# Assessing healthcare software built using IoT and LLM technologies

Gabriele De Vito*
gadevito@unisa.it
Department of Computer Science – University of Salerno
Fisciano, Salerno, Italy

## ABSTRACT

In the fast-paced world of healthcare technology, combining IoT devices with large language models (LLMs) offers a promising path to transform Clinical Decision-Support Systems (CDSS). This Ph.D. project is designed to tap into IoT's extensive data collection ability and LLMs' superior natural language processing skills. It aims to improve clinical decision-making and patient care through a sophisticated DSS that utilizes both technologies' strengths. The project delves into the software engineering challenges and methodologies required to build an effective DSS. It investigates how to smoothly evaluate and integrate IoT and LLMs into healthcare environments, tackling significant issues like data complexity, privacy concerns, and the necessity for high accuracy in medical settings. It underscores the critical role of thorough evaluation and assessment in developing healthcare technologies.

## CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**; **Natural language processing**;

## KEYWORDS

Healthcare Software Assessment, Large Language Models, Clinical Decision Support System

**ACM Reference Format:**

Gabriele De Vito. 2024. Assessing healthcare software built using IoT and LLM technologies. In *28th International Conference on Evaluation and Assessment in Software Engineering (EASE 2024), June 18–21, 2024, Salerno, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3661167.3661202

**Advisor:** Filomena Ferrucci, Department of Computer Science University of Salerno. *Email:* fferrucci@unisa.it

## 1 INTRODUCTION AND MOTIVATION

The healthcare industry constantly seeks innovative solutions to enhance patient care, improve health outcomes, and streamline healthcare processes. Clinical Decision Support Systems (CDSS) have become crucial in enhancing clinical practice and decision-making [27]. However, they encounter several challenges, such as integrating and analyzing unstructured data, navigating the complex healthcare environment, and providing real-time, accurate support that matches the latest medical guidelines and specific patient data. A significant aspect of these challenges is the depth of analysis current systems provide. For example, inadequate analysis, such as overlooking a patient's allergy history in free-text notes, can lead to inappropriate treatment suggestions [20]. These issues highlight the urgent need for improvements that blend smoothly with and enhance CDSS capabilities [27]. We can fully unlock CDSS's potential to support healthcare professionals effectively by overcoming these obstacles through empirical analysis and advanced software engineering methodologies.

Digital technologies, particularly the IoT and Large Language Models (LLMs), are at the forefront of these advancements. IoT technologies allow for real-time patient monitoring and data collection, offering information that could revolutionize clinical decision-making [10, 17, 28]. Meanwhile, LLMs, like OpenAI's GPT-3 [3], excel in processing and generating human-like text, making them ideal for interpreting complex medical data and aiding in diagnostics [8, 21]. However, integrating these technologies into healthcare has been challenging due to the need for high accuracy, the complexity of medical terminology, and privacy and security concerns [11, 12, 23, 31]. For instance, distinguishing between different types of diabetes and their associated complications is crucial for personalized treatment, yet current systems often fall short in capturing these distinctions [20]. Additionally, while IoT's use in healthcare is expanding, more work is needed to effectively merge the vast data it generates with LLMs for practical applications. Most current research focuses on these areas separately, highlighting a gap in integrating real-time health monitoring data with clinical decision-making [13, 14, 22, 30, 32]. The "Assessing healthcare software built using IoT and LLM technologies (HELIOT)" project introduces an innovative approach to overcome these obstacles. Distinctively, HELIOT merges real-time IoT data with the nuanced understanding capabilities of LLMs in a seamless fusion, addressing critical challenges head-on. This integrated approach promises to enhance CDSS by ensuring real-time processing and analysis of diverse data types, embedding advanced privacy-preserving mechanisms, and providing context-sensitive medical data interpretation. In particular, it seeks to enhance the patient anamnesis process through LLMs, making data collection more efficient and reducing the likelihood of overlooking crucial health information. By doing so, HELIOT aims to set a new benchmark for CDSS capabilities in the healthcare sector, overcoming prevalent issues such as data silos, delayed decision-making, and the underutilization of unstructured data.

