substantially improve the potential for longitudinal monitoring of cognitive changes [26]. The convenience of home-based testing facilitates more frequent assessments, enabling:

- monitoring cognitive decline progression;
- evaluating intervention effectiveness;
- detecting subtle changes in cognitive performance;
- supporting early detection and intervention.

This comprehensive view of cognitive health over time allows for more informed decision-making in cognitive health management [7].

A groundbreaking study by Dagum [45] showcased the potential of passive data collection through smartphones for cognitive assessment. By analyzing phone usage patterns and sensor data, researchers were able to detect fluctuations in cognitive function with high precision. This approach opens up new avenues for continuous and unobtrusive cognitive monitoring in real-world environments.

2.2.3 Challenges

Despite their advantages, digital cognitive assessments face several challenges:

- technological accessibility: limited access to hardware and internet in some areas [6];
- user familiarity: difficulties for older adults in navigating digital interfaces [7];
- data security and privacy concerns: need for stringent protection of sensitive health information [48];
- potential exacerbation of healthcare disparities: unequal access due to the digital divide [49];
- lack of personal interaction: potential loss of subtle behavioral cues observed in traditional assessments [6];
- validation concerns: ensuring equivalence with paper-and-pencil counterparts [26];
- consistency of testing environments: variability in settings for web-based tests [19];
- potential for cheating or assistance: difficulty in ensuring test integrity in unsupervised settings [50].

Addressing these challenges is crucial for the widespread adoption and effective use of digital cognitive assessments.

A recent study by Feenstra [50] highlighted the potential for cheating in unsupervised web-based cognitive assessments. The researchers found that a small but significant proportion of participants sought external assistance or used prohibited aids during online tests. This underscores the need for robust measures to ensure test integrity in digital assessment environments.

2.3 AI in Assessment

Artificial Intelligence (AI) can enhance cognitive assessments by automating scoring processes, providing detailed interpretations of test results, predicting cognitive decline, and personalizing assessments based on individual characteristics and performance. Through the analysis of extensive datasets, AI can detect nuanced patterns and associations that might be overlooked by traditional methods, thereby enabling more precise and unbiased evaluations. Machine learning models can be trained to recognize early markers of cognitive decline, potentially supporting early intervention and proactive management of cognitive health [51].

Large Language Models (LLMs), such as GPT-4 [52], have significant potential in cognitive assessment. These models can understand and generate human-like text, facilitating the development of sophisticated and adaptive testing environments, generating a wide range of questions and prompts, and analyzing verbal and written responses, providing detailed insights into language and communication skills. Other AI techniques, including machine learning and neural networks, enhance the accuracy and reliability of cognitive assessments by continually learning from new data. These techniques can improve the sensitivity and specificity of cognitive evaluations, potentially identifying subtle cognitive changes that might be missed by traditional assessment methods.

2.3.1 AI Applications and Ethics

Natural Language Processing (NLP) and computer vision are two AI technologies with transformative potential in cognitive assessment:

- NLP applications include automated analysis of speech patterns, sentiment analysis, extraction of semantic content from free-form responses, and automated scoring of verbal fluency tasks [53].
- some ways computer vision is used are to automatically grade visuospatial tasks, analyze facial expressions and eye movements during cognitive tests, observe test-taking behaviors in unsupervised settings, and recognize gestures for more interactive cognitive tasks.

These technologies enhance the scope and depth of cognitive evaluations, providing comprehensive insights into an individual's cognitive health.

A groundbreaking study by Petti et al. [41] demonstrated the potential of AI-powered speech analysis to detect cognitive impairment. By applying machine learning algorithms to analyze various acoustic and linguistic characteristics of speech samples, the researchers could accurately differentiate between healthy individuals and those with mild cognitive impairment. This approach offers a non-invasive and potentially more sensitive method for the early detection of cognitive decline.

The integration of AI into cognitive assessment offers numerous potential benefits:

- increased objectivity and consistency in scoring;
- enhanced detection of subtle cognitive changes;
- more efficient and cost-effective assessment processes;
- potential for continuous monitoring through passive data collection;
- personalized assessment and intervention strategies;
- improved accessibility through language-agnostic assessment methods.

However, several limitations and ethical concerns must be addressed:

- potential bias in AI algorithms reflecting biases in training data [54];
- lack of transparency in "black box" AI decision-making processes [55];
- privacy concerns related to collecting and storing sensitive cognitive data;
- ethical implications of predictive algorithms (e.g., early detection of dementia risk);
- challenges in ensuring the reliability and validity of AI-based assessments across diverse populations [56].

Addressing these ethical concerns is crucial for the responsible development and implementation of AI in cognitive assessment. Efforts are needed to ensure transparency, fairness, and accountability in AI systems, as well as to develop guidelines for the ethical use of AI in clinical practice.

2.3.2 AI Techniques for Cognitive Assessment

Several case studies illustrate the successful integration of AI in cognitive assessment:

- **Digital Clock Drawing Test (dCDT):** An AI-enhanced version of the traditional clock drawing test, using machine learning to analyze digital pen strokes for a more detailed and objective assessment of cognitive function [57].
- **Speech-based assessments:** AI systems that detect early signs of cognitive decline by analyzing speech patterns, offering a non-invasive and easily administered screening tool [58].

- **Adaptive cognitive testing:** AI-powered platforms that dynamically adjust task difficulty in real-time based on performance, improving assessment efficiency and accuracy [59].
- **Computer vision for facial expression analysis:** AI systems that analyze micro-expressions and eye movements during tasks, providing insights into cognitive processing and emotional states [60].

These studies demonstrate AI's potential to enhance the sensitivity, efficiency, and accessibility of cognitive assessments. For example, Zorluoglu et al. [59] showed that an AI-powered adaptive testing platform could dynamically adjust task difficulty, resulting in more efficient and accurate assessments than traditional fixed-item tests, particularly for individuals at the extremes of cognitive ability.

2.4 Ethical Considerations

Developing and deploying AI tools in cognitive assessment brings significant ethical challenges. Ensuring that these systems are fair, transparent, and accountable is critical to gaining trust and ensuring their effectiveness in clinical settings. Key considerations include:

- rigorous testing of AI systems for biases across diverse populations;
- implementation of mechanisms to explain AI decisions (explainable AI);
- ensuring the use of diverse and representative training data;
- regular auditing and validation of AI algorithms;
- developing clear guidelines for the ethical use of AI in cognitive assessment;
- protecting patient privacy and ensuring data security;
- addressing potential healthcare disparities that could be exacerbated by AI technologies.

Gianfrancesco et al. [56] conducted a comprehensive review of the challenges in ensuring fairness and reliability in AI-based health assessments. They emphasized the importance of diverse representation in training data and proposed a framework for evaluating AI systems across different demographic groups to identify and mitigate potential biases.

2.4.1 Cultural Adaptation Methods

Cultural adaptation and linguistic validation are essential for ensuring the accuracy and relevance of cognitive assessments across diverse populations. Key methods include:

- conducting cross-cultural translation and validation studies to ensure linguistic and cultural appropriateness;
- adapting test items to reflect the cultural norms and values of the target population;
- performing linguistic validation to ensure conceptual and functional equivalence across languages;
- undertaking pilot testing and psychometric validation in diverse samples to confirm reliability and validity;
- ongoing monitoring and refinement based on cross-cultural data to maintain relevance and accuracy;
- developing culture-specific normative data to enhance the precision of cognitive assessments in different populations;
- collaborating with local experts and community representatives to ensure cultural appropriateness.

A recent study by Rosli et al. [61] highlighted the critical importance of cultural adaptation in cognitive assessment. They developed and validated a culturally appropriate version of a cognitive screening tool for use in a Southeast Asian population, significantly improving diagnostic accuracy compared to the original Western version.

2.4.2 Web-Based Assessment Fairness

Web-based cognitive assessments introduce unique fairness and cultural sensitivity challenges, especially given the diverse populations they aim to serve. Strategies to ensure fairness include:

- considering cultural differences in cognitive performance to ensure assessments do not disadvantage any particular group;
- designing assessments that are accessible to individuals with varying levels of technological proficiency to prevent digital literacy from affecting outcomes;
- providing instructions and feedback in multiple languages to cater to non-native speakers;

- implementing adaptive testing techniques that tailor assessments to individual cultural and linguistic contexts;
- conducting validation studies to ensure fairness and reliability across diverse populations;
- incorporating culturally relevant stimuli and test items to engage participants with content that is familiar and appropriate to their cultural context.

A study by Tsoy et al. [62] demonstrated the importance of cultural adaptation in web-based cognitive assessments. They developed a culturally adapted online cognitive assessment platform for Russian-speaking individuals, significantly improving cognitive impairment detection accuracy compared to non-adapted tools.

2.5 Current Technology Gaps

Despite advancements in cognitive assessment technologies, several gaps remain:

- **Ecological validity:** Many assessments struggle to capture real-world cognitive functioning.
- **Longitudinal measurement:** More sensitive tools are needed to detect subtle cognitive changes over time.
- **Multimodal data integration:** Current approaches often fail to fully integrate cognitive test data with neuroimaging and genetic information.
- **Adaptive algorithms:** Potential for more sophisticated, AI-driven algorithms to optimize test efficiency.
- **Virtual reality and passive data collection:** Promising avenues for more immersive and continuous assessments.
- **Diversity and accessibility:** Many tests lack comprehensive normative data for diverse populations.
- **Ethical guidelines and clinical integration:** There is a need for guidelines governing AI use and better integration into clinical practice.

Recent research has begun addressing some gaps. Parsons et al. [63] developed a VR-based assessment tool with improved sensitivity for detecting mild cognitive impairment. Bzdok et al. [64] proposed a machine learning framework integrating neuroimaging, cognitive test scores, and genetic data to enhance impairment prediction.

Future research directions should focus on:

- developing more ecologically valid assessment paradigms [63];
- creating AI-driven adaptive testing algorithms [59];
- establishing diverse normative databases [56];
- formulating ethical frameworks for AI use in assessment [54];
- investigating novel biomarkers and their integration with test data [64].

Addressing these gaps will require interdisciplinary collaboration to enhance the accuracy, accessibility, and real-world applicability of cognitive assessments.

2.6 Summary of Current State

The field of cognitive assessment has evolved significantly, transitioning from traditional paper-and-pencil tests to sophisticated computerized, web-based, and AI-assisted approaches. This evolution has brought increased precision, efficiency, and accessibility while also introducing new challenges and ethical considerations.

Key developments include:

- computerized and web-based assessments offering standardized administration and improved accessibility;
- AI integration enables more accurate scoring, detailed interpretation, and potential prediction of cognitive decline;
- emerging technologies like virtual reality and passive data collection methods showing promise for more ecologically valid assessments [46].

However, significant gaps remain, including:

- need for more ecologically valid tests and better integration of multimodal data [63];
- challenges in ensuring fairness, cultural sensitivity, and ethical use of AI in assessments [54];
- necessity for comprehensive ethical guidelines and improved clinical integration of advanced tools [55].

The future of cognitive assessment lies in interdisciplinary collaboration, balancing technological innovation with ethical considerations and clinical relevance. The goal remains to improve patient outcomes through early detection, accurate diagnosis, and effective monitoring of cognitive function across diverse populations.

3 Research Methodology

This chapter elaborates on the research methodology for the development and evaluation process of an AI-enhanced Web-based MMSE application.

Design Science Research (DSR) is highly suitable for developing and testing such a novel IT artifact in the healthcare sector. DSR is a problem-solving paradigm that seeks to create and evaluate IT artifacts designed to solve identified organizational problems [65, 66]. In the context of AI-powered healthcare solutions, DSR provides a structured framework for addressing complex, multifaceted challenges while ensuring rigorous evaluation and practical relevance [67].

This research adopts a tailored Design Science Research (DSR) methodology, emphasizing four iterative phases, as depicted in Figure 1:

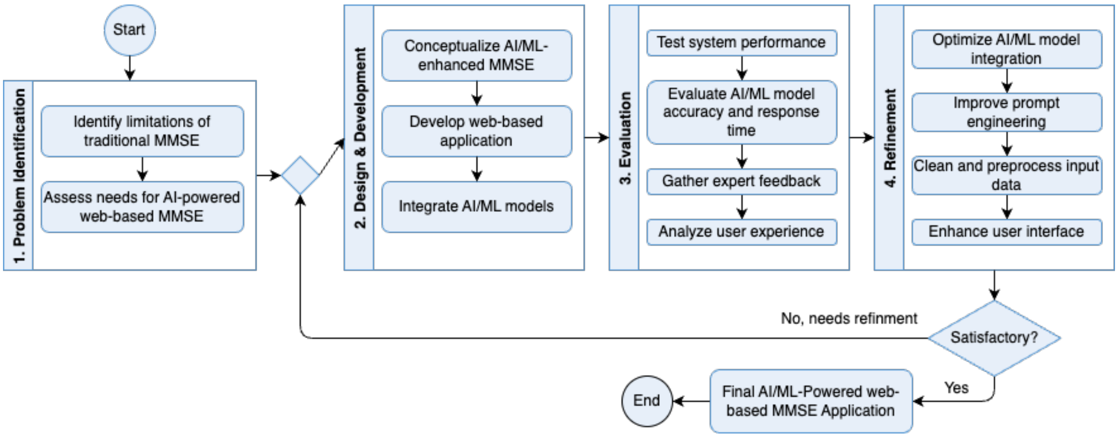

Figure 1. Iterative development process for the AI/ML-powered web-based MMSE application, illustrating the key stages from problem identification through design, evaluation, and refinement.

