EXPLORING THE POTENTIAL OF LARGE LANGUAGE MODELS IN ADDRESSING CHALLENGES OF HOSPITAL INFORMATION SYSTEMS

TEPASIT PONGSABUTR

A THEMATIC PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF MANAGEMENT COLLEGE OF MANAGEMENT MAHIDOL UNIVERSITY 2024

Thematic paper
entitled

EXPLORING THE POTENTIAL OF LARGE LANGUAGE MODELS IN ADDRESSING CHALLENGES OF HOSPITAL INFORMATION SYSTEMS

was submitted to the College of Management, Mahidol University
for the degree of Master of Management
on
25 August 2024

.....
Mr. Tepasit Pongsabutr
Candidate

.....
Assoc. Prof. Nathasit Gerdsri,
Ph.D.
Advisor

.....
Assoc. Prof. Winai Wongsurawat,
Ph.D.
Chairperson

.....
Assoc. Prof. Vichita Ractham,
Ph.D.
Acting Dean
College of Management
Mahidol University

.....
Prof. Kittisak Jermsittiparsert,
Ph.D.
Committee member

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to Associate Professor Nathasit Gerdsri, my thematic advisor, for his invaluable guidance and support throughout the drafting process. His insightful and constructive recommendations have been instrumental in shaping this project.

I am also profoundly thankful to my family and friends for their unwavering support, ideas, and encouragement. Their presence has been a constant source of strength for me. Additionally, I sincerely appreciate all of my participants, colleagues, and friends who generously agreed to be my interviewees. Their contributions have provided me with unique and insightful perspectives that were crucial to the completion of this research.

Finally, I would like to acknowledge the College of Management at Mahidol University (CMMU) for providing an enriching academic environment and extend my gratitude to all the dedicated faculty members at CMMU who have contributed to my educational experience. The knowledge, skills, and mentorship I received have been invaluable for both this research and my professional growth.

Tepasit Pongsabutr

EXPLORING THE POTENTIAL OF LARGE LANGUAGE MODELS IN ADDRESSING CHALLENGES OF HOSPITAL INFORMATION SYSTEMS

TEPASIT PONGSABUTR 6549133

M.M. (HEALTHCARE AND WELLNESS MANAGEMENT)

THEMATIC PAPER ADVISORY COMMITTEE: ASSOC. PROF. NATHASIT GERDSRI, Ph.D., ASSOC. PROF. WINAI WONGSURAWAT, Ph.D., PROF. KITTISAK JERMSITTIPARSERT, Ph.D.

ABSTRACT

This study explores the potential of Large Language Models (LLMs) in addressing challenges within Hospital Information Systems (HIS). Through literature review, interviews with healthcare professionals, and proof-of-concept testing, the research identifies key pain points in current HIS and evaluates LLMs' capabilities in areas such as data summarization, medical coding, error prevention, and interoperability. The study demonstrates LLMs' promising applications in enhancing data retrieval, improving coding accuracy, and facilitating standardized data exchange. However, limitations in consistency, knowledge updating, and data privacy are noted. Recommendations for healthcare organizations and developers are provided, emphasizing the need for continued research and development to fully realize LLMs' potential in improving healthcare delivery and patient outcomes.

KEY WORDS: HOSPITAL INFORMATION SYSTEMS (HIS)/ LARGE LANGUAGE MODELS (LLMS)/ HEALTHCARE DATA MANAGEMENT/ INTEROPERABILITY

68 pages

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Background

Hospital Information Systems (HIS) are an integral part of modern healthcare, serving as the backbone for managing vast amounts of patient data. HIS functions similarly to an Enterprise Resource Planning (ERP) system within a hospital, enabling healthcare providers to streamline and coordinate various administrative and clinical operations. The primary objective of HIS is to enhance the efficiency and effectiveness of healthcare delivery by integrating patient information, improving data accessibility, and supporting decision-making processes.

Despite the numerous benefits HIS offers, there are significant challenges associated with its implementation and utilization. One major issue is the lack of standardized data collection leading to fragmented and unstructured patient records. This fragmentation hampers efficient data retrieval and exchange, complicating efforts to provide cohesive and continuous patient care. Additionally, the presence of multiple vendors providing diverse HIS solutions further exacerbates the interoperability issues, making it difficult to integrate and synchronize data across different platforms.

Another critical challenge lies in the nature of medical information, which is inherently complex and context-rich. The medical field relies heavily on nuanced and detailed documentation, making it a high-context domain for Natural Language Processing (NLP). Traditional rule-based algorithms struggle to effectively process and interpret this data due to the vast variability and specificity inherent in medical language. Consequently, these algorithms often fall short in delivering accurate and reliable results, limiting their practical application in HIS.

Large Language Models (LLMs), such as those developed by OpenAI or Anthropic, offer a promising solution to these challenges. LLMs are advanced AI systems trained on extensive datasets, capable of understanding and generating human-like text. They excel in processing unstructured data, making them well-suited for handling the

complexities of medical information. By leveraging LLMs, HIS can potentially overcome the limitations of rule-based systems, enabling more accurate data processing, improved interoperability, and enhanced decision support.

1.2 Research Objectives

The primary objective of this research is to explore the pain points experienced by healthcare providers in using current HIS and to investigate the potential of LLMs in addressing these challenges, providing a proof of concept. This involves:

1. Identifying Specific Challenges: Through interviews with healthcare professionals, we aim to identify the specific challenges they face with the current HIS.
2. Evaluating LLM Capabilities: Assessing the potential benefits of LLMs in addressing the identified challenges within HIS, based on literature and interview data.
3. Proof of Concept : Using test patient cases to ask existing LLMs to perform specific tasks as a proof of concept, demonstrating their capability of models to solve the pain points gathered from the interviews.
4. Providing Recommendations: Offering recommendations on how LLM technology can be leveraged to improve the quality of patient care and operational efficiency in hospitals.

1.3 Research Questions

To guide this research, the following questions will be addressed:

1. What are the main pain points experienced by healthcare providers with the current HIS and how do these challenges affect patient care?
2. What are potential capabilities of LLMs in addressing these pain points?
3. How effectively do LLMs perform in addressing the identified challenges in test patient cases?
4. What recommendations can be made for leveraging LLM technology in HIS?

CHAPTER II
LITERATURE REVIEW

2.1 Overview of Hospital Information Systems (HIS)

Hospital Information Systems (HIS) are comprehensive, integrated systems designed to manage the administrative, financial, and clinical aspects of a hospital. These systems are essential for managing patient data, supporting decision-making processes, and enhancing the overall efficiency of healthcare delivery. Major components of HIS, particularly on the clinical side, include, but are not limited to, Electronic Medical Records (EMR), patient management systems, Computerized Physician Order Entry (CPOE), and clinical decision support systems. The primary goal of HIS is to improve the quality of patient care by ensuring that relevant information is available to healthcare providers when needed (Wager, Lee, & Glaser, 2017).

2.2 Challenges in Current HIS

2.2.1 Complexity of Medical Language

Medical language is inherently complex and context-rich, posing significant challenges for traditional rule-based Natural Language Processing (NLP) systems. Medical terminology encompasses a vast array of specialized vocabulary, abbreviations, and jargon that vary significantly across different medical disciplines and even among individual practitioners (Kwon et al., 2022; Ji et al., 2024; Berge et al., 2023). Additionally, medical records often include a mixture of structured data (such as vital signs, laboratory results) and unstructured data (such as physician notes), further complicating the interpretation process (Zhang et al., 2020; Berge et al., 2023).

The variability in language use, including synonyms, acronyms, and the contextual nature of medical terms, means that rule-based NLP systems frequently struggle to accurately process and interpret this information. For example, the term "BP"

can mean "blood pressure" or "bronchopneumonia ," depending on the context. Rule-based algorithms rely on predefined sets of rules and patterns which depend on expert knowledge, which are often insufficient to capture the nuances and complexity of medical language. This limitation leads to issues such as less flexibility in data handling, incorrect data extraction, missed critical information, and ultimately, unreliable results (Berge et al., 2023; Ji et al., 2024).

Moreover, the rapid evolution of medical knowledge and terminology presents a constant challenge for maintaining and updating rule-based systems. The introduction of new medical terms, treatments, and procedures requires continuous adjustments to the algorithms, which can be resource-intensive and prone to errors (Berge et al., 2023; Ji et al., 2024).

2.2.2 Data Standardization and Interoperability

One of the most significant challenges in current Hospital Information Systems (HIS) is the lack of standardized data formats, leading to fragmented and unstructured patient records. This fragmentation hampers efficient data retrieval and exchange, complicating efforts to provide cohesive and continuous patient care. The problem is exacerbated by the presence of multiple vendors providing diverse HIS solutions, each with its proprietary data formats and standards (Reisman , 2017; Berge et al., 2023).

Data standardization involves the adoption of uniform formats and definitions for data elements to ensure consistency across different systems and platforms. Without standardized data, integrating information from various sources becomes a daunting task, resulting in data silos where patient information is trapped in isolated systems. This fragmentation not only hinders the seamless exchange of information but also increases the likelihood of errors in patient records (Reisman , 2017).

Interoperability refers to the ability of different HIS to communicate and exchange data effectively. The lack of interoperability between systems can lead to significant inefficiencies, such as duplicated tests, delayed treatments, and increased administrative burdens . For example, if a patient's medical history is not readily accessible from the same or other hospitals to a treating physician due to interoperability issues, it

can result in redundant diagnostic procedures or suboptimal treatment plans (Stewart et al., 2010, Reisman , 2017, Alberta College, 2023).

The presence of multiple vendors with diverse HIS solutions further complicates interoperability. Each system may use different coding standards, data structures, and communication protocols, making it challenging to synchronize and integrate data across platforms (Reisman, 2017). Efforts like the adoption of Health Level Seven (HL7) standards and Fast Healthcare Interoperability Resources (FHIR) are steps toward improving interoperability, but widespread adoption and consistent implementation remain ongoing challenges (Ayaz et al., 2021; Michaels et al., 2021).

2.3 Introduction to Large Language Models (LLMs)

2.3.1 Development and Capabilities

Large Language Models (LLMs), such as those developed by OpenAI, represent a significant advancement in Natural Language Processing (NLP) technology. These models, which include well-known examples like GPT-3 and GPT-4, are trained on extensive datasets that encompass a diverse array of text from the internet. This vast amount of training data allows LLMs to understand and generate human-like text with a high degree of fluency and coherence (OpenAI , 2022). The development of the transformer architecture has been pivotal in enabling the creation of these advanced models, marking a notable departure from earlier NLP techniques (Vaswani et al., 2017).

The advancements in LLMs are not just about size but also about their ability to engage in in-context learning. In-context learning refers to the model's capability to perform tasks by conditioning on examples provided in the input without requiring additional training. This is demonstrated through techniques such as few-shot learning, where the model is given a few examples of a task in the prompt and can then generalize from those examples to perform the task on new data. This flexibility significantly reduces the need for extensive fine-tuning for each specific application (Mosbach et al., 2023).

Another important feature of LLMs is their potential to be fine-tuned for specific tasks or domains. Fine -tuning involves training the pre-trained model further

on a smaller, task-specific dataset. This process allows the model to adapt to particular nuances and requirements of the task, improving its performance. For example, a general-purpose language model can be fine-tuned on medical texts to create a model that is more adept at handling medical inquiries and generating better efficiency and accuracy at medical information (Nawab, 2024; Wang et al., 2023).

The impact of LLMs extends across various industries and applications. In healthcare , they have potential to assist in analyzing and summarizing medical records and providing decision support (Nawab, 2024, Ong et al., 2024). Popular large language model chatbots have been evaluated for their accuracy, comprehensiveness, and self-awareness in answering ocular symptom queries (Wong et al., 2023).