The unique contributions of our project include:


*EASE 2024, June 18–21, 2024, Salerno, Italy*

ACM ISBN 979-8-4007-1701-7/24/06
https://doi.org/10.1145/3661167.3661202

- Development of a unique framework integrating IoT data with LLMs for real-time clinical decision support, featuring a robust evaluation methodology for healthcare effectiveness and reliability.
- The use of advanced LLM techniques to analyze unstructured healthcare data, we transform it into actionable insights for clinical decision-making, focusing on evaluating the accuracy and utility of these methods in handling complex medical information.
- Implementing a holistic data approach improves patient care personalization by customizing treatments with real-time and historical health data.
- Establishing rigorous evaluation and assessment protocols for the integrated IoT and LLM system, focusing on scalability, security, and compliance with healthcare standards.

## 2 RESEARCH QUESTIONS

This PhD project aims to answer the following research questions:

**RQ1 How can LLMs be effectively applied to analyze and interpret the vast amounts of unstructured healthcare data generated by IoT devices?**

This question explores the methodologies and techniques required to leverage LLMs for extracting meaningful insights from patient medical records, clinical notes, and other forms of unstructured data prevalent in healthcare settings.

- Metric 1.1: Precision and recall rates of extracted information.
- Metric 1.2: Time taken to analyze a set volume of data.

**RQ2 How can we develop and define an appropriate evaluation and assessment protocol to ensure the reliability of healthcare software systems integrating IoT and LLM technologies?**

This question aims to identify and establish a comprehensive protocol for evaluating and assessing the reliability of healthcare software systems that leverage the capabilities of both IoT devices and LLMs. The focus will be on identifying specific issues and challenges arising from this integration and proposing methodologies to address them effectively.

- Metric 2.1: Improvement in system reliability, measured by reduced system failures and errors in real-world applications.
- Metric 2.2: Enhancement in the effectiveness of healthcare software systems, measured by improved patient outcomes and satisfaction.

**RQ3 What role do evaluation and assessment play in ensuring the scalability and adaptability of healthcare software systems involving LLMs and IoT?**

This question emphasizes the need for healthcare software systems to be effective upon deployment, scalable, and adaptable to future healthcare challenges and technological advancements. It investigates how evaluation and assessment can facilitate these critical aspects.

- Metric 3.1: Degree of scalability, measured by the system's ability to handle increased loads and integrate additional IoT devices without significant performance degradation.
- Metric 3.2: Degree of adaptability, measured by the system's ability to incorporate new healthcare protocols, LLM updates, and respond to emerging healthcare needs.

**RQ4 How can a DSS, powered by LLMs and IoT data, enhance clinical decision-making processes?**

This question aims to investigate the potential impact of the proposed DSS on the accuracy, efficiency, and safety of clinical decisions, particularly in medication prescribing and patient anamnesis.

- Metric 4.1: Improvement in the accuracy of clinical decisions, measured by reduced error rates.
- Metric 4.2: Reduction in time taken for decision-making processes.

**RQ5 How can integrating LLMs and IoT technologies improve patient outcomes and healthcare delivery efficiency?**

By exploring this question, the project intends to assess the broader implications of its technological innovations on patient care quality, treatment outcomes, and the overall efficiency of healthcare services.

- Metric 5.1: Increase in the percentage of accurately captured medication histories and patient health information.
- Metric 5.2: Reduction in missing or incorrect data points in patient histories.
- Metric 5.3: Reduction in the incidence of adverse drug reactions and interactions.
- Metric 5.4: Increase in the percentage of correct pharmacological decisions supported by the system
- Metric 5.5: Reduction in the time required to complete anamnesis and pharmacological assessments.
- Metric 5.6: Increase in the speed of data processing and analysis for patient histories and medication information

*Hypotheses.* We formalize the research question with the following hypotheses:

**H1 Optimization of Healthcare Software Systems through Advanced Technologies and Methodologies**
Our hypothesis indicates that using LLMs to analyze unstructured healthcare data from IoT devices, alongside thorough evaluation, will notably enhance healthcare software systems' reliability, effectiveness, scalability, and adaptability. Expected benefits include improved data accuracy, reduced system failures, better patient outcomes and satisfaction, and an increased ability to handle larger loads and integrate new protocols seamlessly.