1. **Problem Identification:** The limitations of traditional MMSE were identified and the needs for an AI-powered web-based MMSE were assessed.

2. **Design & Development:** This phase involved conceptualizing the AI/ML-enhanced MMSE, developing the web-based application, and integrating AI/ML models.

3. **Evaluation:** The system's performance was tested, AI/ML model accuracy and response time were evaluated, expert feedback was gathered, and user experience was analyzed.

4. **Refinement:** Based on evaluation results, this phase focused on optimizing AI/ML model integration, improving prompt engineering, cleaning and preprocessing input data, and enhancing the user interface.

The process is iterative, with a decision point after the refinement phase to determine if further development is needed. This approach ensures continuous improvement until a satisfactory AI/ML-powered web-based MMSE Application is achieved.

This adaptation of the traditional six-step DSR process was chosen for several key reasons:

1. **Agility and Iteration:** The four-phase cycle allows for quicker iterations, enabling faster response to new AI developments or changing requirements in cognitive assessment tools.
2. **Focus on Continuous Refinement:** The dedicated 'Refinement' phase and explicit decision point for iteration emphasize the importance of ongoing improvement, which is crucial in AI-powered healthcare applications.
3. **Integration of AI/ML Specifics:** Each phase explicitly includes AI/ML considerations, ensuring that these aspects are central to the development process.
4. **Streamlined Process for Time Constraints:** This condensed model allows for a more focused and efficient research process within the time limitations of a master's thesis.
5. **Alignment with Software Development Practices:** The four-phase cycle mirrors modern software development methodologies like Agile or DevOps, facilitating better integration with real-world development practices.
6. **Enhanced Flexibility:** This model offers greater flexibility to adapt to the unique challenges of developing an AI-powered cognitive assessment tool.
7. **Emphasis on Evaluation and User Feedback:** Significant emphasis is placed on both technical performance assessment and user experience analysis, which are crucial for healthcare applications.

This adaptation of DSR maintains the rigorous, problem-solving approach of the original methodology while tailoring it to the specific needs of AI-powered healthcare solution development. It provides a structured yet flexible framework that is well-suited to the complex, iterative nature of developing an AI-enhanced web-based MMSE application.

The following sections will detail each phase of this adapted DSR methodology as applied to the development of the AI-powered web-based MMSE application.

3.1 Problem Identification

Designing a Web-based AI-powered MMSE requires understanding the existing challenges to cognitive assessment and the needs of stakeholders. The limitations of traditional MMSE identified earlier provide the context for the proposed solution.

3.1.1 MMSE Limitations

Several limitations exist in current MMSE administrations, which reduce its effectiveness and popularity:

- **Time Constraints:** Traditional MMSE administration requires approximately 10–15 minutes per patient, which can be problematic in busy clinical settings [5].
- **Scoring Inconsistencies:** Manual scoring introduces the potential for human error and inconsistencies between administrators [21].
- **Limited Accessibility:** The need for in-person administration restricts access for patients in remote or underserved areas [6].
- **Cultural and Educational Bias:** The standard MMSE may not account for cultural differences or varying educational backgrounds, potentially leading to inaccurate assessments [68].
- **Lack of Longitudinal Tracking:** Traditional methods often fail to efficiently track cognitive changes over time [18].
- **Limited Sensitivity:** The MMSE may not detect subtle cognitive changes, particularly in early stages of cognitive decline [32].

These limitations underscore the need for an innovative approach to cognitive assessment to address these challenges while maintaining or improving the validity and reliability of the test.

3.1.2 Needs Assessment

Based on identified limitations in the literature and analysis of the field, this research compiles a list of key requirements for an AI-powered cognitive assessment tool:

- **Efficiency:** Reduce administration time without compromising test validity.
- **Standardization:** Ensure consistent test administration and scoring across different settings.
- **Accessibility:** Enable remote administration to reach underserved populations.
- **Adaptability:** Accommodate cultural and educational differences through dynamic test adjustments.
- **Longitudinal Tracking:** Facilitate easy monitoring of cognitive changes over time.

- **Data Security:** Implement robust measures to protect patient data and ensure privacy.

These requirements guide the development of the AI-powered web-based MMSE. They ensure the tool addresses key needs and overcomes limitations of traditional cognitive assessments identified in the literature.

3.2 Artifact Design and Development Process

The AI-powered web-based Mini-Mental State Examination (MMSE) application employs a modern, scalable architecture designed to meet the needs identified in the problem identification phase. This section details the system's design, including its overall architecture, key components, and the assessment process flow.

3.2.1 Conceptualizing the MMSE and System Design

The conceptual design of the AI-powered web-based MMSE application involved transforming traditional MMSE tasks into automated versions suitable for digital administration. This process aimed to maintain the integrity of the original assessment [5] while leveraging the capabilities of web-based technologies and artificial intelligence [6, 7]. As cognitive assessment tasks transition from paper-based to web-based administration, it is important to consider the test's psychometric properties and how the digital medium might affect test performance [19]. The details of the traditional MMSE tasks are provided in the appendix[1].

3.2.2 Challenges and Solutions in Automating MMSE Tasks

Automating the MMSE tasks involved several challenges, requiring careful consideration of the limitations and capabilities of current technology. Below are detailed descriptions of the problems encountered and the solutions implemented for each task category:

Orientation: Traditionally, the tester asks the individual verbal questions related to time and place (e.g., "What is the date today?" or "Where are we right now?"). Points are given based on the accuracy of the responses. In the automated version, multiple choice questions are used to assess these aspects. The system presents questions related to time and place, and the user selects the correct answer from a list of options. This approach ensures accurate scoring and immediate feedback, facilitating the assessment process.

Registration: In the traditional MMSE, the individual is asked to repeat three words (e.g., "apple, penny, table"). The automated version uses image recognition and multiple choice to realize this task. The user is shown images corresponding to the words and

---

[1] See Appendix III for the details of the traditional MMSE tasks.

Table 1. Transformation of MMSE Tasks to Automated Versions

Task Category	Traditional MMSE	Automated Version
Orientation	Verbal questions about time and place	Multiple choice questions (QuestionId 2-5, 15), Text input (QuestionId 17-21)
Registration	Verbal repetition of three words	Image recognition and multiple choice (QuestionId 6-8)
Attention and Calculation	Serial subtraction or spelling word backwards	Numeric input for subtraction task (QuestionId 9)
Recall	Recall of three words	Text input (QuestionId 10)
Language	Naming objects	Image recognition with text input (QuestionId 11-12)
Repetition	Repeat a phrase	Voice input and AI transcription (QuestionId 1)
Complex Commands	Three-stage command	Drag and drop interface (QuestionId 16)
Reading	Read and obey a written command	Text input to confirm understanding (QuestionId 13)
Writing	Write a sentence	Text input with AI grammar check (QuestionId 14)
Visuospatial Skills	Copy a design	Drawing interface with AI image analysis (QuestionId 22)

must select the correct ones. If the user incorrectly identifies the images, the system provides immediate feedback and logs the response for further evaluation.

Attention and Calculation: Traditionally, tasks like serial subtraction (e.g., subtracting 7 from 100) or spelling words backward are used. In the automated version, numeric input fields are provided for tasks such as subtraction. The system processes the input, checking for correct and incorrect answers, and provides feedback accordingly. Incorrect answers are noted, and hints or step-by-step guidance can be offered to help the user understand the mistake.

Recall: This task involves recalling three previously mentioned words. In the automated version, the user inputs their answers into text fields. The system uses text recognition to evaluate the responses, allowing for common misspellings or variations in phrasing. Correct answers are scored, and incorrect ones prompt the user to try again or move to the next task with feedback.

Language: Naming objects traditionally involves the individual looking at an object and naming it (e.g., "What is this?"). The automated version utilizes image recognition combined with text input. The user views an image and types the name of the object. The system evaluates the input for correctness, taking into account synonyms and common spelling errors.

Repetition: The traditional task requires the individual to repeat a phrase. The automated version uses voice input, and AI transcribes the spoken response. The transcription is then evaluated for accuracy. Incorrect repetitions are flagged, and the user is asked to try again or move on with feedback.

Complex Commands: In the traditional MMSE, this task involves following a series of verbal instructions, such as "Pick up a sheet of paper with your right hand, fold it, and then set it down on the floor" [5]. The automated version transforms this into an interactive drag-and-drop interface. Users perform the tasks by dragging items to specified locations as directed. The system tracks each user action, scoring them based on accuracy and providing immediate feedback if any errors occur.

Reading: In this task, the user is required to read and execute a simple written instruction, such as "Close your eyes." The automated system displays the command as on-screen text and asks the user to confirm their understanding. The system then assesses whether the correct action was taken, ensuring that the user has comprehended the instruction.

Writing: Traditionally, the individual writes a sentence. In the automated version, the user inputs the sentence into a text field. The system uses AI to check the grammar and structure of the sentence, providing feedback on any errors and scoring the response accordingly.

Visuospatial Skills: The traditional task involves copying a design. The automated version provides a drawing interface where users draw using a mouse or touchscreen. AI analyzes the drawing for accuracy against a reference design. Feedback is provided based on the correctness and completeness of the drawing.

By addressing these challenges, the MMSE application provides a robust, automated solution for cognitive assessments, leveraging modern web technologies and advanced AI capabilities to ensure accuracy and reliability in its assessments.

3.3 Implementation and Development

The implementation process of the AI-powered web-based MMSE application followed a structured approach, utilizing modern software engineering practices and technologies. The application was built using a client-server architecture, comprising the following components:

- Backend: a Java-based Spring Boot application;
- Frontend: a Vue.js single-page application;
- Database: PostgreSQL for persistent data storage;

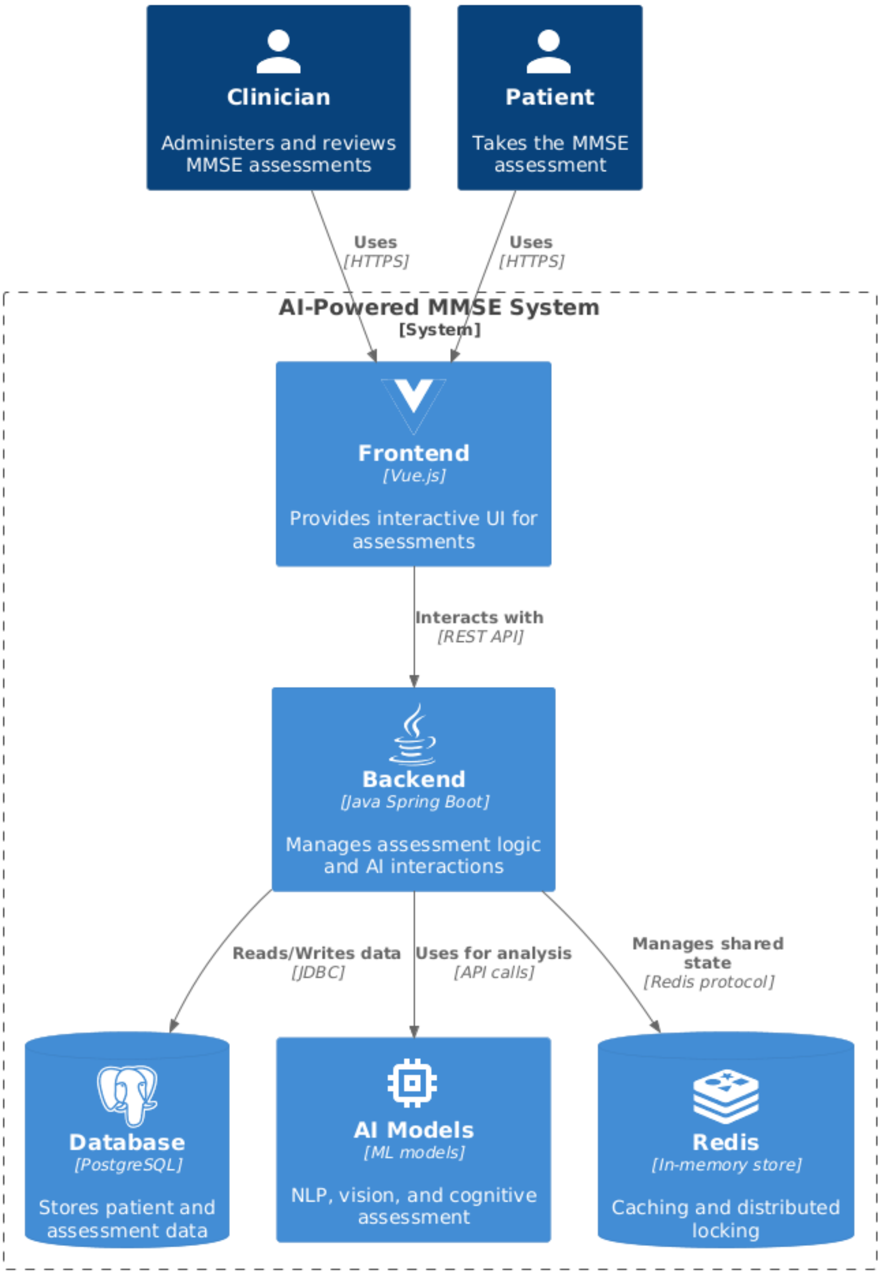

Figure 2. Diagram of the AI-Enabled MMSE Application Structure

- External services: AI models (ChatGPT, Ollama with Llama 3.1:70B) for answer evaluation;
- Caching: Redis for distributed caching and session management;

- Message queue: RabbitMQ for asynchronous processing.