In summary, the advent of LLMs such as GPT-3 and GPT-4 signifies a remarkable leap in NLP capabilities, driven largely by the transformer architecture. These models excel in understanding and generating human-like text, handling complex and unstructured data, and performing a broad spectrum of tasks with minimal task-specific training. Their development marks a transformative era in artificial intelligence, with far-reaching implications across various fields and applications (OpenAI, 2022; Vaswani et al., 2017; Mosbach et al., 2023; Nawab, 2024; Wang et al., 2023; Wong et al., 2023).

2.4 Potential Applications of LLMs in Healthcare

2.4.1 Clinical Decision Support

Large Language Models (LLMs) have transformative potential in healthcare, particularly in clinical decision support systems (CDSS). Models like GPT-3, GPT-4, and ClinicalGPT are designed to enhance the accuracy, efficiency, and reliability of clinical decision-making.

One key application is enhancing diagnostic accuracy. Karabacak and Margetis (2023) discuss how LLMs can improve diagnostic accuracy by analyzing vast amounts of medical data, thereby supporting clinical decision-making. This approach leverages the LLMs' ability to process and understand complex medical information, which can aid in providing more accurate diagnoses and treatment recommendations.

Additionally, LLMs are being used to enhance medication safety. Ong et al. (2023) introduced a Retrieval Augmented Generation (RAG)-based LLM framework for CDSS, which significantly improved the accuracy of detecting drug-related problems when used alongside junior pharmacists. This framework demonstrates the potential of LLMs to support clinical environments where precision and timely decision-making are crucial.

LLMs also show promise in handling unstructured data, such as electronic health records (EHRs). By integrating LLMs with EHR systems, healthcare providers can better manage patient data, ensuring comprehensive and up-to-date records. This integration helps streamline workflows and enhances the overall quality of patient care by providing healthcare professionals with quick and accurate access to patient histories and medical information.

Moreover, LLMs can assist in patient education and engagement. They can generate understandable and relevant information for patients, helping them comprehend their health conditions and treatment options. This can lead to better patient outcomes as individuals become more informed and involved in their healthcare decisions.

Despite their potential, the integration of LLMs in healthcare requires addressing several challenges, including ensuring data privacy, minimizing biases, and establishing robust evaluation metrics and regulatory frameworks (Karabacak & Margetis, 2023). By tackling these challenges, LLMs can be responsibly and effectively integrated into medical practice, ultimately enhancing patient care and improving health outcomes.

In summary, LLMs like GPT-3, GPT-4, and ClinicalGPT can revolutionize clinical decision support by enhancing diagnostic accuracy, improving medication safety, and handling unstructured data. Continued research and fine-tuning of these models are essential for maximizing their utility in healthcare (Karabacak & Margetis, 2023; Ong et al., 2023).

2.4.2 Data Management and EMR

The integration of Large Language Models (LLMs) into data management and electronic medical records (EMR) systems presents substantial opportunities for enhancing the efficiency, accuracy, and utility of healthcare data processing. By leveraging advanced natural language processing (NLP) capabilities, LLMs can significantly

improve various aspects of EMR management, including data standardization, summarization, and security.

- Data Standardization and Integration: A major challenge in EMR systems is the lack of standardized data formats, leading to fragmented patient records that impede efficient data retrieval and exchange. LLMs can assist in standardizing data by mapping diverse data inputs into a unified format, thus facilitating seamless integration across different healthcare systems. This capability is crucial given the complexity of medical language, which includes a variety of synonyms, acronyms, and context-dependent terms (Van Veen et al., 2023).

- Summarization and Information Retrieval: Summarizing electronic health records (EHRs) can greatly reduce the time clinicians spend on documentation, allowing them to focus more on patient care. Recent advancements propose using retrieval augmented generation (RAG) combined with question-answering techniques to extract relevant information from EHRs based on specific queries posed by medical professionals. This approach not only minimizes 'screen time' but also ensures that the summaries are contextually relevant and diverse, addressing the specific needs of different healthcare providers (Saba et al., 2023).

- Improving Clinical Documentation: LLMs can enhance the quality and efficiency of clinical documentation by automatically generating summaries of patient interactions and medical histories. This is particularly beneficial in settings where physicians spend a significant portion of their time on documentation tasks. By employing LLMs for summarizing clinical notes, healthcare providers can maintain comprehensive patient records without the added administrative burden (Van Veen et al., 2023).

- Case Studies and Applications: Several case studies highlight the successful application of LLMs in healthcare data management. For instance, the use of models fine-tuned for medical applications, such as those evaluated for summarizing clinical text, has shown significant improvements in tasks such as named entity recognition and relation extraction. Additionally, the implementation of question-based summarization using RAG has demonstrated promising results in generating accurate and relevant summaries of EHRs (Saba et al., 2023; Van Veen et al., 2023).

In conclusion, integrating LLMs into data management and EMR systems offers a promising avenue for enhancing the efficiency and accuracy of healthcare data

processing. By addressing challenges related to data standardization, summarization, and security, LLMs can significantly improve clinical documentation and information retrieval, ultimately leading to better patient care outcomes. Continued research and development in this field will be essential to fully realize the potential of LLMs in transforming healthcare data management (Saba et al., 2023; Van Veen et al., 2023).

2.4.3 Administrative Efficiency

The integration of Large Language Models (LLMs) such as OpenAI's GPT-3.5 Turbo has shown promising potential in enhancing administrative efficiency within healthcare. Such as in the automation of assigning International Classification of Diseases (ICD) codes to clinical documentation. This process, traditionally performed manually, is time-consuming and prone to errors, significantly impacting workflow efficiency.

Automation of ICD Code Assignment A study by Khalid Nawab et al. (2024) evaluated GPT-3.5 Turbo for automating ICD code assignments using the MIMIC-IV-Note dataset. Initially, the model correctly assigned the target ICD-10 code in 29.7% of cases. However, after fine-tuning with specific training data, accuracy improved to 62.6%. This improvement highlights the potential for fine-tuned models to enhance the accuracy and efficiency of administrative tasks in healthcare.

- Benefits and Challenges

1. Increased Efficiency and Accuracy: Automating ICD code assignments can significantly reduce errors and increase the speed of processing clinical documentation.
2. Cost Reduction: Automation reduces the need for extensive manual coding, lowering operational costs.
3. Time Savings: Healthcare professionals can focus more on patient care instead of administrative tasks.
4. Reimbursement: Accurate ICD coding is crucial for proper reimbursement from insurance companies and Medicare or NHSO, ensuring healthcare providers are compensated correctly for their services.

2.5 Challenges in Using LLM in HIS

While LLMs offer significant benefits, their integration into HIS presents several challenges that need to be addressed to maximize their efficacy and safety.

2.5.1 Hallucination

Hallucinations in Large Language Models (LLMs) refer to instances where these models generate outputs that are factually incorrect or unsupported by the input data. This problem is significant because it affects the reliability and trustworthiness of LLMs in applications that require precise and accurate information.

Hallucinations occur due to the inherent nature of LLMs, which generate text based on patterns learned from vast datasets. These models can produce confident but incorrect statements, leading to misinformation. Addressing hallucinations is crucial for the deployment of LLMs in sensitive areas such as healthcare, law, and education, where accuracy is paramount (Tonmoy et al., 2024).

- Retrieval-Augmented Generation (RAG):

One approach to mitigate hallucinations is Retrieval-Augmented Generation (RAG), which integrates information retrieval mechanisms with generation processes. RAG combines a pre-trained sequence-to-sequence transformer model with a dense vector index of external knowledge, such as Wikipedia. The Dense Passage Retriever (DPR) supplies relevant documents based on the input query, which the sequence-to-sequence model (e.g., BART) uses to generate the final output. This method ensures that the generated text is grounded in real, verifiable information, reducing the likelihood of hallucinations (Lewis et al., 2021).

2.5.2 Challenges in Knowledge Updating for Large Language Models

Updating the knowledge in large language models (LLMs) like GPT-4 is crucial for maintaining their relevance and accuracy, given the dynamic nature of the real world, which is continually generating new knowledge and requiring the unlearning of outdated information. This process presents several challenges:

- Computational Demands

Updating LLMs is computationally intensive due to their extensive parameterization. Even minor updates can require substantial computational resources, making frequent updates impractical and costly (Zhang et al., 2024).

- Integration of New Knowledge

Incorporating new information without disrupting existing knowledge is difficult. LLMs must integrate new facts seamlessly while preserving overall performance across various tasks. Sophisticated techniques are needed to merge new knowledge effectively (Zhang et al., 2024).

New information can conflict with existing knowledge, leading to inconsistencies and errors. Resolving these conflicts without manual intervention is crucial for maintaining reliability. Techniques like knowledge editing and retrieval-augmented generation (RAG) help dynamically update the model's knowledge base (Zhang et al., 2024; Tonmoy et al., 2024).

2.5.3 Data Security and Privacy

Large Language Models (LLMs) present significant challenges regarding data security and privacy, particularly due to their propensity to memorize and potentially leak sensitive information. Studies, such as those by Carlini et al., have demonstrated that LLMs can memorize individual training examples, which can be extracted through targeted queries, posing risks of data leakage (Carlini et al., 2021). These concerns are particularly acute in the biomedical domain, where the handling of Personal Identifiable Information (PII) and other sensitive data is critical.

- Memorization and Leakage: LLMs can unintentionally memorize and disclose training data, leading to privacy breaches. For instance, through malicious querying, attackers can extract verbatim sequences from the model's training set, as shown with GPT-2 (Carlini et al., 2021). This problem is exacerbated in models trained on non-public data, where the risk of exposing confidential information, such as medical records, is high.

- Defense Mechanisms: Several strategies have been proposed to mitigate these risks. One approach is data deduplication during preprocessing, which reduces the amount of memorized text and hence the potential for sensitive data leaks (Das et al.,

2024). Differential Privacy (DP) methods, such as Differentially Private Stochastic Gradient Descent (DP-SGD), can also help protect training data, though they may impair model utility and incur high computational costs (Das et al., 2024). Additionally, Named Entity Recognition (NER) can be employed to scrub datasets of PII, although this method is not foolproof and may reduce dataset utility.

- Challenges and Future Directions: Despite these efforts, achieving comprehensive protection against privacy breaches remains challenging. Existing defense techniques, primarily designed for smaller models, need adaptation and extensive testing for LLMs (Das et al., 2024). More research is required to develop effective, scalable solutions that balance privacy protection with model performance. This includes exploring secure multi-party computation and real-time privacy monitoring.

CHAPTER III
RESEARCH METHODOLOGY

This chapter outlines the methodological approach employed to explore the potential of Large Language Models (LLMs) in addressing the challenges of Hospital Information Systems (HIS). The study adopts a mixed-method approach, combining qualitative interviews with healthcare professionals and proof-of-concept testing of LLM capabilities. This dual approach allows for a comprehensive exploration of both the current challenges in HIS and the potential solutions offered by LLM integration.

The research methodology is designed to address the following key objectives:

1. To identify and analyze the specific pain points experienced by healthcare professionals in their interaction with current HIS.
2. To explore healthcare professionals' perspectives on the potential integration of LLMs into HIS.
3. To evaluate the technical feasibility and potential impact of LLM integration through proof-of-concept testing.
4. To develop practical recommendations for the implementation of LLM-enhanced HIS.

The methodology is structured in three main phases:

1. Data Collection: This phase includes semi-structured interviews with healthcare professionals and a developer, as well as proof-of-concept testing of LLMs based on identified pain points.
2. Data Analysis: A qualitative analysis of interview responses and proof-of-concept results to identify common themes, challenges, and potential solutions.
3. Synthesis and Recommendations: Integration of insights from both data collection methods to develop comprehensive recommendations for LLM integration in HIS.