**H2 Impact of Integrated Technologies on Clinical Decision-Making and Healthcare Delivery**
Our hypothesis posits that a DSS utilizing LLMs and IoT data will boost clinical decision-making and healthcare efficiency, reducing clinical errors, faster and more accurate decisions, improved medication tracking, fewer adverse drug reactions, and quicker patient data processing, enhancing patient outcomes.

## 3 METHODOLOGY AND APPROACH

This section outlines the project's methodology. It starts by reporting the state-of-the-art and delves into work packages, data collection and analysis methods, and evaluation and validation approaches, emphasizing innovative strategies to meet its goals.

## 3.1 State of the Art

Integrating LLMs and IoT promises to leverage the vast data collection capabilities of IoT devices and the sophisticated natural language processing prowess of LLMs to significantly enhance healthcare delivery and patient outcomes. However, realizing this potential necessitates navigating a complex landscape of technological advancements, challenges, and ethical considerations, as outlined in the current state of the art.

**CDSS.** CDSS are pivotal in improving patient safety and clinical efficiency by seamlessly integrating with Electronic Medical Records (EMRs) [27]. Despite their considerable promise, the implementation of CDSS encounters numerous obstacles. Sutton et al. [27] highlight several of these challenges in their systematic literature review. They point out issues such as the difficulty in understanding decisions made by artificial intelligence, the struggle with limited data access, the problem of alert fatigue, and the potential for interference with established clinical practices. Additionally, the complexity of navigating through pharmacology and patient histories adds another layer of difficulty to deploying CDSS. In pharmacology, CDSS manage intricate drug interactions and tailor medication recommendations to individual patients. The sheer volume of possible drug interactions, influenced by unique patient factors such as genetics and lifestyle, demands sophisticated algorithms capable of detecting potential adverse reactions [25]. However, current systems often require substantial manual input to function effectively [18]. Additionally, the task of accurately capturing and interpreting patient histories presents its own set of challenges. This issue is due to the subjective nature of patient-reported information and a reliance on structured data formats, which may inadvertently omit critical details found in patient narratives [15]. These challenges highlight the urgent need for innovative CDSS solutions that integrate effortlessly with existing healthcare workflows, enhance clinical decision-making, and reduce the risks associated with alert fatigue and data fragmentation. Addressing these issues is crucial for developing CDSS to genuinely support healthcare professionals in delivering personalized, safe, and efficient care.

**LLMs in Healthcare.** LLMs have shown remarkable capabilities in generating human-like text, making them particularly suitable for interpreting complex medical data and aiding in diagnostic processes [29]. Their application in healthcare, ranging from analyzing patient records to generating medical reports and providing diagnostic recommendations [8, 21], highlights the potential of LLMs to revolutionize healthcare delivery. However, the deployment of LLMs within healthcare is fraught with challenges, including the need for precision, the complexity of medical terminologies, and significant privacy and security concerns [11, 12, 23, 31]. Moreover, Balch et al. [1] point out the challenges related to the absence of techniques to efficiently analyze medical unstructured data, the need to retrain models to align them with new medical information, the labor-intensive labeling process and the lack of validation on external datasets. These challenges point to the critical need for research and development efforts that can address the nuances of applying LLMs in healthcare, ensuring accuracy, security, and ethical compliance.

**IoT in Healthcare.** IoT technologies have similarly seen a surge in adoption for healthcare purposes, particularly for real-time patient monitoring [10, 17, 28]. While IoT devices can collect extensive data that could profoundly inform clinical decisions, effectively integrating this data with LLMs for healthcare applications remains an area that needs to be explored. In addition, Balch et al., in their systematic review [1], indicate that while some CDSSs utilize data from wearable devices, many existing systems lack comprehensive integration and real-time data processing capabilities. The challenges of interoperability, data processing, analysis, and ethical considerations related to patient data highlight a significant gap in the current research landscape [10, 17, 26, 28].