Figure 2 shows the structure and the AI-powered web-based MMSE application, and also illustrates interactions between clinicians, patients, the frontend and backend components, the database, AI models, and Redis for caching and distributed locking purposes. In the rest of this section, we describe each of these components in more detail.

Backend Design The Spring Boot application forms the core of the system, managing business logic, data repositories, and AI service interactions. Key features include:

- RESTful API with versioning;
- service layer implementing core business logic;
- data access layer using Spring Data JPA;
- integration with external AI services;
- security implementation with JWT-based authentication;
- caching and asynchronous processing for improved performance.

Frontend Design The Vue.js frontend provides a user-friendly interface for both test administrators and test-takers, featuring:

- responsive design for various devices;
- state management using Vuex;
- component-based architecture for maintainability;
- accessibility features adhering to WCAG guidelines;
- progressive loading and offline support.

Development Environment Setup and Version Control GitHub served as the primary version control and collaboration platform, offering robust features that support distributed development and integration with various development tools [69]. The repository structure separated the front- and back-end code, facilitating organized development. JHipster bootstrapped the initial application structure, providing a solid foundation with pre-configured best practices for security, testing, and API development [70].

Iterative Development Approach and Continuous Integration The development process followed an iterative model closely aligned with agile principles. The initial phase involved mapping out all the MMSE questions, which served as the basis for database design and the overall application structure. The development workflow proceeded as follows:

1. question mapping and requirements analysis;
2. database schema design based on question requirements;
3. implementation of backend services and APIs;
4. frontend component development;
5. integration of AI and machine learning models;
6. continuous testing and refinement.

A GitHub Actions workflow automated the continuous integration (CI) pipeline [71]. This configuration ensured that each commit triggered codebase compilation, execution of unit and integration tests, and performance of code quality checks.

Quality Assurance, Testing, and Dependency Management A focused testing strategy maintained high code quality and ensured the reliability of the MMSE application. The testing approach included:

- **Unit Testing:** JUnit tested individual components and functions in isolation for the Java back-end, while Jest handled testing for the JavaScript front-end.
- **Integration Testing:** Tests for API endpoints and service interactions ensured correct behavior when combining components, particularly for validating the integration of AI models with the application logic.
- **Manual Testing:** Regular manual testing sessions verified the user interface, user experience, and general functionality of the application.
- **Accessibility Considerations:** The development process incorporated best practices for web accessibility, guided by the Web Content Accessibility Guidelines (WCAG) [72].

Dependabot integration into the GitHub repository maintained up-to-date dependencies and addressed potential security vulnerabilities [73]. Dependabot automatically created pull requests to update outdated dependencies, ensuring that the project benefited from the latest library versions and security patches.

Code Quality, Monitoring, and Error Tracking Several tools were integrated into the development process to maintain the quality and consistency of the code:

- ESLint for JavaScript linting;
- Prettier for code formatting;
- SonarLint for identifying and fixing code quality issues.

These tools were primarily used within the IntelliJ IDEA development environment, providing real-time feedback during coding. ESLint and Prettier configurations were defined in IntelliJ, allowing immediate code style and quality checks as developers wrote the code.

While not implemented in a production environment, the development process considered future monitoring and error tracking:

- Plans for log aggregation using the ELK stack (Elasticsearch, Logstash, Kibana) aimed at centralized log management and analysis [74].
- Considerations for integration with error tracking services such as Sentry aimed at real-time error monitoring and reporting in production environments [75].

3.3.1 Database and Performance

The PostgreSQL database schema was designed to efficiently store and retrieve test data, user information, and assessment results. Key entities include:

- User management (mmse_user, mmse_authority, mmse_user_authority);
- Test management (test_entity, user_answer, patient_profile, test_entity_hash, media_recording);
- AI integration (dolphin_question, orientation_to_place_answer).

This enhanced schema includes relationships and indexing strategies to ensure optimal performance and data integrity. Figure 3 provides a detailed view of the database schema, illustrating the relationships between key entities such as user management, test management, and AI integration components.

Here is a detailed overview of the primary entities, with the user management section comprising:

- **mmse_user**: Stores user credentials and profile information.
- **mmse_authority**: Contains roles and permissions.
- **mmse_user_authority**: Links users to their roles.

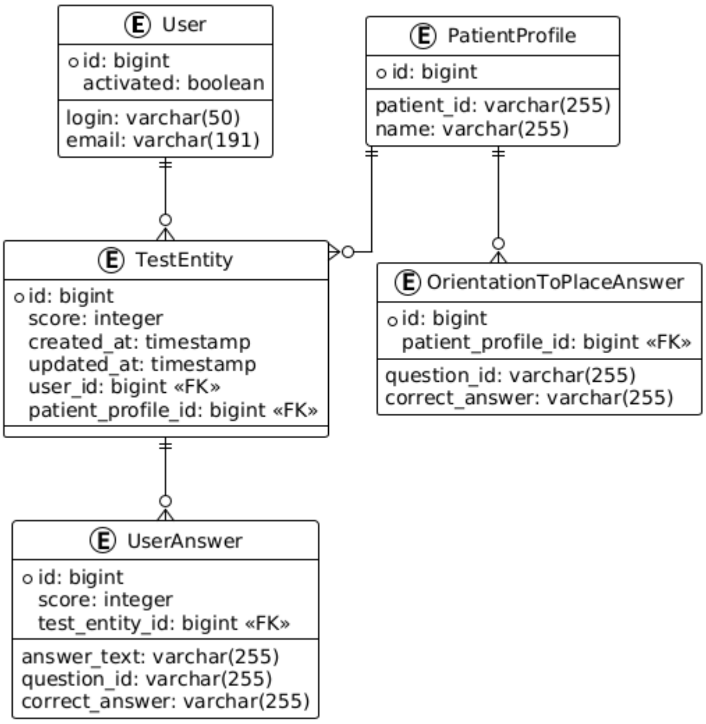

Figure 3. Database Schema for the AI-Powered MMSE Application. This schema includes user management, test management, patient profiles, and AI integration.

This section related to MMSE testing entities:

- **test_entity**: Represents individual test instances.
- **user_answer**: Stores user responses to test questions.
- **patient_profile**: Contains patient information related to test results.
- **test_entity_hash**: Ensures the integrity of the test data with hash values.
- **media_recording**: Links media files (e.g., audio recordings) to test entities and questions.

The **AI Integration** section includes:

- **dolphin_question**: Stores questions used in AI-driven assessments.
- **orientation_to_place_answer**: Contains pre-defined answers and options for orientation questions.

Performance and Security Enhancements While developing the AI-driven web-based MMSE application, performance and security were the highest priorities. This part details the methods employed to enhance database performance, secure data, and sustain solid data flow and integration processes. Additionally, considerations for scalability and upkeep are covered to guarantee the application's efficiency and security over time.

Indexing Indexes are created on frequently queried fields such as user_id, patient_profile_id, and timestamps to enhance query performance. Composite indexes are used where appropriate to speed up multi-column searches.

Security Data at rest are encrypted using PostgreSQL's native encryption features. Access controls are implemented using roles and permissions to restrict access to sensitive data. Regular backups are scheduled to prevent data loss.

Data Flow ETL (Extract, Transform, Load) processes are used to integrate data from various sources into the PostgreSQL database. This ensures a seamless data flow between different components of the system.

Scalability The database design considers future scalability with partitioning and sharding strategies. Maintenance is facilitated through regular monitoring and optimization practices.

Security and Privacy Considerations The implementation includes robust security measures [76]:

- JWT-based authentication and role-based access control;
- strict CORS configuration;
- comprehensive audit logging;
- proactive security monitoring.

This implementation process ensured the development of a secure, efficient, and user-friendly AI-powered web-based MMSE application, leveraging modern technologies to enhance cognitive assessment practices.

In conclusion, the artifact design and development process for the AI-powered web-based MMSE application leveraged modern software engineering practices to ensure efficient and high-quality development. The combination of iterative development, version control, automated testing, and continuous integration enabled the creation of a reliable and scalable application poised to revolutionize cognitive assessment practices.

3.3.2 Assessment Process and AI Integration

The use of the application follows a structured process for administering the MMSE, as illustrated in Figure 4. This process consists of the following steps:

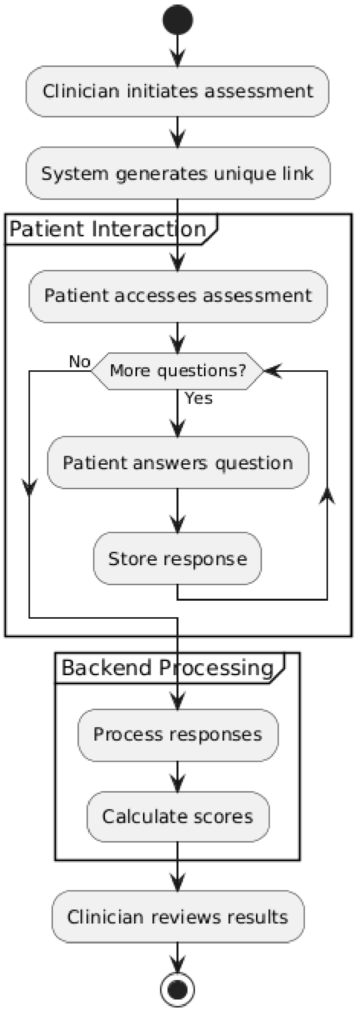

Figure 4. A detailed view of the interactions and data flow during the assessment process.

1. **Test initiation by the healthcare professional**: The clinician initiates the assessment by generating a unique link for the patient.
2. **Patient information collection**: The patient accesses the assessment through the provided link, and the system loads or resumes the current state.

3. **Administration of the MMSE through sequential question presentation**: The patient answers various types of questions, including Multiple Choice, Text, Voice, and Drawing. These interactions are captured in the "Patient Interaction" partition of the diagram.

4. **Answer collection through various input methods**: Once the patient answers a question, the response is temporarily stored on the frontend. Upon submission of the response by the patient, the system initiates the transition to backend processing. This is a critical juncture in the workflow as indicated by the direct link from "Response Submitted" to the decision node (diamond).

5. **AI-powered analysis of patient responses**: At the decision node, the system evaluates whether the submitted response requires AI processing. This decision is based on the type of response and the nature of the question answered by the patient.

   - *If AI processing is required*: The system routes the response to the appropriate AI model. For instance:
     - **Llama 3.1:70B** for complex language comprehension.
     - **ChatGPT 4o** for processing drawings.
     - **Specialized models** for speech-to-text, synonym detection, and semantic understanding.
   - *If AI processing is not required*: The response undergoes standard processing, bypassing the AI models.

   After processing, whether through AI or standard methods, the response is stored in the database. This ensures that all responses are securely saved for future analysis and score calculation.

6. **Generation of clinical insights and result presentation**: Scores are calculated periodically, incorporating concurrency control to manage simultaneous access and updates. This step is crucial for maintaining the integrity and accuracy of the assessment scores. The calculated scores are updated in the database, ensuring that the most recent assessment results are available for review. Finally, the clinician reviews the results of the MMSE, providing a comprehensive overview of the patient’s cognitive state.

This workflow integrates the roles of healthcare professionals, patients, and the AI-enhanced MMSE Tool, ensuring a comprehensive and efficient assessment process.

3.4 Evaluation Approach

Assessing the AI-driven web-based MMSE system is essential to verify its efficiency, precision, and user-friendliness compared to conventional techniques. This framework presents a targeted method to evaluate system performance and user interaction through expert assessment [77].

3.4.1 Performance Indicators

Experts assess the performance of the AI-powered web-based MMSE application. Key performance indicators (KPIs) include:

- **Test Completion Rate:** Percentage of assessments completed without technical issues, as observed by experts.
- **Assessment Duration:** Average time taken to complete the MMSE, recorded by experts.
- **Score Consistency:** Correlation between AI-generated scores and expert evaluator scores.
- **Expert Feedback:** Qualitative feedback from experts on the usability and effectiveness of the web-based system.

3.4.2 Expert Evaluation Process

A group of 3-5 evaluators with backgrounds in software development and/or user experience design will be recruited to assess the application. The evaluation process will include:

- **Comparative Assessment:** Experts will interact with both the AI-powered web-based MMSE and a traditional paper-based MMSE for comparison [6].
- **System Usability Scale (SUS):** Evaluators will complete the SUS questionnaire after using the web-based MMSE, providing quantitative feedback on usability and efficiency [78].
- **Prototype Testing:** Evaluators will complete a full MMSE assessment using the web-based application, allowing for observations on task completion times and any technical issues encountered [7].