3.1 Participant Selection

The study will involve interviews with a group of healthcare professionals, with experience of using HIS for more than three years, to gather comprehensive insights into the challenges and potential improvements in Hospital Information Systems (HIS). The participants will include:

- Four doctors with experience of both private and public hospitals in Thailand.
- One nurse from a public hospital.
- One developer who works with HIS in a public hospital.

3.2 Data Collection

The data collection process will involve two main phases: interviews with healthcare professionals and proof-of-concept testing with LLMs.

3.2.1 Interviews

The interviews will focus on understanding the pain points experienced by these healthcare professionals with the current HIS. The interviews will cover the following aspects:

1. Identifying Pain Points: Participants will be asked to describe the specific challenges they face with the current HIS.
2. Concept of LLM Integration: The concept of integrating Large Language Models (LLMs) into HIS will be introduced to the participants. This will include explaining how LLMs can potentially enhance data processing, improve decision support, and streamline administrative tasks.
3. Perceived Improvements: Participants will be asked to provide their perspectives on how LLMs could address the identified challenges and improve HIS functionality.
4. Developer Concerns: The developer will be asked about technical and implementation concerns regarding the integration of LLMs into HIS.

3.2.2 Proof of Concept Testing

Following the interviews, we will conduct proof-of-concept tests based on the identified pain points. This phase will involve:

1. Deriving Test Cases: Creating specific scenarios based on the pain points highlighted by the interviewees.

2. LLM Testing: Using advanced LLMs to perform tasks relevant to the identified challenges in the test cases. This could involve:

- Processing unstructured data
- Improving data retrieval
- Supporting decision-making processes
- Generating user interfaces for data input

3. Evaluation: Assessing the effectiveness of LLMs in addressing the pain points, its accuracy, and potential user satisfaction.

This approach allows us to:

- Validate the insights gathered from interviews
- Demonstrate practical applications of LLMs in addressing HIS challenges
- Provide concrete examples of how LLMs could improve HIS functionality

By combining these methods, we aim to provide a comprehensive understanding of both the current challenges in HIS and the potential solutions offered by LLM integration.

3.3 Interview Questions

1. For Doctors and Nurse:

- What are the main challenges you face with the current HIS, specifically in clinical operations?
- How do these challenges affect your workflow and patient care?
- What features or improvements would you like to see in HIS?
- How do you think integrating LLMs into HIS could address these challenges?
- Can you provide specific examples of tasks where LLMs could be beneficial?

2. For Developer:

- What are the main technical challenges in maintaining and updating the current HIS?
- What are your concerns regarding data security and privacy when integrating LLMs?
- What technical improvements or support would be necessary to integrate LLMs into HIS?

3.4 Data Analysis

The data from both interviews and proof-of-concept tests will be analyzed qualitatively to identify common themes, specific insights, and practical implications. The analysis will focus on:

- Common Pain Points: Identifying recurring challenges mentioned by multiple participants.
- Potential Solutions and Limitations: Evaluating the feasibility and potential impact of integrating LLMs based on participants' feedback and proof-of-concept results.
- Technical Feasibility: Assessing the practical implementation challenges and opportunities based on the developer's input and proof-of-concept outcomes.
- User Acceptance: Gauging potential user satisfaction and adoption based on healthcare professionals' responses to LLM capabilities demonstrated in the proof-of-concept.

This restructuring provides a more integrated view of your methodology, showing how the interview insights directly inform the proof-of-concept testing, and how both contribute to your overall analysis and findings.

3.5 Expected Outcomes

The study aims to provide:

- An understanding of the challenges faced by healthcare professionals with current HIS.
- Insights into how LLMs can be leveraged to improve HIS functionality.
- Proof of concept demonstrating the potential of LLMs to enhance HIS, validated through test cases derived from real-world pain points.
- Practical recommendations for integrating LLMs into HIS.

From the results of the LLM tests, recommendations will be made based on the identified pain points and the outcomes of the proof of concept. This will guide optimizing HIS using LLM technology to improve patient care quality and operational efficiency in hospitals.

CHAPTER IV
RESEARCH FINDINGS

4.1 Interview Results

1. Interviews result from Doctor A

"... I have 12 years of experience in both public and private hospitals. One major pain point I've encountered with the HIS is the difficulty in inputting ICD-10 coding manually and when reviewing patient past records, especially chronic patients. If AI like ChatGPT were integrated into the HIS, it could be highly beneficial. For example, it could automatically input ICD-10 coding, link seamlessly with knowledge databases and various medical papers to predict common problems or pitfalls. …"

2. Interviews result from Doctor B

" …I have 11 years of experience, mostly in public hospitals. A significant pain point is completing the documentation, which is particularly problematic in the government reimbursement system. If the AI could gather all past patient data, including diseases, treatments, and precipitating factors, and summarize it into a flowchart or timeline, that would be incredibly valuable. Furthermore, summarizing data from evidence-based medical sites into categories like pathophysiology, epidemiology, symptoms, and treatments to provide information for specific patients would be ideal. However, for now, if the AI could just help summarize inpatient and outpatient charts as well as ICD codes to optimize reimbursement from the National Health Security Office (NHSO), it would be a significant improvement… "

3. Interviews result from Doctor C

"...I have 5 years of experience. I have used HIS seriously in two places: Udonthani Hospital during my medical school years, where I used it mainly to view patient history and lab results, and a NHSO clinic with Hospital OS for recording patient history, physical examinations, prescribing, and viewing lab results for OPD cases. I have also assisted a Health Tech company working on HIS and HIE for a year as a business analyst.

The main pain point with HIS is the ‘inputting patient records and doctor orders’ process. Currently, inputting data is quite challenging and time-consuming per case. The quality of the data entered is often poor, with free text or incorrect entries making the data less reliable. If AI like ChatGPT were integrated into the system to simplify data input, it would be immensely helpful. For instance, if the AI could convert free text into structured data, fill in default values, and suggest common patterns based on frequently used data, it would save time and improve data quality. For example, entering ‘right lower abdominal pain for 1 day’ could prompt the AI to suggest an appendicitis pattern with prefilled physical examination fields and management steps.

Improving data input quality with AI would not only enhance the immediate use of HIS but also facilitate future uses like research and patient history summaries. Thus, the key for HIS is quality data input, and AI should assist in making the process easier and more accurate while protecting the privacy of patients. AI should also help screen for potential errors, like misentered birth years or lab values….”

4. Interviews result from Doctor D

“...I have 13 years of experience in both public and private hospitals. One of the major issues is that reviewing patient histories takes a long time and can easily result in omissions. Data entry both patient records and treatment orders are cumbersome due to numerous fields. I would like AI to help summarize patient histories or have an AI chatbot that can summarize histories and assist in reviewing them through conversation. Additionally, the AI should help with differential diagnoses (DDx) based on summarized information provided by doctors in case doctors overlooked some uncommon diseases.

The system should include alerts for incorrect dosages or duplicate prescriptions. It would also be helpful if AI could draft medical certificates , allowing doctors to make final edits. HIS should facilitate cross-hospital patient history reviews, which it currently lacks. An ideal system would also feature a Clinical Decision Support System (CDSS) capable of answering medical questions with references, and it should be able to operate in Thai language….”

5. Interviews result from Nurse A

“...Doctors often refuse to use the HIS because they find data entry too cumbersome, preferring to write things down themselves. This handwriting is often hard

to read, making it difficult for me to understand the patient's case. If the system could make it easier for doctors to enter data, or if it could support voice input for data entry, that would be a significant improvement. Additionally, there are times when I cannot detect errors, such as incorrect medication orders. If the system could alert me about such errors, it would be greatly beneficial. Separately, sometimes I don't fully understand the overall medical picture of a case. If AI could summarize patient history in a way that highlights the key information nurses need to know, it would be very helpful…."

6. Interviews result from Developer A

"...I have been working in a government hospital as an IT support and helping with IT for the past 10 years. One major problem is the lack of standardized structure of data storage. When users request additional fields, I have to manually create them and add columns to the database. The input interface often needs customization for each department.

Exchanging data with other systems is very challenging. We sometimes encounter data leaks, which is a significant concern. If we were to implement AI, we would likely need to send data abroad, raising privacy and PDPA compliance issues.

Additionally, relying heavily on AI, like ChatGPT, poses reliability concerns. For example, ChatGPT recently experienced downtime. We can't afford such disruptions in critical healthcare systems. Another concern is the accuracy of the AI. Medical information must be precise, and any inaccuracies could lead to severe consequences. Therefore, integrating AI requires careful consideration of these potential issues…."

4.2 Interview Data Analysis

The interviews with healthcare professionals and IT staff highlight several key pain points and potential areas for improvement in the current Hospital Information System (HIS) through the integration of AI technologies like ChatGPT.

Table 4.1 Healthcare Professionals' Insights on Key Factors Influencing LLM Integration in HIS

Key Factors	Sub-Factors	Findings from Interviews
Data Retrieval and Summarization	Patient History Review	"If the AI could gather all past patient data, including diseases, treatments, and precipitating factors, and summarize it into a flowchart or timeline, that would be incredibly valuable." - Doctor B "One of the major issues is that reviewing patient histories takes a long time and can easily result in omissions. I would like AI to help summarize patient histories or have an AI chatbot that can summarize histories and assist in reviewing them through conversation." - Doctor D "If AI could summarize patient history in a way that highlights the key information nurses need to know, it would be very helpful." - Nurse A
	Summarizing Complex Information	"If AI like ChatGPT were integrated into the HIS, it could be highly beneficial. For example, it could automatically input ICD-10 coding, link seamlessly with knowledge databases and various medical papers to predict common problems or pitfalls." - Doctor A "Summarizing data from evidence-based medical sites into categories like pathophysiology, epidemiology, symptoms, and treatments to provide information for specific patients would be ideal." - Doctor B

Table 4.1 Healthcare Professionals' Insights on Key Factors Influencing LLM Integration in HIS (cont.)

Key Factors	Sub-Factors	Findings from Interviews
		"The AI should help with differential diagnoses (DDx) based on summarized information provided by doctors in case doctors overlooked some uncommon diseases." - Doctor D
Data Input Challenges	Manual Data Entry	"One major pain point I've encountered with the HIS is the difficulty in inputting ICD-10 coding manually and when reviewing patient past records, especially chronic patients." - Doctor A "A significant pain point is completing the documentation, which is particularly problematic in the government reimbursement system." - Doctor B "The main pain point with HIS is the 'inputting patient records and doctor orders' process. Currently, inputting data is quite challenging and time-consuming per case." - Doctor C "Doctors often refuse to use the HIS because they find data entry too cumbersome, preferring to write things down themselves." - Nurse A
	Complexity and Usability	"Data entry both patient records and treatment orders are cumbersome due to numerous fields." - Doctor D "Doctors often avoid using HIS due to the cumbersome data entry process." - Nurse A

Table 4.1 Healthcare Professionals' Insights on Key Factors Influencing LLM Integration in HIS (cont.)

Key Factors	Sub-Factors	Findings from Interviews
Accuracy and Error Prevention	Alert Systems	"The system should include alerts for incorrect dosages or duplicate prescriptions." - Doctor D "If the system could alert me about such errors, it would be greatly beneficial." - Nurse A
	Error Detection	"AI should help screen for potential errors, like misentered birth years or lab values." - Doctor C
Interoperability and Data Exchange	Cross-Hospital Data Sharing	"HIS should facilitate cross-hospital patient history reviews, which it currently lacks." - Doctor D "Exchanging data with other systems is very challenging." - Developer A
	Standardization Issues	"One major problem is the lack of standardized structure of data storage. When users request additional fields, I have to manually create them and add columns to the database. The input interface often needs customization for each department." - Developer A
Reliability and Privacy Concerns	System Downtime	"Relying heavily on AI, like ChatGPT, poses reliability concerns. For example, ChatGPT recently experienced downtime. We can't afford such disruptions in critical healthcare systems." - Developer A

Table 4.1 Healthcare Professionals' Insights on Key Factors Influencing LLM Integration in HIS (cont.)