The HELIOT project is designed to address these gaps by developing a DSS that synergistically integrates IoT-generated data with the analytical capabilities of LLMs. This endeavor aims to enhance clinical decision-making and patient anamnesis processes and to navigate the ethical, privacy, and security considerations inherent in such technological integration. By addressing the challenges identified in the current state of the art, the HELIOT project positions itself as a critical initiative with the potential to drive forward the innovation in healthcare technology, offering novel solutions that are both technologically advanced and aligned with the ethical standards and needs of healthcare delivery.

## 3.2 Data Collection Approach

The project adopts a comprehensive strategy for gathering data, which is vital for creating and refining the proposed DSS. This strategy includes collecting quantitative and qualitative data to understand healthcare settings and patient needs fully. Our approach is informed by software engineering principles, emphasizing the importance of empirical data collection in developing robust healthcare solutions. We collaborate with healthcare facilities and access specialized databases to gather diverse data types.

We have formed partnerships with several Italian healthcare facilities, which are crucial to acquiring real-time data on patient monitoring and other critical healthcare information. This collaboration mainly provides quantitative data, accessible via API in the form of HL7 CDA2 files, collected via IoT devices, such as vital signs, medication doses, and other measurable factors. This data is crucial for training our models to identify patterns, forecast outcomes, and aid clinical decisions accurately.

Beyond quantitative data, the project strongly emphasizes gathering qualitative data to grasp the intricacies of patient care and clinical decision-making. This data includes patient medical records, clinical notes, and feedback from healthcare professionals obtained through our partnerships. Qualitative data adds depth to the quantitative data, offering insights into patient experiences, the reasoning behind treatments, and the nuances of medical practice that numbers alone cannot convey. For our qualitative data collection, we will utilize drug leaflets from regulatory agencies (e.g., AIFA [1]) and evidence-based guidelines, using freely accessible resources like NICE [2] and subscription-based services such as UpToDate [3]. These sources provide crucial information on drug use, side effects,

[1]https://www.aifa.gov.it/en/home
[2]https://www.nice.org.uk/
[3]https://www.wolterskluwer.com/en/solutions/uptodate/clinical-decision-support?redirect=true

and updated clinical guidelines, which are essential for informed prescribing decisions and understanding complex medical scenarios. By merging real-time patient data with detailed clinical and pharmaceutical research, we aim to develop a DSS that is accurate and sensitive to the complexities of healthcare delivery. This holistic data collection strategy ensures that the DSS can effectively guide healthcare providers in making informed, patient-focused decisions.

## 3.3 Data Analysis Methods and Techniques

In the project, we craft our data analysis methods and techniques to seamlessly blend IoT devices' expansive data collection capabilities with LLMs' understanding of human language. This endeavor requires a subtle approach to data analysis that transcends traditional boundaries, leveraging established statistical methodologies and the frontier of artificial intelligence research. The integration of these methodologies is a testament to the project's commitment to empirical analysis and software engineering excellence. At the core of our data analysis strategy is applying descriptive statistical methods to distill the vast, complex datasets collected from IoT devices into comprehensible insights. These methods involve aggregating data points and identifying patterns and trends to inform subsequent, more complicated analyses [24]. The descriptive analysis serves as the bedrock upon which predictive models are constructed. These models forecast patient outcomes by synthesizing real-time physiological data and historical health records. These predictive models are built using machine learning algorithms, trained and validated against a subset of our data to ensure they can accurately identify potential health events before they occur. Integrating LLMs introduces complexity to our data analysis efforts, particularly in NLP. The project leverages the latest advancements in NLP to parse, understand, and extract meaningful information from unstructured data in clinical notes and patient communications. This process is not merely about extracting information but understanding the context, the implied meanings, and the peculiarities of medical language. It involves sophisticated prompt engineering and fine-tuning techniques to convert narrative data into structured, actionable insights. A critical challenge we address is integrating heterogeneous data sources. The project employs advanced data integration techniques to ensure coherence and integrity across diverse datasets, including real-time data from IoT devices and historical patient records [2, 9, 16]. This integration is pivotal for constructing a holistic view of patient health, enabling our DSS to generate personalized, accurate, and timely recommendations. Evaluating the efficacy of our data analysis methods is conducted with a dual focus on the accuracy of our predictive models and the relevance and depth of insights generated through NLP. We employ a comprehensive suite of metrics designed to rigorously assess the performance of our analytical techniques, ensuring they meet the exacting standards required for clinical application. This evaluation process is underpinned by a commitment to ethical and privacy considerations, with stringent protocols in place to safeguard patient data. We use anonymization techniques, secure data handling practices, and adhere to legal and ethical standards governing patient data privacy.