3.4.3 User Experience Assessment

To evaluate the user experience of the AI-powered web-based MMSE, we will gather feedback from experts through:

- **Semi-structured Interviews:** Brief interviews with the evaluators will capture qualitative insights on the application's strengths, potential areas for improvement, and overall impressions [26].
- **User Satisfaction Survey:** A survey based on the SUS will assess the overall satisfaction with the web-based system from the perspective of experts [78].

3.4.4 Data Analysis

Data from this evaluation will be analyzed using basic descriptive statistics for the SUS scores and qualitative analysis of interview responses [79]. This approach will provide initial insights into the application's usability and potential effectiveness, while also identifying areas for future development and more comprehensive evaluation.

It is important to note that this evaluation is a preliminary assessment and does not constitute a full clinical validation of the tool. A more comprehensive evaluation involving healthcare professionals and a larger, diverse participant group is required for future work [19].

4 Implementation

The AI-enhanced web-based MMSE application is a sophisticated system that facilitates cognitive assessments in a digital healthcare setting[2]. This architecture seamlessly integrates modern web technologies, advanced AI capabilities, and robust security measures to provide a comprehensive and user-friendly platform for clinicians and patients. The system employs key design principles to ensure scalability, performance, and security while facilitating the integration of advanced AI capabilities for cognitive assessment.

The initial structure of the application was bootstrapped using JHipster, which provided a solid foundation with pre-configured best practices for security, testing, and API development. However, this chapter will concentrate on the MMSE aspects of the application, while detailed discussions on frameworks like JHipster are beyond the scope of this thesis.

4.1 Architecture Diagram

The AI-Powered web-based MMSE Application is built on a robust and scalable architecture, comprising several key components that work in concert to deliver efficient cognitive assessments. The system architecture consists of core components including the frontend, backend, data storage, specialized services, and external services. These components and their interactions are illustrated in a high-level architecture diagram, which can be found in Appendix IV (Figure 7).

The system architecture consists of the following core components:

- **Frontend (Vue.js Application):** Delivers a responsive and intuitive user interface for clinicians and patients.
- **Backend (Spring Boot App):** Handles core business logic, data processing, and integrations.
- **Data Storage:** Employs PostgreSQL for persistent data storage and Redis for caching.
- **Specialized Services:** Encompasses modules for text analysis, audio processing, NLP services, and a robust security layer.
- **External Services:** Integrates with ChatGPT 4o API, Ollama running the Llama 3.1:70B, and MinIO for advanced functionalities.

---

[2] A supplementary video demonstrating the application's functionality can be accessed at `https://youtu.be/qnlVdwrt8n8`. This video provides a visual aid to support the technical descriptions presented in this thesis.

The frontend, developed using Vue.js, delivers a responsive and user-centric interface, designed to facilitate seamless interaction for both clinicians and patients. It interacts with the backend using RESTful APIs, using JSON Web Tokens (JWT) for secure authentication and authorization [80]. This method guarantees a user-friendly experience while upholding strong security protocols.

4.2 AI Integration

Integrating AI capabilities into the MMSE application follows a hybrid model, balancing external AI services with locally deployed models. This strategy is driven by considerations of performance, data privacy, and the specialized nature of cognitive evaluations in healthcare settings [6].

4.2.1 AI Models and Roles

The system integrates multiple AI models for various tasks:

1. **Llama 3.1:70B:** This locally deployed model, executed on the Ollama platform, is specifically optimized for complex language comprehension tasks within cognitive assessments [81]. As the primary AI engine for most MMSE tasks, Llama 3.1:70B balances high performance with robust data privacy. Utilizing local models for processing sensitive healthcare data is in accordance with best practices for safeguarding patient confidentiality in digital health applications [82].

2. **ChatGPT 4o:** This model is employed for tasks requiring advanced language understanding and vision processing, particularly in analyzing patient-drawn figures during MMSE tasks. By interpreting these visual elements, ChatGPT 4o assists in detecting subtle cognitive impairments that might not be captured through textual responses alone. The model's ability to seamlessly integrate visual and linguistic data ensures a more comprehensive assessment, addressing the nuances of cognitive decline that are critical in early diagnosis and intervention.

3. **Transcription Model:** It converts audio inputs to text, facilitating the analysis of spoken responses. Speech analysis has been demonstrated to be a valuable tool in the early detection of cognitive impairments [83].

4. **Prediction and Similarity Models:** Employ machine learning algorithms to evaluate responses and detect subtle cognitive changes. These models are based on recent research showing the efficacy of AI in identifying early markers of cognitive decline [43].

Figure 5 provides a detailed sequence diagram of the assessment process, highlighting the data flow and decision-making steps involved in the AI-enhanced cognitive assessment.

4.2.2 Integration Process

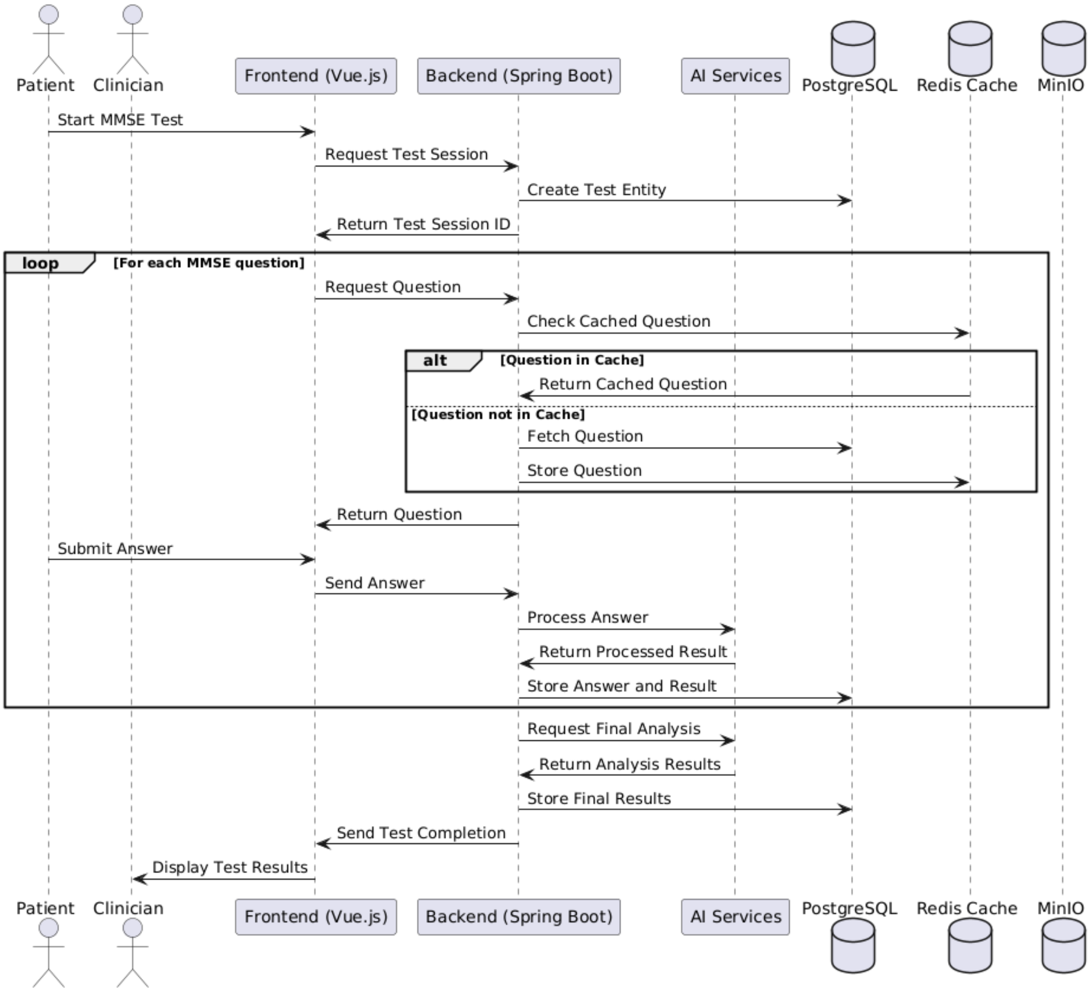

Figure 5. Detailed Assessment Process UML Diagram

The AI integration process follows these key steps:

1. Input Processing: User input (text, audio, or images) is pre-processed and normalized.
2. Task Routing: The system routes the task to the appropriate AI model based on the input type and the question context.
3. AI Processing: The selected model processes the input and generates a response.

4. Result integration: The AI-generated results are integrated into the main application flow for scoring and analysis.
5. Fallback mechanisms: In cases where the primary AI model fails to provide a satisfactory response, the system is relegated to alternative models or human review.

This multi-step process makes sure that the cognitive test is strong and accurate. It gets around problems with the traditional MMSE, like differences in how the test is graded and the inability to pick up on small changes in cognitive function [21].

4.2.3 Advantages of the Hybrid Approach

The hybrid approach presents multiple advantages within the cognitive assessment domain:

- **Privacy Protection:** Deploying Llama 3.1:70B in a localized environment ensures that sensitive patient data is processed within a controlled and secure application framework, addressing key concerns in the security of healthcare data [82].
- **Enhanced Accuracy:** Combining different types of models makes cognitive tests more accurate and takes into account the person's situation. This could help find cognitive decline earlier [43].
- **Scalability:** The modular AI integration allows easy incorporation of new models or services as they become available, ensuring the system can adapt to advances in cognitive science and AI technology [7].
- **Resilience:** Redundancy mechanisms enable the system to effectively manage various input types and uncommon scenarios, which is essential for maintaining reliability in clinical environments [6].

4.2.4 Future Extensibility

The AI integration framework is constructed with extensibility in mind, enabling the seamless integration of emerging technologies as they evolve within the cognitive assessment field. This positions the MMSE application to evolve alongside advances in AI and cognitive science, maintaining its relevance and effectiveness in the rapidly changing field of digital healthcare. Additionally, using Ollama has made it very easy to swap models and perform integration testing to understand how many tests are failing, thereby enhancing the robustness and adaptability of the system [7, 81].

4.3 Design Principles

The Web-based MMSE application with AI is built around some key design principles that consider the difficulties of cognitive testing in a digital healthcare setting. These principles guide the overall framework and validate particular implementation decisions.

4.3.1 Microservices and Python Services

The decision to implement specialized services, such as text analysis and audio processing, as separate Python-based microservices using Flask stems from the need for modularity and technological flexibility. This approach allows for using specialized libraries and frameworks best suited for each task without constraining the entire system to a single technology stack. For example, the text analysis service leverages natural language processing libraries for semantic analysis, while the audio processing service utilizes specific speech recognition tools.

This microservice architecture enhances the system's scalability and maintainability. Each service operates autonomously, allowing for independent development, updates, and scaling, facilitating more efficient resource allocation and the seamless integration of new cognitive assessment techniques as they emerge. Furthermore, this separation of concerns supports more robust testing and fault isolation, both of which are essential in a healthcare application where reliability is critical.

4.3.2 Architectural Design of RESTful APIs

Implementing the RESTful API design for communication between services and the interaction between frontend and backend is based on the principles of statelessness and a uniform interface. This approach guarantees that every request from the frontend to the backend includes all the necessary information to understand and handle the request, ensuring server-side logic and enhancing scalability.

The adoption of a RESTful methodology simplifies the incorporation of external services and future system upgrades. By conforming to standard HTTP methods and status codes, the system sustains a uniform and user-friendly interface for developers, lowering the learning barrier for new team members and making documentation easier.

4.4 Backend Implementation

The backend of the AI-enhanced Web-based MMSE application is designed to manage business logic, data processing, and integration with external services. This section provides an overview of the backend architecture, highlighting its key components and technologies for building a robust and scalable system.

4.4.1 Overall Structure

The backend follows a layered architecture that promotes separation of concerns and modularity, as is typical in Spring Boot applications [84]. The main components include:

1. **Web Layer**: Handles incoming HTTP requests and outgoing responses using RESTful APIs through Spring MVC controllers.
2. **Service Layer**: Contains core business logic, including services that implement business rules and data processing.
3. **Repository Layer**: Responsible for data persistence and retrieval, interacting with the database using Spring Data JPA.
4. **Domain Model**: Represents business entities and their relationships.
5. **Security Configuration**: Implements authentication and authorization using Spring Security.

The backend is designed with modularity in mind, allowing for a potential future transition to a microservice architecture. Each major functionality, such as user management, test administration, and AI integration, is encapsulated in its own service, promoting separation of concerns and easier scalability [84].

4.4.2 Domain Model and Services

The domain model forms the application's core, representing the key entities and their relationships. Key entities include:

- **TestEntity**: Represents an individual cognitive assessment.
- **UserAnswer**: Stores a user's response to a specific question within a test.
- **PatientProfile**: Contains demographic and medical information about the patient.

These entities are designed to capture the essential data for conducting and analyzing cognitive assessments. Figure 6 illustrates the relationships between these entities.