Key Factors	Sub-Factors	Findings from Interviews
	Privacy and Compliance	"If we were to implement AI, we would likely need to send data abroad, raising privacy and PDPA compliance issues." - Developer A "AI should assist in making the process easier and more accurate while protecting the privacy of patients." - Doctor C
Language and Communication	Language Support	"An ideal system would also feature a Clinical Decision Support System (CDSS) capable of answering medical questions with references, and it should be able to operate in Thai language." - Doctor D
	Enhanced Communication	"If the system could make it easier for doctors to enter data, or if it could support voice input for data entry, that would be a significant improvement." - Nurse A

1. Data Retrieval and Summarization:

- Patient History Review: Doctor B, D and Nurse A noted that reviewing patient histories is time-consuming and can easily result in omissions. AI could help by summarizing patient histories and assisting in reviews. Doctor B specifically mentioned the value of summarizing patient data into flowcharts or timelines for better understanding and government reimbursement processes.
- Summarizing Complex Information: Doctors A, B, and D highlighted the need for AI to link with knowledge databases and provide differential diagnoses.

2. Data Input Challenges:

- Manual Data Entry: Doctors and Nurse A find manual data entry time-consuming and prone to errors, which impacts the quality and reliability of patient records. Both Doctors A and B emphasized the difficulty and time required for inputting ICD-10 codes and treatment orders.
- Complexity and Usability: Doctor D mentioned the cumbersome nature of inputting patient records due to numerous fields. Nurse A also pointed out that doctors often avoid using HIS due to the cumbersome data entry process.

3. Accuracy and Error Prevention:

- Alert Systems: Doctors D and Nurse A stressed the importance of AI providing alerts for incorrect dosages, duplicate prescriptions, and other potential errors.
- Error Detection: Doctor C suggested that AI could help screen for potential errors in data entry, such as misentered birth years or lab values.

4. Interoperability and Data Exchange:

- Cross-Hospital Data Sharing: Doctor D and Developer A highlighted the challenge of reviewing patient histories across different hospitals. An AI-enabled HIS could facilitate better data exchange and interoperability.
- Standardization Issues: Developer mentioned the lack of standardized data storage structures, making it difficult to add new fields and customize the input interface for different departments.

5. Reliability and Privacy Concerns:

- System Downtime: Developer A expressed concerns about the reliability of AI systems, citing recent downtime with ChatGPT. Relying heavily on AI could pose risks if the system is not consistently available.
- Privacy and Compliance: Both Developer A and Doctor C raised concerns about privacy and compliance with regulations like PDPA if data needs to be sent abroad for AI processing.

6. Language and Communication:

- Language Support: Doctor D emphasized the need for AI systems to operate in Thai to be effective in the local healthcare context.

- Enhanced Communication: Nurse A suggested that AI could help improve communication by supporting voice input and summarizing key patient information for easier understanding.

4.3 Findings for Proof of Concept

4.3.1 Data Summarization and Thai Language Support

The proof of concept involved testing the ability of an AI system to read and summarize patient data from a mock-up dataset adapted from patients in JSON format. This dataset contained multiple visit records structured based on a free version of a popular HIS software in Thailand. The findings demonstrate the AI's capabilities in processing complex medical data providing valuable insights into the potential integration of AI in Hospital Information Systems (HIS).

4.3.1.1 Test Case: AI's Ability to Summarize JSON Input

We attached the JSON file to Claude 3 Opus due to its larger context windows, as ChatGPT's context length was insufficient, to summarize patient records.

Table 4.2 Example Input JSON Excerpt from one of a Hospital Visit


```
{
    "visitDateTime": "2025-01-12T03:00:11.000Z",
    "diagnoses": [
    {
    "icd10": "I10",
    "icd10Name": "Essential (primary) hypertension",
    "icd10ThaiName": "โรคความดันโลหิตสูง",
    "diagTypeName": "Principal Diagnosis",
    },
    {
    "icd10": "E782",
    "icd10Name": "Mixed hyperlipidaemia",
    "icd10ThaiName": "ภาวะที่มีไขมันหลายชนิดในเลือดสูง",
    "diagTypeName": "Comorbidity"
    },
```

Table 4.2 Example Input JSON Excerpt from one of a Hospital Visit (cont.)


```
                {
                "icd10": "N182",
                "icd10Name": "Chronic kidney disease, stage 2",
                "icd10ThaiName": "โรคไตเรื้อรังระยะที่ 2",
                "diagTypeName": "Comorbidity"
                }
                ],
"medications": [
                {
                        "drugNondugName": "amlodipine (GPO) 5 mg. tablet",
                        "drugUsage": "Take 2 tablets once a day after breakfast",
                        "rxDateTime": "2025-01-12T04:06:30.000Z"
                },
                {
                        "drugNondugName": "enalapril (pink blister pack) 5 mg. tablet",
                        "drugUsage": "Take 1 tablet twice a day after meals (morning and
evening)",
                        "rxDateTime": "2025-01-12T04:06:31.000Z"
                },
                {
                        "drugNondugName": "SIMVASTATIN (GPO) 20 mg. tablet",
                        "drugUsage": "Take 1 tablet before bedtime",
                        "rxDateTime": "2025-01-12T04:07:17.000Z"
                }
  ]
  …
}
```

- AI's Summarized Output:

Table 4.3 Underlying Diseases Summarization:

Input(Th): สรุป underlying diseases ผู้ป่วยรายนี้ให้หน่อย **Input(En):** Summarize the underlying diseases for this patient.
Response (Th): จากประวัติการรักษาของคนไข้รายนี้ พบว่ามีโรคประจำตัวที่สำคัญดังนี้ 1. โรคความดันโลหิตสูง (Essential hypertension) 2. ภาวะไขมันในเลือดสูงหลายชนิด (Mixed hyperlipidemia) 3. โรคไตเรื้อรังระยะที่ 2 (Chronic kidney disease, stage 2) นอกจากนี้ยังพบว่ามีความผิดปกติของหลอดเลือดแดงที่ขา (Atherosclerosis) ซึ่งทำให้เกิดภาวะเนื้อตายของนิ้วเท้าขวาจากการขาดเลือด (Gangrene) สรุปโรคร่วมหรือโรคประจำตัวที่สำคัญของผู้ป่วยคนนี้ ได้แก่ ความดันโลหิตสูง, ไขมันในเลือดสูง และโรคไตเรื้อรัง ซึ่งเป็นปัจจัยเสี่ยงต่อการเกิดโรคหลอดเลือดแดงตีบตันที่ขา ส่งผลให้เกิดภาวะแทรกซ้อนของการขาดเลือดที่เท้าขวาในที่สุด **Response (En):** Based on the patient's medical history, the significant underlying diseases include: 1. Essential hypertension 2. Mixed hyperlipidemia 3. Chronic kidney disease, stage 2 Additionally, there is a condition of atherosclerosis in the leg, which has led to gangrene of the right toe due to lack of blood supply. In summary, the comorbidities or significant underlying diseases for this patient are hypertension, hyperlipidemia, and chronic kidney disease, which are risk factors for arterial occlusion in the legs, leading to complications of ischemia in the right foot.

Table 4.4 Chronic Disease Medication Summary

Input(Th): สรุป timeline ของยาโรคเรื้อรังให้หน่อย

Input(En): Summarize the timeline of medications for chronic diseases.

Response (Th): สรุป timeline ของยารักษาโรคเรื้อรังได้ดังนี้

17 พฤศจิกายน 2567

- Amlodipine 5 mg ไม่ระบุขนาดการใช้
- Enalapril 5 mg 1 เม็ด เช้า (รับประทานวันละครั้ง แตกต่างจากเดิม)

12 มกราคม 2568

- Amlodipine 5 mg 2 เม็ดต่อวัน (ปรับขนาดยาขึ้น)
- Enalapril 5 mg 1 เม็ด วันละ 2 ครั้ง
- Simvastatin 20 mg 1 เม็ดก่อนนอน

9 มีนาคม 2568

- Amlodipine 5 mg 2 เม็ดต่อวัน
- Enalapril 5 mg 1 เม็ด วันละ 2 ครั้ง
- Atorvastatin 40 mg ครึ่งเม็ดก่อนนอน (แพทย์เปลี่ยนยาจาก Simvastatin เป็น Atorvastatin)

ดังนั้น สรุปการใช้ยารักษาโรคความดันและไขมันในเลือดสูง ณ ปัจจุบัน (ข้อมูลล่าสุดเมื่อ 9 มีนาคม 2568) คือ

- Amlodipine 5 mg รับประทาน 2 เม็ดต่อวัน
- Enalapril 5 mg รับประทาน 1 เม็ด วันละ 2 ครั้ง
- Atorvastatin 40 mg รับประทานครึ่งเม็ดก่อนนอน

Response (En): The timeline of chronic disease medications is as follows:

November 17, 2024

- Amlodipine 5 mg, unspecified dosage
- Enalapril 5 mg, 1 tablet in the morning (taken once a day, different from previous usage)

January 12, 2025

- Amlodipine 5 mg, 2 tablets per day (dosage increased)
- Enalapril 5 mg, 1 tablet twice a day
- Simvastatin 20 mg, 1 tablet before bedtime

March 9, 2025

- Amlodipine 5 mg, 2 tablets per day
- Enalapril 5 mg, 1 tablet twice a day
- Atorvastatin 40 mg, half a tablet before bedtime (physician changed medication from Simvastatin to Atorvastatin)

Therefore, the current medication regimen for hypertension and hyperlipidemia (latest information as of March 9, 2025) is:

- Amlodipine 5 mg, take 2 tablets per day
- Enalapril 5 mg, take 1 tablet twice a day
- Atorvastatin 40 mg, take half a tablet before bedtime

4.3.1.2 Test Cases: AI's Ability to Facilitate Medical Coding:

We utilized two state-of-the-art models, Claude Opus and ChatGPT-4o, for their ability to facilitate medical coding and yielded similar results and limitations.

Table 4.5 Medical Coding

Input: The patient has had a high fever and chills for one day. The examination revealed a temperature of 39 °C, a respiratory rate of 26 breaths per minute, a pulse rate of 100 beats per minute, and blood pressure of 120/70 mmHg. The patient experienced tenderness in the left costovertebral angle. Urine analysis showed a large number of white blood cells, and E. coli was found in both urine and blood cultures. The doctor summarized the discharge summary as follows: ● **Primary Diagnosis**: Acute Pyelonephritis ● **Secondary Diagnosis**: Sepsis due to E. coli Please provide the ICD codes for the diagnoses.
Response : Based on the provided information, the ICD codes for the patient can be summarized as follows: **Primary Diagnosis**: Acute Pyelonephritis **ICD-10 Code**: N10 (Acute pyelonephritis) **Secondary Diagnosis**: Sepsis due to E. coli **ICD-10 Code**: A41.51 (Sepsis due to Escherichia coli [E. coli]) Therefore, the ICD codes for the patient are: **N10** (Acute pyelonephritis) **A41.51** (Sepsis due to Escherichia coli [E. coli])
Input: A pregnant woman at 32 weeks gestation comes for her second prenatal visit. No abnormalities were found during the examination.