## 3.4 Work Packages

The HELIOT project contains five main work packages (WP) described below.

**WP1: Systematic Literature Review.** This WP aims to comprehensively review the current state of research at the intersection of LLMs, IoT, and healthcare. It will identify gaps in the literature, potential applications of LLMs in healthcare IoT, and challenges in integrating these technologies. The findings will guide the project's subsequent phases.

**WP2: Experiments with LLM Techniques.** WP2 experiments with various LLM techniques, such as prompt engineering, fine-tuning, and Retrieval Augmentation Generation (RAG) [19], to optimize LLMs for healthcare data analysis. This package aims to determine the most effective strategies for leveraging LLMs to interpret complex medical information and generate actionable insights.

**WP3: DSS Development.** This WP involves the development of a DSS that utilizes LLMs to analyze patient medical records and IoT data. WP3 covers designing the system architecture, developing algorithms for data analysis, and integrating the system with existing healthcare IT infrastructure to aid in medication prescribing and patient anamnesis.

**WP4: Evaluation and Validation.** This WP is committed to rigorously evaluating the DSS's impact on clinical decision-making and healthcare delivery. It will use a detailed evaluation framework, focusing on the importance of thorough assessment in healthcare technology. The WP will conduct real-world trials, usability studies, and efficiency analyses to confirm the impact of the system's effectiveness, safety, and healthcare outcomes, showcasing the project's dedication to evidence-based healthcare technology development.

**WP5: Addressing Ethical, Privacy, and Security Considerations.** This WP investigates the ethical implications, privacy concerns, and security challenges associated with applying LLM technologies in healthcare. It aims to develop robust strategies for safeguarding patient information, ensuring compliance with legal and ethical standards, and addressing potential biases in LLM-generated insights.

*Methodology and Risk Management.* We adopt an iterative methodology, ensuring responsiveness to findings at each stage for more effective and reliable outcomes. Risks related to technology integration, data privacy, and ethical considerations are mitigated through careful planning, stakeholder engagement, and adherence to ethical guidelines and security standards.

## 3.5 Evaluation and Validation

The evaluation and validation processes are designed to ensure that the project's outcomes meet the research objectives. To this aim, we adopt a mixed-methods approach. We integrate quantitative and qualitative techniques to achieve a detailed understanding of the CDSS effectiveness and suitability for clinical use.

We employ quantitative empirical studies to address R1, R2, and RQ3, using different performance metrics tailored to assess specific aspects of the system. For RQ1, we measure accuracy, precision, recall, and F1 score to evaluate the system's effectiveness in processing and interpreting large volumes of data. In addressing RQ2, we focus on system reliability, quantifying error rates, system uptime, and failure frequencies to assess system robustness under various operational conditions. For RQ3, stress testing protocols assess the system's performance under increased loads. We examine scalability and adaptability through metrics like throughput, response time, and resource utilization at different load levels. These metrics provide insights into the system's capability to maintain performance under stress.