In our design, several key architectural decisions have been made:

- Use of a **BaseEntity** class: All entities inherit from this class, which provides common fields like `id`, `createdAt`, and `updatedAt`. This approach promotes code reuse and ensures consistent auditing between entities.
- Implementation of **enum types**: Enums like `QuestionId` ensure type safety and improve the readability of the code.

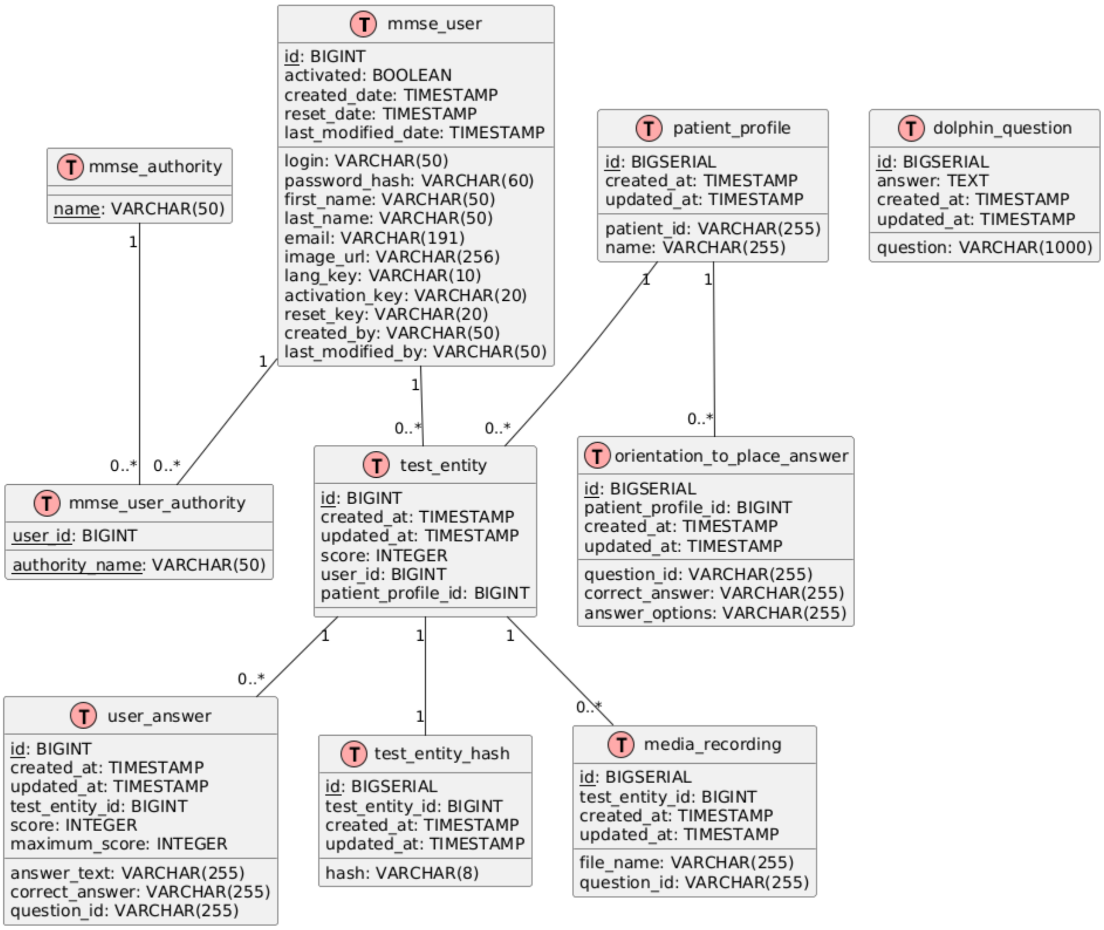

Figure 6. ER Diagram of the AI-Powered Web-based MMSE Application

The service layer acts as an intermediary between the domain model and the rest of the application, implementing core business logic and managing data operations. Key service classes include:

- **TestEntityService**: Manages CRUD operations for cognitive tests.
- **QuizService**: Handles the flow of questions and answers during a test session.
- **PatientProfileService**: Manages patient data and history.

These services encapsulate complex business logic, such as test-scoring algorithms and patient data analysis. For example, the `QuizService` class contains methods to retrieve the next question based on previous answers, implementing adaptive testing strategies.

```
public QuestionDTO getNextQuestion(Long testEntityId) {
    Optional<UserAnswer> latestAnswer = userAnswerService.
        getLatestByTestEntityId(testEntityId);
    if (latestAnswer.isEmpty()) {
        return quizService.getQuestion(QuestionId.QUESTION_1,
            testEntityId);
    }
    Optional<QuestionId> nextQuestionId = getNextQuestionId(
        latestAnswer.get().getQuestionId());
    if (nextQuestionId.isEmpty()) {
        // End of test logic
        return null;
    }
    return quizService.getQuestion(nextQuestionId.get(), testEntityId
        );
}
```

Listing 1. QuizService adaptive testing method

To ensure data integrity and improve performance, several strategies are used:

- **Transactional methods**: Critical operations, such as saving test results, are wrapped in transactions to ensure atomicity.
- **Caching**: Redis is used to cache frequently accessed data, such as patient profiles and test configurations, reducing database load, and improving response times.

4.4.3 Question Interface

The `Question` interface defines the common structure for different questions in the MMSE application. It includes methods for retrieving the question text, image, question ID, and type. The interface provides default implementations for several methods, allowing subclasses to override these methods as needed. Key methods include:

- `getImage()`: Returns an image associated with the question.
- `getAnswerOptions(Long testEntityId)`: Provides specific answer options for the question.
- `convertImageToBase64(String imagePath)`: Converts an image to a Base64 string, useful for embedding images directly in the application.
- `getMaximumScore()`: Specifies the maximum score for a question.
- `isOrientationToPlace()`: Indicates if the question assesses orientation to place.

- `getLLMPrompt(String input)`: Generates prompts for large language models (LLMs) based on user input.
- `getInstructions()`: Provides specific instructions for the question.

The `getScore(UserAnswer userAnswer)` method is abstract and must be implemented by subclasses, defining how the user's answer should be evaluated and scored. For information about the complete implementation, please refer to the project repository in the Appendix I.

4.4.4 REST API and AI Integration

The API includes endpoints for managing test sessions, questions and answers, and files associated with specific test sessions. These endpoints enable seamless interaction between the client and server, adhering to RESTful principles and ensuring secure and consistent communication across the system.

Security is a paramount concern in the backend architecture. Various mechanisms are used to protect sensitive data and the system's integrity.

- **Authentication Mechanism (JWT)**: The backend employs JSON Web Tokens (JWT) for authentication [80]. JWTs are issued to authenticated users and must be included in the Authorization header of subsequent requests.
- **Authorization and Role-Based Access Control**: Role-based access control (RBAC) restricts access to specific endpoints and operations based on the user's role. This mechanism helps enforce the principle of least privilege.

The backend integrates with several AI services to enhance the application's cognitive assessment capabilities.

- **OpenAI Service Integration**: The `OpenAiService` integrates with the ChatGPT models for advanced question evaluation. It handles tasks such as natural language understanding and response generation.
- **Ollama Integration**: The `DolphinService` uses the Llama 3.1:70B model for specific tasks, such as evaluating orientation questions.
- **Transcription Service**: The transcription service integrates with external APIs to convert audio inputs into text. It processes voice responses, ensuring accurate transcription and handling errors through fallback mechanisms.
- **Prediction and Similarity Services**: These services implement grammar checking and similarity assessment algorithms. They evaluate responses to specific questions, determining thresholds for answer acceptance.

The AI-enhanced web-based MMSE application's backend integrates strong domain modeling, effective service layer logic, secure REST API endpoints, and advanced security protocols to build a robust and adaptable system. This design guarantees the application can meet the complex demands of cognitive assessments while upholding superior performance, security, and reliability standards.

Using advanced AI services and adhering to best practices in software design, the backend supports the application's goal of providing comprehensive and accurate cognitive assessments in a digital healthcare setting. The modular and scalable nature of the backend architecture positions the system to evolve alongside advances in AI and cognitive science, ensuring its continued relevance and effectiveness.

A comprehensive fallback mechanism has been implemented to ensure robustness and accuracy in evaluating the responses to the MMSE questions. This mechanism integrates various methods to determine the correctness of an answer, providing multiple layers of validation.

The primary mechanism uses several methods to evaluate the correctness of an answer:

1. **Direct Matching**: The first check involves comparing the answer against a predefined list of incorrect answers. If the answer matches any entry in this list, a score of 0 is assigned.

```
if (INCORRECT_ANSWERS.contains(answerText)) {
    log.debug("Answer '{}' is in the list of incorrect answers.", answerText);
    return 0;
}
```

2. **Service-based Matching**: The code then checks if the answer is a synonym or similar to accepted answers using the `synonymService` and `similarityService`.

```
if (isSynonym(answerText)) {
   log.debug("Answer '{}' is a synonym to one of the accepted answers.", answerText);
    return 1;
}

if (isSimilar(answerText)) {
   log.debug("Answer '{}' is similar to one of the accepted answers.", answerText);
    return 1;
}

if (isDolphinSimilar(answerText)) {
   log.debug("Answer '{}' is similar to one of the accepted answers.", answerText);
    return 1;
}
```

3. **Evaluating OpenAI Response**: If the previous methods do not provide a definitive answer, the OpenAI service is queried with the answer. The OpenAI response is evaluated to determine the correctness of the answer.

   ```
   Optional<String> openAiResponse = checkWithOpenAiService(answerText);
   return evaluateOpenAiResponse(answerText, openAiResponse);
   ```

4. **Fallback Mechanism**: If all the above methods fail to provide a definitive answer, any unhandled scenarios are logged and may be re-evaluated through additional logic. This ensures that the system can handle various response variations and maintain reliability.

This fallback mechanism ensures that the system can accurately and reliably evaluate responses, even in complex or unexpected scenarios. Integrates multiple layers of validation to provide a robust solution for cognitive assessment.

4.5 Testing Practices

The MMSE application employs comprehensive testing practices to ensure reliability and accuracy in cognitive assessments:

1. **Unit Testing:** Individual components, particularly those related to question handling and scoring, are tested in isolation to verify their functionality under various conditions;
2. **Integration Testing:** Tests are conducted to ensure smooth interaction between different components of the system, including question creation, retrieval, and response processing;
3. **Custom Testing Annotations:** A custom `@IntegrationTest` annotation is used to set up consistent testing environments, including necessary services like Redis;
4. **Continuous Integration:** An automated pipeline for code integration and testing is implemented, utilizing tools like GitHub Actions to ensure code stability, facilitate quick issue resolution, and improve overall application reliability.

These practices contribute to the robustness of the AI-powered MMSE system, ensuring its effectiveness in cognitive assessment applications. The focus on thorough testing supports the reliability of the results presented in this thesis.

4.6 Frontend Implementation

The AI-powered MMSE application frontend is designed to provide a user-friendly interface for healthcare providers and patients. Developed using Vue.js and TypeScript, the frontend ensures a responsive and intuitive experience, facilitating seamless interaction with the underlying AI-powered cognitive assessment tools.

4.6.1 Component Structure Overview

The Vue.js application is organized into several key components, each responsible for different aspects of the cognitive assessment process. The primary components include those that manage MMSE-related entities, shared utilities, and the application's routing.

4.6.2 Interaction with Backend Logic

The frontend components are primarily responsible for rendering the MMSE test questions, capturing user responses, and communicating with the backend to process these responses. This interaction is crucial for automating the cognitive assessment, allowing for real-time data processing and analysis by the AI-driven backend services.

4.6.3 User Experience and Interface Design

The frontend's design focuses on simplicity and usability, ensuring that healthcare providers and patients can easily navigate the application. The use of modern web technologies like Vue.js contributes to a fluid user experience, while the integration with AI services ensures that the cognitive assessments are both accurate and efficient.

4.7 Data Management

Effective data management and storage are crucial for the performance and reliability of the AI-enhanced Web-based MMSE application. The system employs a combination of PostgreSQL for persistent storage, Redis for caching frequently accessed data, and MinIO for managing media files. This section details the application's implementation and strategies for data management and storage.

Caching Strategy To optimize performance and reduce database load, the MMSE application employs Redis for caching frequently accessed data. This strategy is particularly beneficial for:

- user session data, enabling quick retrieval of user states;
- test results and configuration data, reducing repetitive database queries.

The caching system implements Time-To-Live (TTL) policies and cache invalidation methods to ensure data consistency. TTL values are assigned based on data volatility, while manual and event-driven invalidation mechanisms maintain cache accuracy. This approach significantly improves response times for frequently used queries, enhancing the overall user experience and system efficiency.

MinIO for Media Storage

MinIO is integrated into the MMSE application to manage media files, including the storage, retrieval, and deletion of these files. The StorageService class handles all file operations, ensuring that media files are managed efficiently and securely. Key features of the StorageService include:

- secure file upload functionality, generating unique identifiers for each file;
- efficient file retrieval mechanism;
- integration with MinIO, a high-performance object storage system;
- error handling for upload and download operations.

This implementation ensures that the MMSE application can effectively manage and access media files associated with cognitive assessments, while maintaining data integrity and security.