Table 4.5 Medical Coding (cont.)

Response : ICD-10 Codes:

1. **Routine Prenatal Visit:**
   - **Z34.82** (Encounter for supervision of normal pregnancy, second trimester)
   - For the third trimester, it would be Z34.83, but since the visit is for 32 weeks, it should be coded for the appropriate trimester.
2. **Normal Pregnancy:**
   - Z34.80 (Encounter for supervision of normal pregnancy, unspecified trimester)
   - Depending on local coding standards, you might specify Z34.8x with the appropriate trimester. Here, it's Z34.82 for the second trimester.

Input: Give me ICD-10 code for these cases

1. **Acute bronchitis**
   - A 12-year-old child has had a cough and chest pain for 7 days.
2. **Pneumothorax**
   - A 34-year-old man fell off his motorcycle, his chest hit the handlebar, and he feels chest tightness.
3. **Burn**
   - A 28-year-old woman was scalded by hot water. She has third-degree burns on her arm and second-degree burns on her thigh.
4. **Acute cholecystitis**
   - A 44-year-old woman had a fever and right lower abdominal pain. An ultrasound showed multiple gallstones with cholecystitis.
5. **Internal hemorrhoid**
   - A 38-year-old pregnant woman, 35 weeks, has internal hemorrhoids with bleeding during defecation.
6. **Blunt trauma abdomen**
   - A 25-year-old man was in a car accident and underwent abdominal surgery. He had a ruptured liver and a ruptured spleen.
7. **Mitral stenosis with mitral regurgitation**
   - A 28-year-old woman with a history of heart disease presented with shortness of breath.

Table 4.5 Medical Coding (cont.)

Response:

1. **Acute bronchitis**
   - ICD-10 Code: J20.9 (Acute bronchitis, unspecified)
2. **Pneumothorax**
   - ICD-10 Code: J93.9 (Pneumothorax, unspecified)
3. **Burn**
   - ICD-10 Codes:
     - Third-degree burn on arm: T22.30XA (Burn of third degree of unspecified site of right upper limb, initial encounter)
     - Second-degree burn on thigh: T24.212A (Burn of second degree of left thigh, initial encounter)
4. **Acute cholecystitis**
   - ICD-10 Code: K81.1 (Chronic cholecystitis)
5. **Internal hemorrhoid**
   - ICD-10 Codes:
     - Internal hemorrhoids with bleeding: K64.8 (Other specified hemorrhoids)
     - Pregnancy, 35 weeks: Z3A.35 (35 weeks gestation of pregnancy)
6. **Blunt trauma abdomen**
   - ICD-10 Codes:
     - Ruptured liver: S36.112A (Laceration of liver, initial encounter)
     - Ruptured spleen: S36.032A (Laceration of spleen, initial encounter)
7. **Mitral stenosis with mitral regurgitation**
   - ICD-10 Code: I08.0 (Rheumatic disorders of both mitral and aortic valves)

Please verify the specific details with the patient's medical records, as certain nuances may require different codes.

Adapted from: http://thcc.or.th/download/ICD-10.pdf, https://www4.si.mahidol.ac.th/th/department/cancer/file_shortcourse/2018/ICD10_Problem.pdf

4.3.1.3 Analysis of AI's Performance:

A. Comprehensive Data Extraction:

- The AI successfully parsed the JSON structure containing multiple patient visits, accurately extracting relevant medical information from each record.
- Despite the complexity and variability of medical terminology, the AI correctly identified key details such as diagnosis, medications, visit dates, and treatment plans.

B. Accurate Summarization:

- The AI produced clear and concise summaries of the patient's underlying diseases and medication timelines.
- It accurately identified chronic conditions like essential hypertension, mixed hyperlipidemia, and chronic kidney disease, along with associated complications such as atherosclerosis and gangrene.
- The medication timeline generated by the AI reflected the correct sequence and adjustments in treatment, showcasing its ability to track and summarize longitudinal medical data.

C. AI's Proficiency in Thai Language

- The AI demonstrated strong proficiency in generating responses in Thai, maintaining both fluency and contextual accuracy.
- Medical terms and patient details were accurately translated and conveyed in Thai, ensuring clarity and comprehensibility for Thai-speaking healthcare professionals.

D. Medical Coding:

- Medical coding is a crucial part of the healthcare process, particularly for reimbursement purposes. Accurate coding ensures that healthcare providers are appropriately compensated for the services they deliver. In our study, the coding outcomes were classified into three categories: correct, partially correct, and wrong. Partially correct coding is defined as being correct in major details but having minor inaccuracies.
- Out of 9 test cases, the AI produced 4 correct, 2 partially correct, and 3 wrong coding outcomes. In the wrong coding cases, even when the correct ICD code and details were matched, there were occasional errors in the AI's medical

knowledge. This indicates that the AI sometimes misunderstands the medical context or specifics, leading to inaccurate coding.

- While the AI has demonstrated partial success in providing accurate ICD-10 codes, it is essential to recognize that medical coding practices can vary by country. For instance, Thailand uses the ICD-10 TM (Thai Modification), which includes specific guidelines and codes tailored to the local context. To ensure the AI's outputs are fully compatible with local coding standards, further localization and customization are necessary. This involves adapting the AI to understand and apply country-specific modifications of the ICD-10 coding system.
- At present, the AI should be used with caution, serving as a suggestion tool for coders, similar to autocomplete functions. However, the potential for improvement is significant. Developing specialist models tailored specifically to medical coding could enhance the accuracy and reliability of the AI, making it a more valuable tool for healthcare providers in the future.

4.3.2 Accuracy and Error Prevention

Ensuring the accuracy of data entry and preventing errors in Hospital Information Systems (HIS) is crucial for maintaining the integrity of patient records and ensuring patient safety. This section explores the use of AI to enhance accuracy and error prevention in HIS.

1. Alert Systems

Doctors D and Nurse A highlighted the importance of AI systems providing alerts for incorrect dosages, duplicate prescriptions, and other potential errors. For example, an AI system could be programmed to:

- Detecting Medication Errors: If a prescribed dosage exceeds the recommended amount, the AI can alert the healthcare provider to verify and correct the dosage. This is particularly important for medications with narrow therapeutic windows. Also, the AI can detect when the same medication is prescribed multiple times or the patient is having allergy to medications, to prevent overmedication and potential adverse drug interactions.

2. Error Detection

Doctor C suggested that AI could help screen for potential errors in data entry, such as misentered birth years or lab values. This could involve:

- Data Validation Checks: The AI can cross-verify patient data with existing records to ensure consistency. For example, if the birth year entered does not match the patient's ID or previous records, the AI can flag this for review.

4.3.2.1 Test Case: Accuracy and Error Prevention

Three mock up patient records are used to evaluate the AI's ability to detect and alert healthcare providers about incorrect dosages, duplicate prescriptions, and data entry errors. We used both Claude Opus and ChatGPT-4o and yielded similar results and limitations.

Table 4.6 Accuracy and Error Prevention Case 1

Input: You are a HIS assistant helping doctors input data. Any suggestions on the following input? (The hospital is using Buddhism Era) **Patient Information:** ● **Name:** Pichit Sukjai ● **Date of Birth:** 10 มีนาคม 1982 ● **Patient ID:** 789012 ● **Allergies:** Sulfa drugs (difficulty breathing) **Visit Information:** ● **Visit Date:** 11 มิถุนายน 2567 (2024-06-11) ● **Diagnosis:** ○ Type 2 Diabetes Mellitus (E11) ○ Chronic Obstructive Pulmonary Disease (J44) **Medications Prescribed:** ● Metformin 500 mg, take 1 tablet twice a day ● Salbutamol 100 mcg, inhale 2 puffs every 6 hours ● Metformin 500 mg, take 1 tablet twice a day ● Hydrochlorothiazide 25 mg, take 1 tablet once a day
Response : **Suggestions for Improving Patient Data:** ● Correct the patient's date of birth to the Buddhist calendar to align with the hospital's system (10 มีนาคม 2525). ● Check for duplicate medications (Metformin has been prescribed twice). ● Verify the prescription of Hydrochlorothiazide, as the patient has a history of allergy to sulfa drugs, which includes Hydrochlorothiazide. Consider switching to another medication. ● Add the date and time of the lab results to provide better context. ● Consider adding other relevant information such as the patient's chief complaints, family history, and social history (e.g., smoking status, alcohol consumption) for a more comprehensive record.

Table 4.7 Accuracy and Error Prevention Case 2

Input: You are a HIS assistant helping doctors input data. Any suggestions on the following input?

Patient Information:

- **Name**: Pichit Sukjai
- **Date of Birth**: March 10, 1982
- **Patient ID**: 789012
- **Drug Allergy**: Amoxicillin (rash reaction)

Prescribed Medications:

- **Metformin 500 mg**: Take 1 tablet twice daily
- **Cephalexin 400 mg**: Take 1 capsule three times daily
- **Ibuprofen 400 mg**: Take 1 tablet three times daily

Response :

According to the provided information, the patient has a history of allergy to Amoxicillin, which is a Penicillin group drug. Therefore, prescribing Cephalexin, a Cephalosporin group drug, may not be appropriate. Patients allergic to Penicillin have a chance of being allergic to Cephalosporins as well (approximately 10% incidence).

If antibiotic use is necessary, consider choosing another class of antibiotics to which the patient has no known allergies, such as Macrolides (Erythromycin, Azithromycin), Fluoroquinolones (Ciprofloxacin, Levofloxacin), or Tetracyclines (Doxycycline). The choice depends on the indication and appropriateness for the individual case.

Although Ibuprofen is not an antibiotic, its indication for use should be reviewed carefully, as the patient has diabetes. NSAIDs can affect kidney function and should be used with caution in patients with chronic diseases.

Table 4.8 Accuracy and Error Prevention Case 3

Input: You are a HIS assistant helping doctors input data. Any suggestions on the following input?

Prescribed Medications:

- **Metformin 500 mg**: Take 1 tablet twice daily
- **Cephalexin 400 mg**: Take 1 capsule three times daily
- **Ibuprofen 400 mg**: Take 1 tablet 4 times daily

Additional Info:

- **Cephalexin Dosage and Administration**: If using a 250 mg dose, take every 6 hours (4 times a day) or if using a 500 mg dose, take every 12 hours (2 times a day) for 7-10 days.
- **Ibuprofen Dosage for Pain Relief and Fever Reduction**:
  - **Adults**: Take 200-400 mg every 4-6 hours. The maximum dose should not exceed 1.2 grams per day unless under medical supervision.
  - **Children 6 months to 12 years**: Take 5-10 mg per kg of body weight every 6-8 hours. The maximum dose should not exceed 40 mg per kg of body weight per day.

Response : **Suggestions for Recording Medication Information for This Patient:**

1. **Cephalexin Dosage**: The prescribed dose is 400 mg, 1x3 (1 tablet three times daily). According to additional information, Cephalexin 250 mg should be taken every 6 hours (4 times a day) or Cephalexin 500 mg should be taken every 12 hours (2 times a day). Adjust the dosage accordingly as per the recommended guidelines.
2. **Ibuprofen Dosage**: The prescribed dose is 400 mg, 1x4 (1 tablet four times daily). According to additional information, Ibuprofen for pain relief and fever reduction in adults should be taken 200-400 mg every 4-6 hours, with a maximum dose not exceeding 1.2 grams per day. The prescribed dosage is appropriate but ensure to advise the patient not to exceed the maximum daily dose.
3. **Medication Indications and Precautions**: Consider the indications and precautions for all three medications, especially Metformin, which is for diabetes management. Adjust the dosage based on blood sugar levels and kidney function. Also, advise the patient on diet and exercise for effective diabetes control.
4. **Patient Education**: Inform the patient about the correct way to take the medications, possible side effects, and symptoms that warrant a visit to the doctor. This ensures patient compliance and monitoring for adverse drug reactions.