In addressing RQ4 and RQ5, we use qualitative empirical studies to delve deeper into the system's real-world application. For RQ4, simulations of clinical scenarios and subsequent detailed feedback via in-depth interviews and focus groups assess the CDSS's usability, relevance, and integration into clinical workflows. Observational studies enhance our understanding by showing how the system improves clinical decision-making. For RQ5, these observational studies extend to examine the impact on patient outcomes and healthcare delivery efficiency, providing practical insights into the CDSS's effectiveness in clinical settings. Before deploying the CDSS, physicians will be fully briefed on its purpose and operation, and their consent will be obtained for transparency and ethical adherence. Patient consent will also be secured according to institutional and legal standards, ensuring they understand their data's role in enhancing care. This process prioritizes participant rights and privacy, maintains trust and compliance with ethical and legal norms, and fosters ongoing communication to address any concerns about the CDSS and data use. Ethical and privacy issues are paramount, with strict adherence to data anonymization and protection laws, alongside efforts to identify and mitigate biases for equitable healthcare outcomes. Security assessments by cybersecurity experts aim to fortify the system against threats, enhancing data safety. A comprehensive report will conclude this phase, outlining findings, feedback from healthcare professionals, and recommendations for DSS improvement, focusing on accuracy, usability, and clinical relevance.

# 4 PRELIMINARY RESULTS

The initial phase of the HELIOT project has been dedicated to laying the groundwork through a systematic literature review (WP1) and conducting preliminary experiments with LLMs techniques (WP2), aiming to answer RQ1. These efforts have already begun to yield valuable insights and outcomes, contributing significantly to the project's overarching goals.

## 4.1 Systematic Literature Review (WP1)

WP1 has played a crucial role in exploring the integration of LLMs and IoT technologies within the healthcare sector. This comprehensive review aimed to uncover the potential applications, challenges, and future directions for LLMs and IoT to enhance healthcare delivery and patient outcomes. Key findings from the review include identifying transformative applications of LLMs in healthcare IoT, such as real-time patient monitoring and the automation of clinical documentation. However, it highlighted significant challenges, including the complexity of medical terminologies and data privacy concerns, alongside the technical hurdles of combining IoT-generated data with LLMs for actionable insights. Moreover, the insights from WP1 highlight the critical need for rigorous evaluation and assessment methodologies in healthcare software systems. By pinpointing the challenges and opportunities presented by LLMs and IoT in healthcare, the review sets the stage for developing frameworks that ensure these technologies are effectively assessed for their impact on healthcare delivery and patient outcomes. All this underscores the project's dedication to advancing healthcare technology through meticulous evaluation and assessment, ensuring that innovations are impactful and meet the sector's needs. These insights have been synthesized into a paper titled "The Role of Large Language Models in Addressing IoT Challenges: A Systematic Literature Review," prepared for submission.

## 4.2 Experiments with LLM Techniques (WP2)

WP2 has been marked by a series of experiments designed to explore the capabilities of LLMs in processing complex data, focusing on enhancing their functionality through various techniques. The experiments conducted have led to significant contributions, each aimed at addressing specific challenges and objectives within the project's broader goal. The ECHO approach [7] was pivotal in experimenting with prompt and co-prompt engineering techniques. By leveraging LLMs as prompt engineers, the approach sought to improve the prompts iteratively, guiding the LLM toward generating more accurate and relevant answers. This experiment underscored the potential of fine-tuning LLM interactions to meet specific user needs, demonstrating a novel way to enhance the utility of LLMs in understanding and responding to complex queries. Building on the insights from ECHO, the C4SE experiment [5] delved into the RAG technique, aiming to endow LLMs with dynamic memory capabilities. This approach facilitated the creation of smart agents based on LLMs capable of executing specific tasks. Given the initial prompt, using specific prompt engineering techniques, we can exploit the LLM to break down the request into atomic tasks to perform, activate the most suitable LLM-based agents for addressing these tasks, and then combine and synthesize the results to provide a comprehensive final response. This methodology showcased the ability of LLMs to handle multi-faceted queries by breaking them down into manageable tasks, thereby enhancing their applicability in complex problem-solving scenarios. On the other hand, ENEA [4] focused on developing a methodology capable of analyzing large volumes of data to extract meaningful knowledge. This experiment aimed to harness the power of LLMs in sifting through extensive datasets, identifying patterns, and deriving insights that could inform decision-making processes. ENEA highlighted the potential of LLMs in data analysis and knowledge extraction, providing a foundation for leveraging these models in data-intensive applications. Lastly, the COMET initiative [6] tested the application of few-shots learning techniques, exploring how these methods could generate appropriate responses to user requests. This experiment revealed how few-shot learning can enhance LLM responses for more personalized and accurate user interactions, highlighting its strengths and limitations. Together, these experiments have contributed to the academic discourse on applying LLMs in healthcare and laid the groundwork for their practical implementation, guiding the development of the DSS in WP3.