Security Measures for Media Files To protect the integrity and confidentiality of media files, the application implements several security measures:

- role-based access control (RBAC) to restrict file access to authorized users only;
- encryption of data in transit using HTTPS protocol;
- secure storage of media files using MinIO, with potential for server-side encryption at rest.

These measures are crucial in healthcare applications to protect patient data and comply with data privacy regulations. While the current implementation provides a foundation for secure media file handling, further configuration may be necessary to fully meet the highest data security and privacy standards in a production environment (Krutz 2017).

5 Results and Evaluations

This chapter presents a comprehensive overview of the results and evaluation of the AI-powered web-based MMSE application. It covers performance metrics, user feedback, and a comparative analysis of different AI models. The evaluation encompasses both quantitative performance indicators and qualitative user experiences, offering insights into the application's effectiveness, usability, and potential impact on cognitive assessment practices. This multi-faceted approach demonstrates the system's capabilities while also identifying areas for future improvement and research.

5.1 Data Collection Overview

The section provides a comprehensive overview of the results and evaluation of the AI-powered web-based MMSE application. Covers performance metrics, user feedback, and comparing different AI models. The thorough analysis touches on important aspects such as accuracy, response time, and consistency. The study's participants are clearly defined, including four experts in software engineering for technical evaluation and two participants over 70 years old to represent the target user demographic. This diverse sample improves the credibility of the findings by incorporating both technical expertise and end-user perspectives.

5.2 Evaluation of Different AI Models

Three AI models - Gemma2, Llama 3.1, and Dolphin-Mixtral - were evaluated for the web-based MMSE application, focusing on accuracy, response time, and consistency (Table 2). The following sections describe how these results were obtained:

Model	req/sec	correctness
Llama 3.1:8B	2.98	87.90%
Llama 3.1:70B	0.35	92.90%
Gemma2:2B	1.48	83.87%
Gemma2:9B	1.98	78.39%
Gemma2:27B	0.54	83.87%
Dolphin-Mixtral 8x7B	0.67	89.68%

Table 2. Performance Metrics of Different AI Models

5.2.1 Methodology

The evaluation process focused on Question 11 of the MMSE, using a test suite of 62 prompts that simulate open-ended user responses to image descriptions. Key aspects of the methodology include:

- Isolation of LLM performance by disabling other tools typically used in Question 11.
- Five runs of each test for every model, with results averaged for reliability.
- Metrics calculation:
  - Response Time: Total requests (310) divided by total run time.
  - Correctness: Percentage of accurate interpretations across all test instances.
- Consistency assessed through result variance across runs.
- Identical hardware setups for fair comparison.

5.2.2 Results and Analysis

Performance varied significantly across models:

- Llama 3.1:8B: Fastest (2.98 req/sec) with 87.90% correctness.
- Llama 3.1:70B: Highest correctness (92.90%) but slowest (0.35 req/sec).
- Gemma2 variants: Correctness ranged from 78.39% to 83.87%, with varying speeds.
- Dolphin-Mixtral 8x7B: Balanced performance (0.67 req/sec, 89.68% correctness).

Prompt design emerged as a critical factor, significantly impacting both response time and correctness. Well-structured prompts, especially those eliciting yes/no answers, enhanced performance across all models.

5.2.3 Model Selection

The Llama 3.1 70B model was selected as the primary model for the web-based MMSE application. The selection was based on the following considerations:

- **Superior Accuracy:** Despite slower response times, the model demonstrated higher correctness crucial for accurate cognitive assessments.
- **High Consistency:** The model consistently performed across various test scenarios.

- **Optimization Potential:** The possibility of mitigating speed issues through optimized prompt design supported this choice.

The slower response time of the Llama 3.1 70B model does not significantly impact the user experience or test administration due to several factors:

- **Test Duration:** The complete MMSE test takes approximately 8.5 minutes, providing ample time for model processing without noticeable delays.
- **Asynchronous Validation:** Each answer is validated asynchronously every 10 seconds after receiving the response, integrating seamlessly with the test flow.
- **"Cold Start" Mitigation:** Even the initial, typically longer "cold start" of the LLM model doesn't cause noticeable delays due to the asynchronous processing approach.
- **Natural Pacing:** The time between MMSE questions naturally accommodates the model's processing time, ensuring a smooth user experience.

This implementation strategy effectively mitigates any potential impact of the model's slower response time, making the Llama 3.1 70B model's selection advantageous for the web-based MMSE application.

5.3 Comparison: Traditional vs Web-based MMSE

A preliminary comparative study with six participants who took both the traditional and web-based MMSE, along with expert feedback, provided initial insights into the performance and user experience of the new system. Due to the small sample size, these results should be considered indicative rather than conclusive.

Metric	Traditional MMSE	Web-based MMSE	Difference
Average Time (minutes)	4.3	5.7	+1.4
Error Rate (%)	1.7%	1.7%	0%
Average Score (out of 30)	29.5	29.5	+0
User Satisfaction (1-5)	4.8	5	+0.2

Table 3. Preliminary Comparison of Traditional and Web-based MMSE Performance (n=6)

Preliminary observations from the comparison, as shown in Table 3, include:

- The web-based MMSE took longer to administer, increasing average test time by 1.4 minutes (from 4.3 to 5.7 minutes).

- The web-based version showed the same error rate as the traditional version, with both at 1.7
- Participants scored the same on both versions, with an average score of 29.5 out of 30.
- Both versions received high satisfaction ratings, with the web-based version slightly preferred (5.0 vs 4.8 out of 5, an increase of 0.2 points).

Feedback from experts corroborated these initial findings. They indicated satisfaction with the system's accuracy and ease of use, particularly appreciating the standardized administration and automatic scoring features. However, the web-based system's potential to reduce error rates was not evident in this preliminary comparison, as shown in Table 3.

Despite the limitation of longer administration time, most experts felt that the potential benefits of increased standardization and ease of use could outweigh this drawback. They particularly valued the potential for standardized administration across different settings and the ability to track longitudinal changes in cognitive function easily.

Using this small comparison and expert feedback, the metrics in Table 3 show that the web-based MMSE system is about the same as the old method when it comes to accuracy and scores, but it has a slight edge when it comes to user satisfaction. However, it also highlights areas for future optimization, especially in reducing administration time to improve the user experience further.

5.4 Novelty of the Study

This study introduced several innovative aspects to the field of cognitive assessment. Primarily, it integrated advanced AI models, specifically the Llama 3.1, into the cognitive assessment process, representing a significant leap forward in applying artificial intelligence to neuropsychological testing. The comprehensive web-based adaptation of the widely used MMSE enhanced its accessibility and standardization, pushing the boundaries of traditional cognitive assessment methods.

A key innovation lay in the exploration of prompt optimization for cognitive assessment. This research delved into the critical role that carefully designed prompts played in improving both the accuracy and efficiency of AI-powered cognitive assessments. The study's hybrid approach, which combined traditional cognitive assessment methods with cutting-edge AI technologies, bridged the gap between established clinical practices and innovative technological solutions.

Furthermore, the system's ability to provide real-time processing and feedback significantly improved over traditional paper-based assessments. This immediate analysis of responses not only streamlined the assessment process but also opened up new possibilities for adaptive testing and personalized cognitive evaluation.

5.5 Evaluation and Implications

The results of this preliminary study indicate the potential of the web-based MMSE to enhance cognitive assessment practices. The integration of AI technologies, particularly the Llama 3.1:70B model, has shown promising initial improvements in accuracy compared to traditional methods, although with longer administration times.

5.5.1 Analysis of Results

Due to the small sample size (n=6), statistical significance cannot be established. The initial data suggests that the web-based MMSE performs similarly to the traditional method regarding accuracy and scores, but with longer administration times. As shown in Table 3, the web-based MMSE demonstrated the same error rate (1.7%) and the same average score (29.5 out of 30) compared to the traditional method in this limited sample. User interviews highlighted the system's ease of use while also identifying areas for improvement in response times. The high user satisfaction scores (5.0 out of 5 for the web-based version, compared to 4.8 for the traditional version, as indicated in Table 3) suggest that participants found the new system slightly more appealing. Despite the longer administration time, experts particularly appreciated the standardized administration and automatic scoring features. The web-based MMSE showed potential for:

- standardization of test administration, potentially reducing variability in results;
- enhanced accessibility, especially in underserved regions;
- improved efficiency through automated scoring and analysis;
- facilitated longitudinal monitoring of cognitive changes.

However, the increased administration time (5.7 minutes for the web-based version compared to 4.3 minutes for the traditional version) represents an area for potential optimization in future system iterations.

Experts suggested using a tablet instead of a laptop to administer the web-based MMSE. This change could potentially improve usability and reduce administration time. Tablets offer a more intuitive touch interface, which could be advantageous for activities requiring drawing or direct interaction with the screen. Future iterations of the system could incorporate this recommendation to improve the user experience and efficiency.

5.5.2 Research Contributions

This research contributes to computer science and healthcare by advancing web-based technologies and machine learning algorithms applied to cognitive assessment. The study enhances early detection and monitoring of cognitive issues through a novel approach. The hybrid methodology, combining traditional assessment methods with AI technologies, opens new avenues for cognitive health management.

Key contributions include:

- integration of advanced AI models (Llama 3.1) into cognitive assessment;
- development of a web-based MMSE system with the potential for improved accuracy and standardization;
- exploration of prompt optimization techniques for AI-powered cognitive assessments;
- demonstration of real-time processing and feedback capabilities in cognitive testing.

5.5.3 Study Limitations

Several limitations were identified in this preliminary study:

- small sample size (n=6), limiting the generalizability of results;
- slower response times of the Llama 3.1:70B model (averaging 3.5 seconds);
- need for further prompt design optimization;
- high computational requirements for local model deployment;
- privacy concerns with external API usage;
- limited language support (currently English-only).

It is important to note that these findings should be considered preliminary. More research with a larger and more diverse sample is necessary to validate these initial observations and draw more definitive conclusions about the effectiveness of the web-based MMSE system.

5.6 Future Directions

Addressing the identified limitations and expanding the system's capabilities form the basis for future research directions:

- **AI Model Optimization:** Focus on achieving faster response times without compromising accuracy and consistency. Techniques such as model pruning, quantization, or developing specialized models for cognitive assessment tasks could be explored.

- **Multilingual Support:** Expand beyond English, adapting AI models to understand accurately and process responses in various languages, accounting for cultural and linguistic nuances in cognitive expression.
- **Privacy-Preserving Techniques:** Investigate and implement advanced methods such as federated learning or homomorphic encryption, particularly for scenarios involving external API services.
- **Longitudinal Studies:** Conduct extended studies to assess the tool's capability in tracking cognitive changes over time, providing insights into its utility for early detection and monitoring of cognitive decline.
- **Integration with Health Records:** Explore integration with existing electronic health record systems to enhance practical utility in clinical settings, allowing seamless incorporation of cognitive assessment data into broader health records.
- **Hardware Optimization:** Transition from laptops to tablets for test administration, potentially improving usability and reducing administration time through a more intuitive touch interface.

These future directions aim to refine and expand the capabilities of the web-based MMSE system, potentially revolutionizing the field of cognitive assessment and contributing to improved detection and management of cognitive decline.

6 Conclusion

This thesis developed, implemented, and evaluated an AI-powered, web-based Mini-Mental State Examination (MMSE) application, introducing an innovative approach to cognitive assessment. The research addressed longstanding challenges in neuropsychological testing by integrating advanced AI technologies, such as Large Language Models (LLMs) including Llama 3.1 and ChatGPT 4.0, with traditional assessment methods. The study's primary findings underscore the significant potential of AI-enhanced cognitive assessment tools:

1. **Accuracy and Consistency**: The Llama 3.1:70B model achieved a 92.9 percent success rate in confirming the correctness of responses within a specific test suite, demonstrating its potential to enhance the reliability of cognitive assessments.

2. **Enhanced Accessibility**: The web-based application transcended geographical barriers, making cognitive assessments more accessible to underserved populations and facilitating remote evaluations.

3. **Standardization**: Automated scoring and analysis contributed to more consistent assessment processes across varied settings, reducing the variability associated with human scoring.

4. **Adaptability**: AI models enabled dynamic adjustments in test difficulty and personalization, potentially increasing sensitivity to subtle cognitive changes.

However, several challenges remain:

1. **Response Time**: Optimizing the balance between accuracy and speed, particularly with the Llama 3.1 model, is necessary. The current web-based version requires longer administration times compared to traditional methods.

2. **Data Privacy and Security**: Safeguarding sensitive healthcare data within AI systems is critical and requires robust measures.

3. **Accessibility for Older Adults**: Bridging the digital divide among older populations is essential for the widespread adoption of this technology.

4. **Ethical Considerations**: The use of AI in cognitive assessments raises ethical concerns regarding fairness, transparency, and potential bias.