4.3.2.2 Analysis & of AI's Performance:

Case 1

- Correctly Suggested Changing the Date of Birth to the Buddhist Calendar: Successfully identified the need to correct the patient's date of birth to align with the hospital's system.
- Detected Duplicate Prescription: Successfully identified and alerted about the duplicate prescription of Metformin.
- Identified Hydrochlorothiazide as a Sulfa Drug: Correctly noted that Hydrochlorothiazide is a sulfa drug, which is important since the patient has a sulfa allergy.

Case 2

- Correctly Identified Cross-Reactivity Between Penicillin and Cephalosporin: Successfully noted the potential cross-reactivity between Amoxicillin (Penicillin) and Cephalexin (Cephalosporin).
- Good Suggestion for Alternative Antibiotics: Provided a good recommendation for considering alternative antibiotics to which the patient has no known allergies.
- Reviewed NSAID Use in Diabetic Patients: Appropriately flagged the use of NSAIDs in diabetic patients due to potential kidney issues.
- Failed to Detect Overdosage of Cephalexin: Did not identify that the prescribed dosage of Cephalexin might be too high.

Case 3

- Correctly Suggested Adjusting Cephalexin Dosage: Provided correct recommendations for adjusting the dosage of Cephalexin based on additional information.
- Failed to Adjust Ibuprofen Dosage: Did not identify that the prescribed dosage of Ibuprofen (400 mg, 4 times daily) exceeds the maximum recommended daily dose of 1.2 grams.

In analyzing the AI's performance across these three cases, the AI demonstrated substantial capability in improving patient safety by identifying medication errors and suggesting appropriate alternatives. However, there were areas where the AI failed to

detect dosage errors, highlighting the need for further refinement and enhancement of the AI's capabilities in dosage verification.

Implementing Retrieval-Augmented Generation (RAG) can enhance the AI's accuracy by combining information retrieval with text generation. RAG integrates external knowledge bases, such as clinical guidelines and drug databases, into the decision-making process. This approach helps ensure that the AI's suggestions are grounded in verified information, reducing the likelihood of errors.

While RAG can improve accuracy and reduce errors, the system is not yet perfect. Domain-specific models or additional fine-tuning may be required. Fine-tuning the model on specialized medical data can help the AI better understand and interpret complex medical terminology and guidelines, further enhancing its ability to provide accurate and reliable suggestions.

4.3.3 Data Structuring, Interoperability and Data Exchange

- Cross-Hospital Data Sharing

Doctor D and the Developer A highlighted the challenge of reviewing patient histories across different hospitals. An AI-enabled Hospital Information System (HIS) could facilitate better data exchange and interoperability, allowing healthcare providers to access comprehensive patient records seamlessly.

- Standardization Issues

The Developer A mentioned the lack of standardized data storage structures, making it difficult to add new fields and customize the input interface for different departments. Implementing standardized formats such as Fast Healthcare Interoperability Resources (FHIR) can address these challenges by providing a consistent framework for data exchange.

4.3.3.1 Test Case: Data Structuring, Interoperability and Data Exchange

A mock up case in table 4.5 is used to demonstrate AI’s capabilities on data standardization. We utilized two state-of-the-art models, ChatGPT 4o and Claude 3 Opus, to convert patient records into standardized formats like FHIR.

Table 4.9 Testing Conversion of Patient Records into FHIR as a Proof of Concept

Patient Information:

- **Name:** Pichit Sukjai
- **Date of Birth:** 1982-03-10
- **Patient ID:** 789012
- **Allergies:** Sulfa drugs (difficulty breathing)

Visit Information:

- **Visit Date:** 2024-06-11
- **Diagnosis:**
  - Type 2 Diabetes Mellitus (E11)
  - Chronic Obstructive Pulmonary Disease (J44)

Medications Prescribed:

- Metformin 500 mg, take 1 tablet twice a day
- Salbutamol 100 mcg, inhale 2 puffs every 6 hours
- Metformin 500 mg, take 1 tablet twice a day
- Hydrochlorothiazide 25 mg, take 1 tablet once a day

Lab Results:

- Blood Sugar Level: 180 mg/dL (Normal range: 70-140 mg/dL)
- FEV1 (Forced Expiratory Volume in 1 second): 60% (Normal range: >80%)

Example of FHIR Conversion from AI

1. Patient Resource:

```
{
 "resourceType": "Patient",
 "id": "789012",
 "identifier": [
                {
                "use": "usual",
                "system": "http://hospital.org/patients",
                "value": "789012"
                }
 ],
 "name": [
                {
                "use": "official",
                "family": "Sukjai",
                "given": ["Pichit"]
                }
```

Table 4.9 Testing Conversion of Patient Records into FHIR as a Proof of Concept (cont.)


```
 ],
 "gender": "male",
 "birthDate": "1982-03-10",
 "extension": [
              {
              "url": "http://hl7.org/fhir/StructureDefinition/patient-birthTime",
              "valueDateTime": "1982-03-10T00:00:00+07:00"
              }
 ],
 "active": true
}
```

2. AllergyIntolerance Resource:

```
{
 "resourceType": "AllergyIntolerance",
 "id": "sulfa-allergy",
 "patient": {
              "reference": "Patient/789012"
 },
 "substance": {
              "coding": [
              {
              "system": "http://www.nlm.nih.gov/research/umls/rxnorm",
              "code": "Sulfa drugs",
              "display": "Sulfa drugs"
              }
              ]
 },
 "reaction": [
              {
              "manifestation": [
              {
              "coding": [
              {
              "system": "http://snomed.info/sct",
              "code": "422587007",
              "display": "Difficulty breathing"
              }
              ]
              }
              ]
              }
 ]
}
```

Table 4.9 Testing Conversion of Patient Records into FHIR as a Proof of Concept (cont.)

3.Condition Resource:


```
{
 "resourceType": "Condition",
 "id": "diabetes",
 "patient": {
              "reference": "Patient/789012"
 },
 "code": {
              "coding": [
              {
              "system": "http://hl7.org/fhir/sid/icd-10",
              "code": "E11",
              "display": "Type 2 Diabetes Mellitus"
              }
              ]
 },
 "verificationStatus": "confirmed",
 "onsetDateTime": "2024-06-11"
}
```


4. MedicationStatement Resource:


```
{
 "resourceType": "MedicationStatement",
 "patient": {
              "reference": "Patient/789012"
 },
 "medicationCodeableConcept": {
              "coding": [
              {
              "system": "http://www.nlm.nih.gov/research/umls/rxnorm",
              "code": "860975",
              "display": "Metformin 500 mg"
              }
              ]
 },
 "dosage": [
              {
              "text": "Take 1 tablet twice a day",
              "timing": {
              "repeat": {
              "frequency": 2,
              "period": 1,
              "periodUnits": "d"
              }
              },
```

Table 4.9 Testing Conversion of Patient Records into FHIR as a Proof of Concept (cont.)


```
                "route": {
                "coding": [
                {
                "system": "http://snomed.info/sct",
                "code": "26643006",
                "display": "Oral"
                }
                ]
                },
                "doseQuantity": {
                "value": 500,
                "unit": "mg",
                "system": "http://unitsofmeasure.org",
                "code": "mg"
                }
                }
 ]
}
```


5. Observation Resource:


```
{
 "resourceType": "Observation",
 "id": "blood-sugar",
 "status": "final",
 "patient": {
                "reference": "Patient/789012"
 },
 "code": {
                "coding": [
                {
                "system": "http://loinc.org",
                "code": "2339-0",
                "display": "Blood sugar level"
                }
                ]
 },
 "valueQuantity": {
                "value": 180,
                "unit": "mg/dL",
                "system": "http://unitsofmeasure.org",
                "code": "mg/dL"
 },
 "referenceRange": [
                {
                "low": {
```

Table 4.9 Testing Conversion of Patient Records into FHIR as a Proof of Concept (cont.)


```
                "value": 70,
                "unit": "mg/dL",
                "system": "http://unitsofmeasure.org",
                "code": "mg/dL"
                },
                "high": {
                "value": 140,
                "unit": "mg/dL",
                "system": "http://unitsofmeasure.org",
                "code": "mg/dL"
                }
                }
 ]
}
```


Table 4.10 Comparison of ChatGPT 4o and Claude 3 Opus

	ChatGPT 4o	Claude 3 Opus
Patient Information	● Using Patient Resource Correctly	● Using Patient Resource Correctly
Drug Allergy	● Using AllergyIntolerance Resource Correctly ● Wrongly use SNOMED-CT concept of difficulty breathing	● Using AllergyIntolerance Resource Correctly ● Using SNOMED-CT concept of sulfa drug allergy and dyspnea correctly
VIsit Information	● Fail to use Encounter Resource (using Condition Resource instead) ● Using ICD-10 code for diagnosis correctly	● Using Encounter Resource correctly ● Using SNOMED-CT concept for diagnosis correctly
Medications Prescribed	● Fail to use MedicationRequest Resources (using Medication Statement Resource instead) ● Using older version of medication CodeableConcept content ● Using RxNorm codes for Metformin correctly ● Using SNOMED-CT concept for route of administration correctly	● Using MedicationRequest Resource correctly ● Using older version of medication CodeableConcept content ● Partially using RxNorm codes correctly (2 out of 3 medications)

Table 4.10 Comparison of ChatGPT 4o and Claude 3 Opus (cont.)

	ChatGPT 4o	Claude 3 Opus
Lab Results	• Using Observation Resource correctly • Using LOINC Code Correctly	• Using Observation Resources correctly • Using LOINC Codes Correctly
Others	• Use separate resources without bundling	• Use Bundle Resource to combine resources together

4.3.3.2 Analysis of AI's Performance:

The table 4.9 compares the performance of ChatGPT 4o and Claude 3 Opus in using various FHIR resources and concepts. Key observations include:

- Patient Information: Both models use the Patient Resource correctly.
- Drug Allergy: Both models use the AllergyIntolerance Resource correctly, but ChatGPT 4o incorrectly uses the SNOMED-CT concept for difficulty breathing.
- Visit Information: ChatGPT 4o fails to use the Encounter Resource correctly, opting for the Condition Resource instead. Claude 3 Opus handles it correctly.
- Medications Prescribed: ChatGPT 4o fails to use the Medication Request Resource and uses an older version of medication Codeable Concept content, while Claude 3 Opus uses the Medication Request Resource correctly but only partially uses RxNorm codes correctly.
- Lab Results: Both models use the Observation Resource and LOINC codes correctly.
- Other Resources: ChatGPT 4o uses separate resources without bundling, whereas Claude 3 Opus uses the Bundle Resource to combine resources.

Potential Uses of LLMs in Data Structuring, Interoperability and Data Exchange

1. Facilitate Inputting and Structuring Data: Large Language Models (LLMs) have significant potential in assisting with the input of natural language data

and converting it into structured data using FHIR. This can facilitate better data exchange and support subsequent queries and research purposes.

2. Data Migration: LLMs can help in the migration of existing unstructured or less structured data into more standardized formats. This includes cleaning the data and ensuring it adheres to standardized structures like FHIR, which is crucial for interoperability and effective data management.

3. Medical Coding: LLMs can be utilized to facilitate accurate medical coding by interpreting clinical data and assigning appropriate ICD-10 codes. The AI's performance in medical coding can be evaluated by classifying coding outcomes into three categories—correct, partial, and wrong. In our test, out of 9 cases, the AI provided 4 correct, 2 partial, and 3 wrong coding outcomes. While the AI demonstrated the ability to provide accurate ICD-10 codes, there were instances where errors in medical knowledge led to inaccuracies, even when the correct codes were matched. This highlights the need for continuous refinement and validation of the AI's medical knowledge to ensure consistent accuracy.