# 5 WORK PLAN

In the first year (2022-2023), we conducted a systematic literature review and initial experiments with LLMs to build a solid theoretical and experimental base. Presenting at conferences like the 49th Euromicro SEAA in Albania enabled community engagement and feedback, shaping our evaluation and assessment methods.

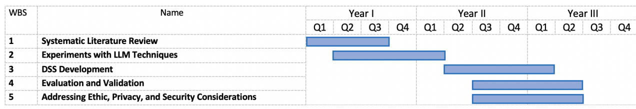


**Figure 1: Project GANTT**

As we move into the second year (2023-2024), our efforts will deepen in exploring LLM techniques, specifically targeting pre-training and fine-tuning to tailor these models for nuanced health-care data analysis. This phase is critical for ensuring that the models we develop are robust and capable of handling the complexities of healthcare data, a prerequisite for practical evaluation and assessment.

- **I Quarter**: Initiating further experiments on LLM techniques to refine the models for superior performance in healthcare contexts, setting the stage for rigorous evaluation.
- **II Quarter**: Transitioning to developing the DSS under WP3, focusing on system architecture and integration with health-care IT infrastructure, ensuring that the system is built with evaluation and assessment in mind from the ground up.
- **III Quarter**: Conducting initial evaluations of the DSS in real-world settings to assess its impact on clinical decision-making and patient outcomes, directly applying our commitment to thorough evaluation and assessment.
- **IV Quarter**: Preparing for an extensive evaluation of the DSS, which will rigorously test its effectiveness, usability, and impact, underscoring the project's dedication to evaluation and assessment.

The final year (2024-2025) will focus on finishing the DSS evaluation and tackling ethical, privacy, and security issues in WP5. This critical phase ensures the DSS meets technical requirements and the highest ethical and data protection standards necessary for our evaluation framework.

- **I Semester**: Continuing the evaluation of the DSS, focusing on its usability, effectiveness, and impact on healthcare delivery. Mid-year, we will delve into the ethical implications, privacy concerns, and security challenges, ensuring that our evaluation and assessment processes consider these critical dimensions.
- **II Semester**: Finalizing the doctoral thesis to encapsulate project findings, including evaluation and assessment details, conference presentations, and publications. This document will highlight contributions to healthcare technology and the significance of thorough evaluation and assessment for future implications.