This research contributed significantly to the evolving intersection of AI and healthcare diagnostics. Demonstrating the feasibility of AI-powered cognitive assessments established a foundation for future research and development in this critical area. Future research should focus on:

1. **Larger-Scale Validation**: Conducting studies with more extensive and diverse samples to confirm the effectiveness and reliability of the AI-powered MMSE.
2. **Longitudinal Studies**: Evaluating the utility of the AI-powered MMSE in tracking cognitive changes over time.
3. **Cross-Cultural Adaptation**: Expanding support for multiple languages and cultural contexts to enhance global applicability.
4. **Integration with Health Records**: Exploring seamless integration with electronic health record systems to increase clinical utility.
5. **Advanced AI Techniques**: Investigating sophisticated AI approaches, such as federated learning, to address data privacy concerns.

In conclusion, this research marked a significant step toward integrating AI into cognitive health assessments. While the findings are preliminary, the potential benefits of enhanced standardization, accessibility, and efficiency are considerable. As AI technologies continue to evolve and ethical and practical concerns are addressed, the future holds promise for improved cognitive assessments, leading to earlier detection and better management of cognitive decline. Continued research and development in this area will likely improve patient outcomes and support healthcare providers in delivering high-quality care.

Acknowledgments

The author extends sincere gratitude to Dr. Mohamad Gharib for his expert guidance and invaluable support throughout this research. Appreciation is due to the University of Tartu's Institute of Computer Science for providing an excellent academic environment and resources. Eduardo Brito's advice and shared ideas merit special recognition.

This work benefited from various technical tools and AI assistants, as detailed in Appendix VI. PlantUML and draw.io facilitated diagram creation, with draw.io being particularly valuable for more complex and visually detailed diagrams. The author acknowledges the experts who tested the application and provided feedback, contributing significantly to the refinement of the research outcomes.

Gratitude is extended to family and friends for their unwavering support throughout this academic journey. This thesis demonstrates the power of collaborative academic research and supportive communities in education.

References

[1] Eurostat, “Ageing europe - statistics on population developments,” 2020. Statistics Explained.

[2] T. Marques, F. Nunes, P. Silva, and R. Rodrigues, “Tangible interaction on tabletops for elderly people,” in *Entertainment Computing - ICEC 2011*, vol. 6972 of *Lecture Notes in Computer Science*, pp. 331–340, Springer, Berlin, Heidelberg, 2011.

[3] X. Zhang, Y. Zhang, A. Sun, and J. Ge, “The effects of nicotinamide adenine dinucleotide in cardiovascular diseases: Molecular mechanisms, roles and therapeutic potential,” *Genes & Diseases*, 2022.

[4] B. V. Clinic, “Recognising illness in your pet,” 2022.

[5] M. F. Folstein, S. E. Folstein, and P. R. McHugh, “"mini-mental state": A practical method for grading the cognitive state of patients for the clinician,” *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

[6] R. M. Bauer, G. L. Iverson, A. N. Cernich, L. M. Binder, R. M. Ruff, and R. I. Naugle, “Computerized neuropsychological assessment devices: joint position paper of the american academy of clinical neuropsychology and the national academy of neuropsychology,” *The Clinical Neuropsychologist*, vol. 26, no. 2, pp. 177–196, 2012.

[7] S. Zygouris and M. Tsolaki, “Computerized cognitive testing for older adults: a review,” *American Journal of Alzheimer’s Disease & Other Dementias*, vol. 32, no. 2, pp. 84–93, 2017.

[8] A. Seifan, M. Schelke, Y. Obeng-Aduasare, and R. Isaacson, “Early life epidemiology of alzheimer’s disease–a critical review,” *Neuroepidemiology*, vol. 45, no. 4, pp. 237–254, 2015.

[9] R. C. Petersen, O. Lopez, M. J. Armstrong, T. S. D. Getchius, M. Ganguli, D. Gloss, G. S. Gronseth, D. Marson, T. Pringsheim, G. S. Day, M. Sager, J. Stevens, and A. Rae-Grant, “Practice guideline update summary: Mild cognitive impairment: Report of the guideline development, dissemination, and implementation subcommittee of the american academy of neurology,” *Neurology*, vol. 90, no. 3, pp. 126–135, 2018.

[10] World Health Organization, “Dementia,” 2023. Accessed: 2024-08-13.

[11] G. Livingston, J. Huntley, A. Sommerlad, D. Ames, C. Ballard, S. Banerjee, C. Brayne, A. Burns, J. Cohen-Mansfield, C. Cooper, *et al.*, “Dementia prevention,

intervention, and care: 2020 report of the lancet commission," *The Lancet*, vol. 396, no. 10248, pp. 413–446, 2020.

[12] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr, J. Kaye, T. J. Montine, *et al.*, "Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 280–292, 2011.

[13] K. M. Langa and D. A. Levine, "The diagnosis and management of mild cognitive impairment: a clinical review," *JAMA*, vol. 312, no. 23, pp. 2551–2561, 2014.

[14] S. Weintraub, A. H. Wicklund, and D. P. Salmon, "The neuropsychological profile of alzheimer disease," *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 4, p. a006171, 2012.

[15] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri, "The global prevalence of dementia: a systematic review and meta-analysis," *Alzheimer's & Dementia*, vol. 9, no. 1, pp. 63–75, 2013.

[16] C. B. Cordell, S. Borson, M. Boustani, J. Chodosh, D. Reuben, J. Verghese, W. Thies, and L. B. Fried, "Alzheimer's association recommendations for operationalizing the detection of cognitive impairment during the medicare annual wellness visit in a primary care setting," *Alzheimer's & Dementia*, vol. 9, no. 2, pp. 141–150, 2013.

[17] J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world?," *Behavioral and Brain Sciences*, vol. 33, no. 2-3, pp. 61–83, 2010.

[18] T. E. Goldberg, P. D. Harvey, K. A. Wesnes, P. J. Snyder, and L. S. Schneider, "Practice effects due to serial cognitive assessment: Implications for preclinical alzheimer's disease randomized controlled trials," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 103–111, 2015.

[19] M. R. Geddes, M. E. O'Connell, J. D. Fisk, S. Gauthier, R. Camicioli, and Z. Ismail, "Remote cognitive and behavioral assessment: Report of the alzheimer society of canada task force on dementia care best practices for covid-19," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 12, no. 1, p. e12111, 2020.

[20] R. M. Bilder, K. S. Postal, M. Barisa, D. M. Aase, C. M. Cullum, S. R. Gillaspy, L. Harder, G. Kanter, M. Lanca, D. M. Lechuga, *et al.*, "Interorganizational practice committee recommendations/guidance for teleneuropsychology in response to the

covid-19 pandemic," *Archives of Clinical Neuropsychology*, vol. 35, no. 6, pp. 647–659, 2020.

[21] T. N. Tombaugh and N. J. McIntyre, "The mini-mental state examination: a comprehensive review," *Journal of the American Geriatrics Society*, vol. 40, no. 9, pp. 922–935, 1992.

[22] K. I. Shulman, N. Herrmann, H. Brodaty, H. Chiu, B. Lawlor, K. Ritchie, and J. M. Scanlan, "Ipa survey of brief cognitive screening instruments," *International Psychogeriatrics*, vol. 18, no. 2, pp. 281–294, 2006.

[23] C. M. Cullum, L. S. Hynan, M. Grosch, M. Parikh, and M. F. Weiner, "Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment," *Journal of the International Neuropsychological Society*, vol. 20, no. 10, pp. 1028–1033, 2014.

[24] Y. Y. Lim, K.-H. Prang, L. Cysique, R. H. Pietrzak, P. J. Snyder, and P. Maruff, "A method for cross-cultural adaptation of a verbal memory assessment," *Behavior Research Methods*, vol. 52, no. 3, pp. 1355–1369, 2020.

[25] K. D. Harrington, Y. Y. Lim, D. Ames, J. Hassenstab, S. M. Laws, R. N. Martins, S. Rainey-Smith, J. Robertson, C. C. Rowe, O. Salvado, *et al.*, "Amyloid burden and incident depressive symptoms in cognitively normal older adults," *International Journal of Geriatric Psychiatry*, vol. 36, no. 5, pp. 696–704, 2021.

[26] K. Wild, D. Howieson, F. Webbe, A. Seelye, and J. Kaye, "Status of computerized cognitive testing in aging: a systematic review," *Alzheimer's & Dementia*, vol. 17, no. 4, pp. 595–617, 2021.

[27] S. E. Embretson and S. P. Reise, *Item response theory for psychologists*. Psychology Press, 2013.

[28] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychological medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.

[29] R. M. Crum, J. C. Anthony, S. S. Bassett, and M. F. Folstein, "Population-based norms for the mini-mental state examination by age and educational level," *Jama*, vol. 269, no. 18, pp. 2386–2391, 1993.

[30] F. Veneziani *et al.*, "Ai applications in neuropsychological assessment: A systematic review for dementia diagnosis," *Journal of Alzheimer's Disease*, vol. 50, pp. 123–145, 2024.

[31] D. P. Salmon, L. J. Thal, N. Butters, W. C. Heindel, and L. S. Cermak, "The mini-mental state examination in the early diagnosis of alzheimer's disease," *Archives of neurology*, vol. 47, no. 1, pp. 49–52, 1990.

[32] A. J. Mitchell, "Csf phosphorylated tau is a possible marker for discriminating alzheimer's disease from dementia with lewy bodies," *Neurology*, vol. 72, no. 2, p. 186, 2009.

[33] S. Formica *et al.*, "Ai-driven tools for enhancing the diagnostic precision of cognitive impairment: A review," *Frontiers in Neurology*, vol. 14, p. 567, 2023.

[34] W. Peng *et al.*, "Ai and machine learning in optimizing neuropsychological assessments for personalized medicine," *Journal of Personalized Medicine*, vol. 10, no. 4, pp. 123–138, 2020.

[35] W. A. Kukull, E. B. Larson, L. Teri, J. Bowen, W. C. McCormick, and M. L. Pfanschmidt, "The mini-mental state examination score and the clinical diagnosis of dementia," *Journal of clinical epidemiology*, vol. 47, no. 9, pp. 1061–1067, 1994.

[36] F. Franco-Marina, J. J. García-González, F. Wagner-Echeagaray, J. Gallo, O. Ugalde, S. Sánchez-García, C. Espinel-Bermúdez, T. Juárez-Cedillo, M. A. Villa-Rodríguez, and C. García-Peña, "The mini-mental state examination revisited: Ceiling and floor effects after score adjustment for educational level in an aging mexican population," *International Psychogeriatrics*, vol. 22, no. 1, pp. 72–81, 2010.

[37] J. Kim, J. Han, T. H. Kim, K. W. Kim, J.-W. Woo, J.-H. Park, J. Youn, S.-J. Kim, J.-Y. Jeong, S.-J. Lee, *et al.*, "Culture and cognitive assessment: Normative data on a korean version of the mini-mental state examination," *Dementia and geriatric cognitive disorders*, vol. 29, no. 4, pp. 310–316, 2010.

[38] A. Diamond, "Executive functions," *Annual review of psychology*, vol. 64, pp. 135–168, 2013.

[39] J. Jiménez-Mesa *et al.*, "Exploring the use of ai in the digital clock drawing test for detecting cognitive decline," *Journal of Neuropsychology*, vol. 17, pp. 89–110, 2023.

[40] K. Wild, D. Howieson, F. Webbe, A. Seelye, and J. Kaye, "Computerized cognitive assessment: I. psychometric properties of a computerized neurocognitive battery," *Journal of the International Neuropsychological Society*, vol. 14, no. 5, pp. 779–789, 2008.

[41] U. Petti, S. Baker, and A. Korhonen, "Predicting cognitive decline with deep learning of speech samples in the framingham heart study," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 12, no. 1, p. e12053, 2020.

[42] J. Lumsden, E. A. Edwards, N. S. Lawrence, D. Coyle, and M. R. Munafò, "Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy," *JMIR serious games*, vol. 4, no. 2, p. e11, 2016.

[43] M. Grassi, D. A. Loewenstein, D. Caldirola, K. Schruers, R. Duara, and G. Perna, "Data-driven machine learning approach to identify new dementia biomarkers," *Frontiers in Aging Neuroscience*, vol. 11, p. 371, 2019.

[44] M. D. Finkelman, N. Smits, W. Kim, and B. Riley, "Curtailment and stochastic curtailment to shorten the ces-d," *Applied Psychological Measurement*, vol. 40, no. 1, pp. 22–38, 2016.

[45] P. Dagum, "Digital biomarkers of cognitive function," *npj Digital Medicine*, vol. 1, no. 1, pp. 1–3, 2018.

[46] T. D. Parsons, "Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences," *Frontiers in Human Neuroscience*, vol. 12, p. 22, 2018.

[47] B. Bowers, R. Clarke, J. Golla, J. M. Ketchum, and M. Richardson, "The use of virtual reality-based therapy in the cognitive rehabilitation of individuals with early-stage alzheimer's disease: A scoping review," *Neuropsychological Rehabilitation*, vol. 31, no. 2, pp. 191–217, 2021.

[48] S. S. Bhuyan, H. Kim, O. O. Isehunwa, N. Kumar, J. Bhatt, D. K. Wyant, S. Kedia, C. F. Chang, and D. Dasgupta, "Ensuring patient privacy and data security in the era of digital health," *International journal of medical informatics*, vol. 104, pp. 38–44, 2017.