- Current Limitations

1. Reliability Concerns:

A. Inconsistent Accuracy: Despite demonstrating the ability to provide accurate ICD-10 codes, the AI models occasionally produce incorrect or partially correct codes. This inconsistency can be attributed to errors in medical knowledge or the complexity of certain cases. Continuous refinement and validation of the AI's knowledge base are essential to enhance its accuracy.

B. Medical Knowledge Validation: AI-generated outputs need rigorous testing and validation. While AI can assist in data structuring and coding, healthcare providers must verify the accuracy and reliability of the information before using it in clinical decision-making. This requires a collaborative effort between AI developers and healthcare professionals to establish trust in AI systems.

2. Localization: Medical coding practices may vary by country, such as Thailand using the ICD-10 TM (Thai Modification). To ensure the AI's outputs are fully compatible with local coding standards, further localization and customization may be necessary.

3. Regular Updates and Use of RAG: To maintain accuracy and keep up with evolving standards like SNOMED-CT, which is updated quarterly, LLMs will require regular updates. Incorporating Retrieval-Augmented Generation (RAG) can help by integrating current knowledge and providing contextually relevant information during data processing. Regular updates and RAG integration are essential to ensure the AI stays current with medical standards and practices.

In summary, while AI-enabled HIS offers promising solutions for data structuring, interoperability, and data exchange, addressing the current limitations is critical for successful implementation and widespread adoption in healthcare. Continuous improvement, rigorous validation, and collaborative efforts between AI developers and healthcare providers are essential to overcome these challenges and harness the full potential of AI in healthcare.

4.3.4 AI-Augmented User Interface for Data Input

One of the key challenges identified by healthcare professionals was the cumbersome nature of data entry in current Hospital Information Systems. AI systems like Claude demonstrate potential in addressing this issue by generating user-friendly interfaces for clinical data input.

4.3.4.1 Test Case: Data Input

In a test scenario, Claude 3.5 Sonnet was asked to create an interface for inputting the ABCD2 score, which is used to assess stroke risk in patients who have experienced a transient ischemic attack (TIA).

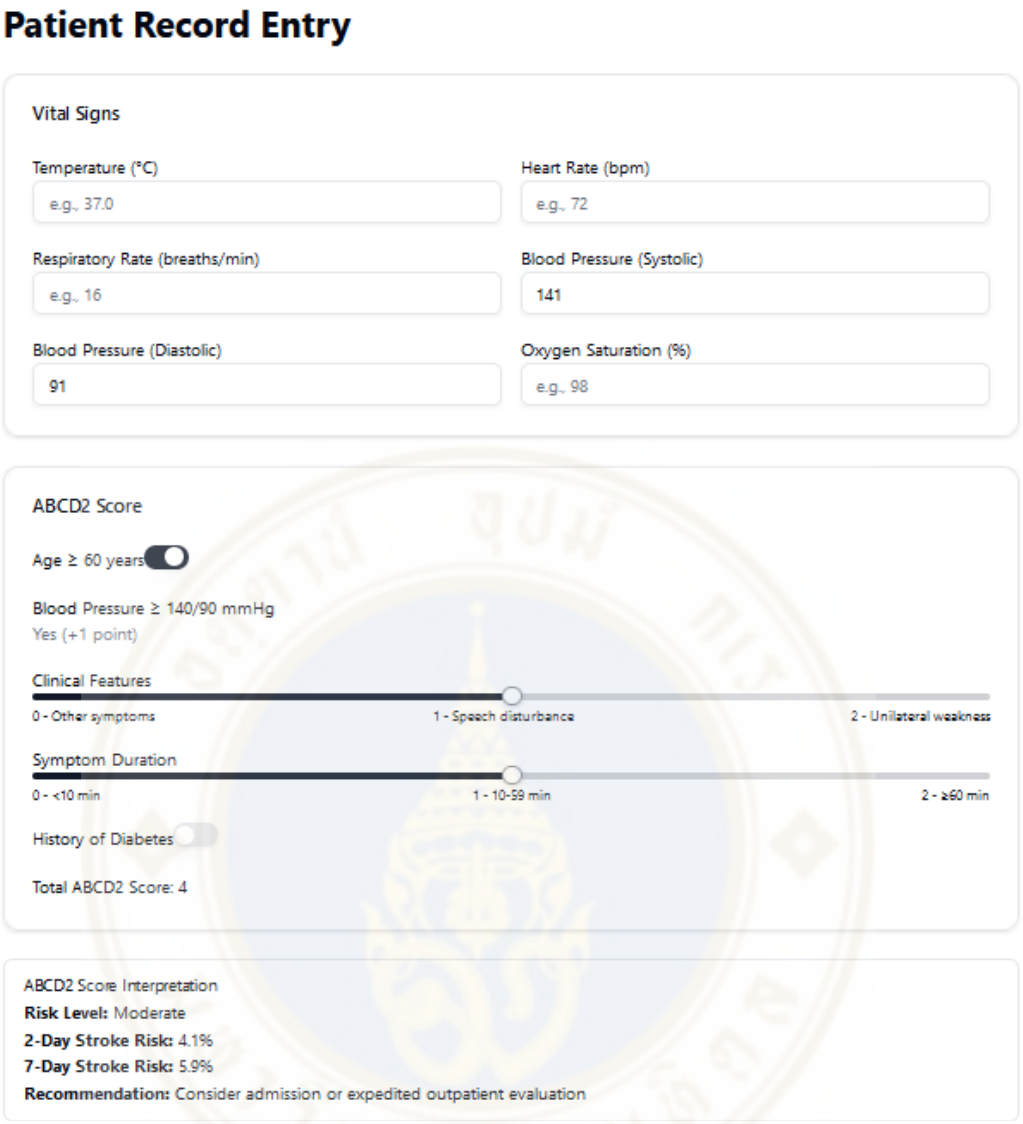

Figure 4.1 AI generated user interface with ABCD2 score for TIA patient

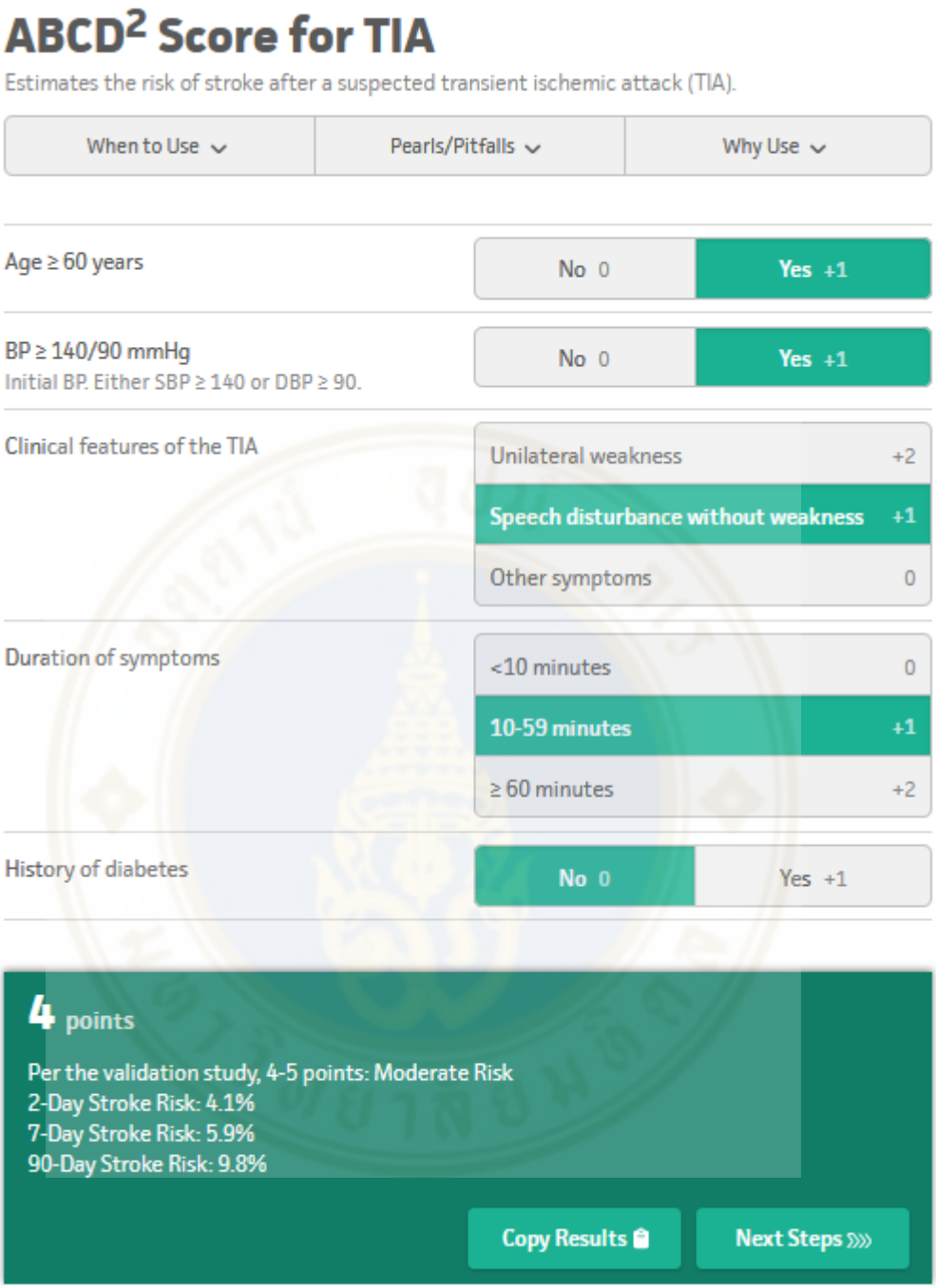

Figure 4.2 Figure 4.2 Reference of same setting ABCD2 score

source: https://www.mdcalc.com/calc/357/abcd2-score-tia

4.3.4.2 Analysis of AI's Performance:

The AI successfully generated a React component (figure 4.1) that includes:

- Relevant input fields for patient information (name, age, gender)
- Clinical data inputs (blood pressure, clinical features, duration of symptoms, diabetes status)
- A button to calculate the ABCD2 score
- Display of the calculated score and risk interpretation

The generated interface simplifies the data entry process by:

- Presenting only the necessary fields for the specific clinical task (ABCD2 score calculation) and automatically input blood pressure into the score without inputting it twice
- Using its UI of choice (sliding scale and toggle switches) for categorized inputs to reduce errors and standardize data entry
- Automatically calculating the score based on inputs, reducing manual computation errors
- Providing immediate risk interpretation, enhancing clinical decision support

Additionally, it's important to highlight that the AI-generated interface correctly implemented all components of the ABCD2 score calculation and also accurately interpreted the risk based on the calculated score as moderate risk (figure 4.2).

This example demonstrates how AI can help address the pain point of complex data entry by creating intuitive, task-specific interfaces. Such AI-generated interfaces could potentially:

- Reduce the time required for data entry
- Minimize errors in data input
- Improve standardization of data collection
- Enhance user satisfaction by simplifying complex clinical scoring systems

However, it's important to note that while AI can generate these interfaces, they would still require thorough testing and validation by healthcare professionals and UI/UX experts before implementation in a clinical setting. Additionally, the ability to

customize these interfaces for specific clinical needs and integrate them with existing systems would be crucial for widespread adoption.

This approach shows promise in addressing the data input challenges highlighted by healthcare professionals in the interviews, potentially leading to more efficient and accurate clinical documentation.