## REFERENCES

[1] Jeremy A Balch and et al. 2023. Machine learning–enabled clinical information systems using fast healthcare interoperability resources data standards: scoping review. *JMIR Medical Informatics* 11 (2023), e48297.
[2] Kornelia Batko and Andrzej Ślęzak. 2022. The use of Big Data Analytics in healthcare. *Journal of big Data* 9, 1 (2022), 3.
[3] T. Brown, B. Mann, et al. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle et al. (Ed.), Vol. 33. Curran Associates, Inc., 1877–1901.
[4] Gabriele De Vito, Stefano Lambiase, Dario Di Nucci, Fabio Palomba, and Filomena Ferrucci. 2023. Efficient Knowledge Extraction Using Large Language Models and Vector Databases: An Integrated Approach.
[5] Gabriele De Vito, Stefano Lambiase, Fabio Palomba, and Filomena Ferrucci. 2023. Meet C4SE: Your New Collaborator for Software Engineering Tasks. In *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 235–238.
[6] Gabriele De Vito, Fabio Palomba, Carmine Gravino, Sergio Di Martino, and Filomena Ferrucci. 2023. Automatic COSMIC Measurement of Use Case Models Employing Large Language Models.
[7] Gabriele De Vito, Fabio Palomba, Carmine Gravino, Sergio Di Martino, and Filomena Ferrucci. 2023. Echo: An approach to enhance use case quality exploiting large language models. In *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 53–60.
[8] Zachary Englhardt et al. 2023. From Classification to Clinical Insights: Towards Analyzing and Reasoning About Mobile and Behavioral Health Data With Large Language Models. *arXiv preprint arXiv:2311.13063* (2023).
[9] AK Kock-Schoppenhauer et al. 2021. Medical Data Engineering–Theory and Practice. In *Advances in Model and Data Engineering in the Digitalization Era: MEDI 2021 International Workshops: DETECT, SIAS, CSMML, BIOC, HEDA, Tallinn, Estonia, June 21–23, 2021, Proceedings 10*. Springer, 269–284.
[10] M. Al-Khafajiy et al. 2019. Remote health monitoring of elderly through wearable sensors. *Multimedia Tools and Applications* 78, 17 (2019), 24681–24706.
[11] M. Cascella et al. 2023. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *Journal of medical systems* 47, 1 (2023), 33.
[12] R. Yang et al. 2023. Large language models in health care: Development, applications, and challenges. *Health Care Science* 2, 4 (2023), 255–263.
[13] Y. Wang et al. 2023. Aligning Large Language Models with Human: A Survey. arXiv:2307.12966 [cs.CL]
[14] Lizhou Fan and et al. 2023. A Bibliometric Review of Large Language Models Research from 2017 to 2023. arXiv:2304.02020 [cs.DL]
[15] Marta Fernandes and et al. 2020. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. *Artificial Intelligence in Medicine* 102 (2020), 101762.
[16] Panagiota Galetsi, Korina Katsaliaki, and Sameer Kumar. 2020. Big data analytics in health sector: Theoretical framework, techniques and prospects. *International Journal of Information Management* 50 (2020), 206–216.
[17] Bonnie Kaplan. 2020. Revisiting health information technology ethical, legal, and social issues and evaluation: telehealth/telemedicine and COVID-19. *Int. J. of medical informatics* 143 (2020), 104239.
[18] Laura Légat and et al. 2018. Clinical decision support systems for drug allergy checking: systematic review. *Journal of medical Internet research* 20, 9 (2018).
[19] Patrick Lewis et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020).
[20] Irene Li et al. 2022. Neural natural language processing for unstructured data in electronic health records: a review. *Computer Science Review* 46 (2022), 100511.
[21] Jing Miao et al. 2024. Chain of thought utilization in large language models and application in nephrology. *Medicina* 60, 1 (2024), 148.
[22] B. Min and et al. 2023. Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. *ACM Comput. Surv.* (jun 2023).
[23] Javad Pool and et. al. 2024. Large language models and generative AI in telehealth: a responsible use lens. *Journal of the American Medical Informatics Association* (2024), ocae035.
[24] Wullianallur Raghupathi and Viju Raghupathi. 2014. Big data analytics in health-care: promise and potential. *Health information science and systems* 2 (2014), 1–10.
[25] Romy Roy et al. 2022. Drug-food Interactions in the Era of Molecular Big Data, Machine Intelligence, and Personalized Health. *Recent Advances in Food, Nutrition & Agriculture* 13, 1 (2022), 27.
[26] S. Selvaraj and S. Sundaravaradhan. 2019. Challenges and opportunities in IoT healthcare systems: a systematic review. *SN Applied Sciences* (2019).
[27] Reed T Sutton and et al. 2020. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine* 3, 1 (2020), 17.
[28] Olutosin Taiwo and Absalom E Ezugwu. 2020. Smart healthcare support for remote patient monitoring during covid-19 quarantine. *Informatics in medicine unlocked* 20 (2020), 100428.
[29] Jason Wei et al. 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682* (2022).
[30] L. Wu and et al. 2023. A Survey on Large Language Models for Recommendation. arXiv:2305.19860 [cs.IR]
[31] Qianqian Xie and et al. 2023. Faithful AI in Medicine: A Systematic Review with Large Language Models and Beyond. *medRxiv* (2023), 2023.04.18.23288752.
[32] Wayne Xin Zhao and et al. 2023. A Survey of Large Language Models. arXiv:2303.18223 [cs.CL]

Received 15 March 2024