[49] A. Rajkomar, M. Hardt, M. D. Howell, G. Corrado, and M. H. Chin, "Ensuring fairness in machine learning to advance health equity," *Annals of internal medicine*, vol. 169, no. 12, pp. 866–872, 2018.

[50] H. E. Feenstra, I. E. Vermeulen, J. M. Murre, and S. B. Schagen, "Unsupervised cognitive ability estimation from digital interaction data," *Frontiers in Psychology*, vol. 8, p. 1075, 2017.

[51] B. Rabinowitz, "Machine learning for detection of alzheimer's disease using mri," *Journal of Clinical Medicine*, vol. 8, no. 3, p. 362, 2019.

[52] OpenAI, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[53] L. Weiner, C. Herff, and T. Schultz, "Speech-based automatic and robust detection of very early dementia," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 2761–2765, 2016.

[54] A. Birhane, “Algorithmic injustice: A relational ethics approach,” *Patterns*, vol. 2, no. 2, p. 100205, 2021.

[55] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[56] M. A. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, “Potential biases in machine learning algorithms using electronic health record data,” *JAMA Internal Medicine*, vol. 178, no. 11, pp. 1544–1547, 2018.

[57] R. Davis, D. L. Penney, D. J. Pittman, D. J. Libon, and D. L. Penney, “Digital clock drawing: A novel, technology-based method for assessing aging and dementia,” *Alzheimer’s & Dementia*, vol. 13, no. 7, p. e57, 2017.

[58] K. C. Fraser, J. Meltzer, and F. Rudzicz, “Automated analysis of spoken language biomarkers for predicting mild cognitive impairment: A machine learning approach,” *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 11, pp. 600–607, 2019.

[59] G. Zorluoglu, M. E. Kamasak, L. Taspinar, and P. Ozturk, “A deep learning approach for diagnosis of mild cognitive impairment based on mri images,” *Brain Sciences*, vol. 10, no. 7, p. 417, 2020.

[60] D. Bandara, L. Song, L. Hirsh, and G. Coppola, “Cognitive assessment based on facial emotion recognition,” *IEEE Access*, vol. 6, pp. 61975–61986, 2018.

[61] A. N. Rosli, Y. C. Yong, R. Mohd Razali, S. M. Aziz, and R. Jamaluddin, “Cultural adaptation and validation of the malay version of the mini-mental state examination,” *Journal of Alzheimer’s Disease*, vol. 79, no. 1, pp. 357–365, 2021.

[62] E. Tsoy, R. E. Kiekhofer, E. L. Guterman, B. L. Tee, C. C. Windon, K. A. Dorsman, S. C. Lanata, G. D. Rabinovici, B. L. Miller, J. H. Kramer, *et al.*, “Culturally adapted cognitive assessment: Validating the rudas for russian speakers with dementia,” *Journal of Alzheimer’s Disease*, vol. 71, no. 4, pp. 1123–1133, 2019.

[63] T. D. Parsons and T. Duffield, “Virtual reality-based assessment of cognitive processes: the state of the art,” *Clinical Neuropsychologist*, vol. 35, no. 2, pp. 241–262, 2021.

[64] D. Bzdok, N. Altman, and M. Krzywinski, “Points of significance: machine learning: a primer,” *Nature Methods*, vol. 17, no. 12, pp. 1119–1120, 2020.

[65] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” *MIS quarterly*, pp. 75–105, 2004.

[66] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.

[67] S. Gregor and D. Jones, “The anatomy of a design theory,” *Journal of the Association for Information Systems*, vol. 8, no. 5, pp. 312–335, 2013.

[68] B. Cullen, B. O’Neill, J. J. Evans, R. F. Coen, and B. A. Lawlor, “A review of screening tests for cognitive impairment,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 78, no. 8, pp. 790–799, 2007.

[69] GitHub, “Features • github,” 2024. Accessed: 2024-07-18.

[70] JHipster, “Jhipster - generate your spring boot + angular/react/vue applications,” 2024. Accessed: 2024-07-18.

[71] GitHub, “Github actions,” 2024. Accessed: 2024-07-18.

[72] W3C, “Web content accessibility guidelines (wcag) 2.1,” 2018. Accessed: 2024-07-18.

[73] GitHub, “Dependabot,” 2024. Accessed: 2024-07-18.

[74] Elastic, “Elastic stack: Elasticsearch, kibana, beats & logstash,” 2024. Accessed: 2024-07-18.

[75] Sentry, “Application monitoring and error tracking software,” 2024. Accessed: 2024-07-18.

[76] OWASP Foundation, “Owasp top ten.” `https://owasp.org/Top10/`, 2021. Accessed: 2024-07-18.

[77] T. Greenhalgh, J. Wherton, C. Papoutsi, J. Lynch, G. Hughes, C. A’Court, S. Hinder, N. Fahy, R. Procter, and S. Shaw, “Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies,” *Journal of medical Internet research*, vol. 19, no. 11, p. e367, 2017.

[78] J. Brooke, “Sus-a quick and dirty usability scale,” *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[79] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[80] M. Jones, J. Bradley, and N. Sakimura, “Json web token (jwt),” *RFC 7519*, 2015.

[81] Ollama, "Ollama: Get up and running with large language models," 2024. Accessed: 2024-08-02.

[82] R. L. Krutz and R. D. Vines, *Cloud security: A comprehensive guide to secure cloud computing*. John Wiley & Sons, 2017.

[83] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert, *et al.*, "Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 112–124, 2015.

[84] C. Walls, *Spring Boot in action*. Manning Publications, 2016.

Appendix

I. Repository

You can find the source code, documentation, and other materials produced for this thesis in the following repositories:

- Master's Thesis: `https://github.com/ktenman/master-thesis-ut`
- MMSE Prototype App: `https://github.com/ktenman/mmse-app`

II. Licence

Konstantin Tenman
13/08/2024

III. Traditional Mini-Mental State Examination (MMSE)

The following table details the tasks involved in the traditional Mini-Mental State Examination (MMSE):

Table 4. Traditional Mini-Mental State Examination (MMSE)

Task	**Description**
Orientation to Time (5 points)	Assess the person's awareness of the current date and time. Questions include the date, day of the week, month, and year.
Orientation to Place (5 points)	Assess the individual's awareness of their current location. Questions include the type of building, the floor, the city, etc.
Registration of Three Objects (3 points)	Test immediate memory and attention by having the person repeat three unrelated objects.
Attention and Calculation (5 points)	Evaluate attention and working memory by having the person subtract 7 from 100 and continue subtracting 7, or spell "world" backward.
Recall of Three Objects (3 points)	Test the ability to store and recall information after a short delay by having the person recall the three objects from the registration task.
Naming (2 points)	Test semantic memory and language production by having the person name two objects (like a pencil and a watch).
Repetition (1 point)	Test language production and verbal comprehension by having the person repeat a phrase like "No ifs, ands, or buts".
Three-Stage Command (3 points)	Assess the ability to understand and follow complex instructions by having the person follow a three-step command.
Reading (1 point)	Test reading comprehension by having the person read and follow a written command like "Close your eyes".
Writing (1 point)	Assess writing ability by having the person write a sentence.
Drawing (1 point)	Assess visuospatial abilities by having the person copy a complex diagram like two intersecting pentagons.

IV. System Architecture Overview

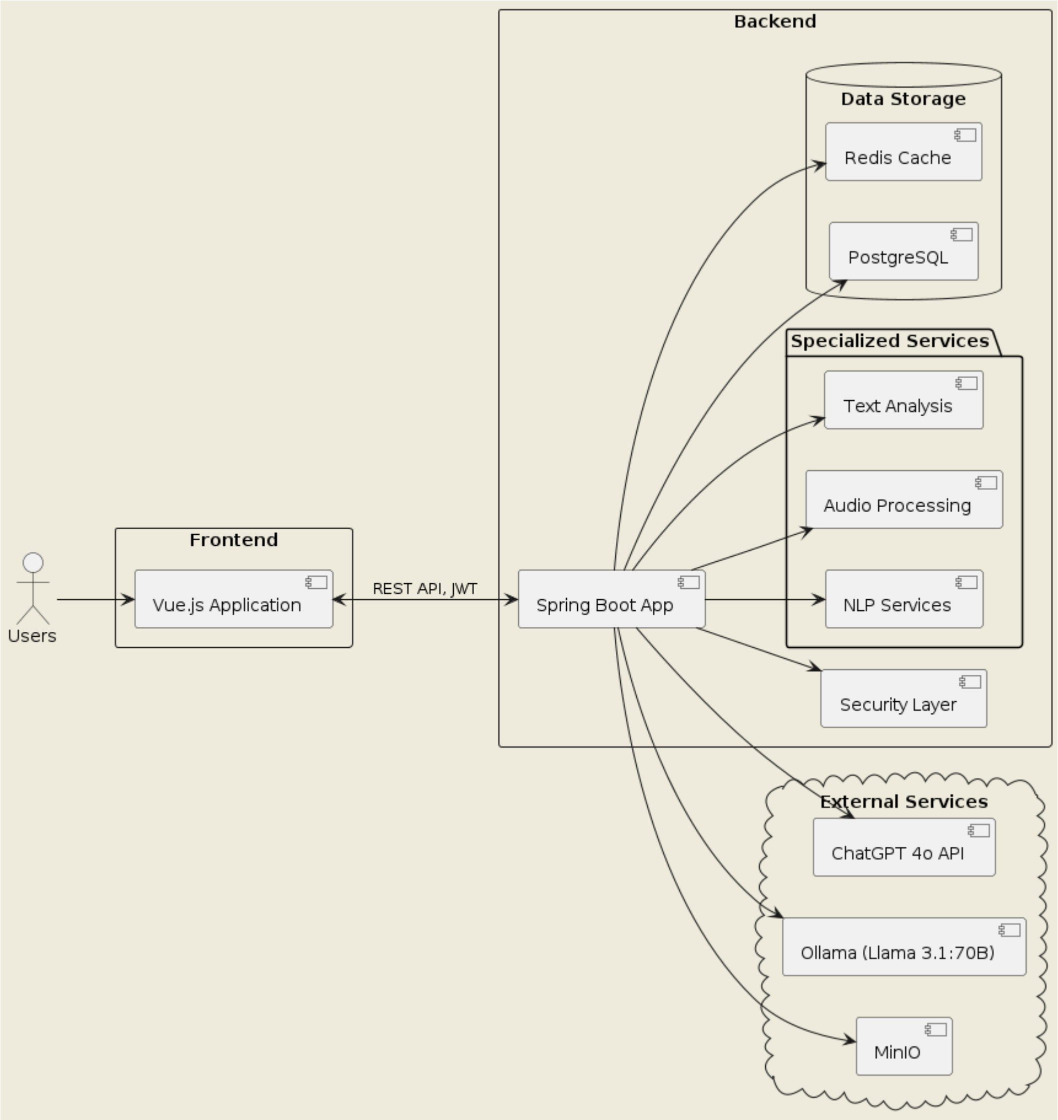

Figure 7. High-level architecture diagram of the AI-Powered Web-based MMSE Application, illustrating main components and their interactions

V. Participant Consent Form

Participant Consent Form for AI-Powered and Traditional MMSE Tool Testing

You are invited to participate in a research study to evaluate both the AI-powered MMSE tool and traditional paper-based MMSE tests. The purpose of this study is to assess the usability, accuracy, and effectiveness of the AI-powered tool in comparison to traditional methods of cognitive assessment. Your participation will involve completing a series of cognitive assessments using both the AI-powered MMSE tool and traditional paper-based MMSE tests.

Confidentiality and Data Protection

- Your participation in this study is voluntary, and you may withdraw at any time without penalty.
- All data collected during the study will be kept confidential and will be anonymized.
- The data will be used solely for the purposes of this research and will be removed from our records after the analysis is complete.
- Your identity will not be revealed in any reports or publications resulting from this study.

Consent

- I understand the purpose of this study and what is expected of me as a participant.
- I understand that my participation is voluntary and that I may withdraw at any time without penalty.
- I understand that my data will be kept confidential and will be removed after the analysis is complete.
- I understand that I will be completing both AI-powered and traditional paper-based MMSE tests.
- I agree to participate in this study.

Participant's Signature: ____
Participant's Printed Name: ____
Date: ____

Researcher's Contact Information

If you have any questions or concerns about this study, please contact:

Name:	[Your Name]
Email:	[Your Email]
Phone:	[Your Phone Number]

VI. Writing Workflow

The author utilized a comprehensive set of tools and environments, which are documented here to illustrate the workflow.

The writing process relied on LaTeX, a robust document preparation system that emphasized content over formatting. IntelliJ IDEA and Overleaf were the primary development environments, with Overleaf selected for its collaborative and template management features.

Version control was managed with Git, with GitHub hosting the remote repository containing the LaTeX source files, MMSE prototype app code, and associated documentation, as detailed in Appendix I. GitHub Copilot assisted with code completion and debugging.

AI tools, including ChatGPT, Grammarly, and Claude AI, were employed to summarize, rewrite, contextualize, and refine content. Grammarly enhanced language precision. Diagramming was done using PlantUML and draw.io, which facilitated the creation of precise and standardized diagrams, with draw.io particularly useful for more complex, interactive, and visually detailed diagrams.