CHAPTER V
CONCLUSION AND RECOMMENDATION

5.1 Summary of Findings

This research aimed to explore the potential of Large Language Models (LLMs) in addressing the challenges associated with Hospital Information Systems (HIS). Through a combination of literature reviews, interviews with healthcare professionals, and proof-of-concept tests, several key insights and recommendations have emerged.

5.1.1 Pain Points in Current HIS

- Data Retrieval and Summarization: Healthcare professionals spend significant time reviewing patient histories and often encounter omissions. AI can aid by summarizing patient histories and assisting in reviews.
- Data Input Challenges: Manual data entry is time-consuming and prone to errors. Simplifying this process with AI can improve data quality and reliability.
- Accuracy and Error Prevention: Alerts for incorrect dosages and duplicate prescriptions, as well as error detection in data entry, are crucial for maintaining accurate patient records.
- Interoperability and Data Exchange: Standardizing data formats and improving cross-hospital data sharing are essential for cohesive patient care.
- Reliability and Privacy Concerns: Ensuring the consistent availability of AI systems and addressing data privacy and compliance issues are critical for safe AI integration.

5.1.2 Potential of LLMs

- Data Summarization and Thai Language Support: LLMs demonstrated strong proficiency in summarizing complex medical data and processing information in Thai, ensuring accurate and contextually relevant outputs for healthcare providers.

- Medical Coding: LLMs showed the ability to facilitate accurate medical coding. However, out of 9 test cases, the AI produced 4 correct, 2 partial, and 3 wrong coding outcomes, indicating room for improvement. Errors in medical knowledge and the need for localization to country-specific coding standards, such as ICD-10 TM for Thailand, were notable limitations.
- Accuracy and Error Prevention: The AI system successfully identified medication errors, albeit not perfect, such as duplicate prescriptions and incorrect dosages, and flagged potential issues with drug allergies. This capability enhances patient safety and improves the accuracy of medical records but still currently needs close supervision.
- Data Structuring, Interoperability, and Data Exchange: By converting patient records into FHIR format, LLMs facilitate seamless data exchange between different HIS platforms, data migration when changing HIS vendor, addressing standardization issues and improving cross-hospital data sharing.
- AI-Augmented User Interface for Data Input: LLMs demonstrated the ability to generate user-friendly interfaces for clinical data input, as shown in the ABCD2 score interface test case. The AI successfully created an intuitive interface with relevant input fields, automated calculations, and risk interpretation. This approach has the potential to simplify data entry, reduce errors, and improve standardization in clinical documentation.

5.1.3 Current LLMs Limitations

Despite the significant potential of Large Language Models (LLMs) in improving Hospital Information Systems (HIS), several limitations must be addressed to fully realize their benefits:

5.1.3.1 Accuracy and Reliability:

- Hallucinations: LLMs can generate factually incorrect or unsupported outputs, known as hallucinations. These inaccuracies can be particularly problematic in medical contexts where precision is critical.
- Inconsistent Performance: The accuracy of LLMs can vary, sometimes producing incorrect or partially correct outputs. This inconsistency can be due to errors in the AI's medical knowledge or the complexity of certain cases.

5.1.3.2 Knowledge Updating:

- Dynamic Nature of Medical Knowledge: The rapid evolution of medical knowledge presents a challenge for maintaining up-to-date AI models. Frequent updates are needed to ensure the AI reflects current medical standards and practices.
- Computational Demands: Updating LLMs is computationally intensive and resource-consuming, making frequent updates impractical and costly.

5.1.3.3 Data Privacy and Security:

- Memorization and Data Leakage: LLMs can unintentionally memorize and disclose training data, leading to potential privacy breaches. This is a significant concern in healthcare, where sensitive patient information is handled.
- Compliance with Regulations: Ensuring compliance with data privacy regulations, such as the Personal Data Protection Act (PDPA) in Thailand, is essential to protect patient information when integrating AI systems.

5.1.3.4 Localization and Customization:

- Adapting to Local Standards: Medical coding practices and standards vary by country. LLMs need to be localized and customized to ensure compatibility with local practices, such as the ICD-10 TM (Thai Modification).
- Language Support: Ensuring the AI supports the local language and accurately processes medical terminology in that language is critical for effective implementation.

5.1.3.5 Service Level Reliability:

System Downtime: AI systems must be reliable and consistently available. Any downtime or disruptions can negatively impact healthcare operations.

5.2 Implications for Healthcare

The integration of LLMs into HIS holds significant promise for improving healthcare delivery. By addressing the pain points identified, LLMs can enhance the efficiency, accuracy, and reliability of healthcare systems. This has the potential to:

- Improve Patient Care: Accurate data retrieval and error prevention can lead to better clinical outcomes and more cohesive patient care.

- Enhance Operational Efficiency: Automating administrative tasks can reduce the workload on healthcare professionals, allowing them to focus more on patient care.
- Facilitate Interoperability: Standardizing data formats and improving data exchange can lead to a more integrated healthcare system, enabling better coordination and continuity of care.

5.3 Recommendations

1. For Healthcare Organizations

- Consider implementing LLM-assisted tools for data entry and summarization to reduce healthcare providers burden, reduce waiting time and enhance efficiency.
- Utilize AI for preliminary medical coding, but ensure human verification for accuracy.
- Explore the use of LLMs for converting legacy data into standardized formats to improve interoperability.

2. For Medical LLMs and HIS Developers

- Focus on enhancing the accuracy and consistency of LLMs in medical contexts through specialized training and continuous updates.
- Develop robust validation processes to ensure the reliability of AI-generated outputs in clinical settings.
- Investigate the integration of Retrieval-Augmented Generation (RAG) to improve the contextual accuracy of LLMs in healthcare applications.
- Address the inconsistency in medical coding accuracy through more extensive training on diverse medical cases.
- Enhance localization efforts to better align with country-specific medical coding practices, such as ICD-10 TM in Thailand.
- Develop more sophisticated error detection capabilities, particularly for medication dosages and contraindications.

5.4 Suggestions for Further Research

1. Conduct trials of LLM integration in real hospital settings to assess practical impacts on workflow and patient care.
2. Investigate the effects of LLM assistance on healthcare provider efficiency and job satisfaction.
3. Explore the data privacy and security implications of using LLMs in HIS.

5.5 Conclusion

The integration of Large Language Models into Hospital Information Systems represents a transformative opportunity for healthcare. By addressing the identified challenges and leveraging the capabilities of LLMs, healthcare providers can enhance the quality of patient care, improve operational efficiency, and achieve better health outcomes. However, their integration requires careful consideration of accuracy, reliability, privacy and security concerns. Continued research and development, coupled with close collaboration between AI specialists and healthcare professionals, will be crucial in realizing the full potential of this technology in improving healthcare delivery and patient outcomes.

REFERENCES

Alberta College of Family Physicians & Alberta Medical Association. (2023). *Supporting comprehensive primary care in Alberta: Decreasing administrative burden*. Retrieved from https://acfp.ca/wp-content/uploads/2024/02/ACFP-AMA_Decreasing-Administrative-Burden-Report.pdf

Ayaz, M., Pasha, M., Alzahrani, M., Budiarto, R., & Stiawan, D. (2021). The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities. *JMIR Medical Informatics, 9* (7), e21929. https://doi.org/10.2196/21929

Berge, G. T., Granmo, O. C., Tveit, T. O., Munkvold, B. E., Ruthjersen, A. L., & Sharma, J. (2023). Machine learning-driven clinical decision support system for concept-based searching: a field trial in a Norwegian hospital. *BMC Medical Informatics and Decision Making, 23*(5). https://doi.org/10.1186/s12911-022-01921-5

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Baghery, S., ... & Song, D. (2021). Extracting training data from large language models. In *Proceedings of the 30th USENIX Security Symposium* (pp. 2633-2650).

Das, B. C., Amini, M. H., & Wu, Y. (2024). Security and privacy challenges of large language models: A survey. arXiv preprint arXiv:2402.00888 . https://doi.org/10.48550/arXiv.2402.00888

Ji, Y., Yu, Z., & Wang, Y. (2024). Assertion detection in clinical natural language processing using large language models. arXiv. https://arxiv.org/pdf/2401.17602

Karabacak, M., & Margetis, K. (2023). Embracing large language models for medical *applications: Opportunities and challenges. Cureus, 15*(5), e39305. https://doi.org/10.7759/cureus.39305

Kwon, S., Yao, Z., Jordan, H., Levy, D., Corner, B., & Yu, H. (2022). MedJEx: A Medical Jargon Extraction Model with Wiki's Hyperlink Span and Contextualized Masked Language Model Score. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 11733–11751). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics.

REFERENCES (Cont.)

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. arXiv preprint arXiv:2005.11401v4.

Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: A review of recent research. *IMIA Yearbook of Medical Informatics*, 2008, 128-144.

Michaels, M., Syed, S., & Lober, W. B. (2021). Blueprint for aligned data exchange for research and public health. *Journal of the American Medical Informatics Association, 28*(12), 2702-2706. https://doi.org/10.1093/jamia/ocab210

Mosbach, M., Pimentel, T., Ravfogel, S., Klakow, D., & Elazar, Y. (2023). Few-shot fine-tuning vs. in-context learning: A fair comparison and evaluation. Findings of ACL 2023 . arXiv:2305.16938.

Nawab, K., Fernbach, M., Atreya, S., Asfandiyar, S., Khan, G., Arora, R., Hussain, I., Hijjawi, S., & Schreiber, R. (2024). *Fine-Tuning for Accuracy: Evaluation of GPT for Automatic Assignment of ICD Codes to Clinical Documentation.* Research Square. https://doi.org/10.21203/rs.3.rs-4392229/v1

Ong, J. C. L., Jin, L., Elangovan, K., Lim, G. Y. S., Lim, D. Y. Z., Sng, G. G. R., Ke, Y., Tung, J. Y. M., Zhong, R. J., Koh, C. M. Y., Lee, K. Z. H., Chen, X., Chng, J. K., Than, A., Goh, K. J., & Ting, D. S. W. (2024). Development and Testing of a Novel Large Language Model-Based Clinical Decision Support Systems for Medication Safety in 12 Clinical Specialties. arXiv. https://arxiv.org/abs/2402.01741

OpenAI. (2022, November 30). Introducing ChatGPT. Retrieved June 17, 2024, from https://openai.com/index/chatgpt/

Reisman, M. (2017). EHRs: The Challenge of Making Electronic Data Usable and Interoperable. *P&T, 42*(9), 572-575. PMID: 28890644; PMCID: PMC5565131.

REFERENCES (Cont.)

Saba, W., Wendelken, S., & Shanahan, J. (2024). Question-Answering Based Summarization of Electronic Health Records using Retrieval Augmented Generation. *arXiv.* https://arxiv.org/abs/2401.01469

Stewart, B. A., Fernandes, S., Rodriguez-Huertas, E., & Landzberg, M. (2010). A preliminary look at duplicate testing associated with lack of electronic health record interoperability for transferred patients. *Journal of the American Medical Informatics Association, 17*(3), 341-344.

Tonmoy, S. M. T. I., Zaman, S. M. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models. *arXiv.* https://arxiv.org/abs/2401.01313

Van Veen, D., Van Uden, C., Blankemeier, L., & Delbrouck, J.-B. (2023). Clinical text summarization: Adapting large language models can outperform human experts. Preprint . https://doi.org/10.21203/rs.3.rs-3483777/v1

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In A*dvances in Neural Information Processing Systems* (Vol. 30).

Wager, K. A., Lee, F. W., & Glaser, J. P. (2017). *Health Care Information Systems: A Practical Approach for Health Care Management.* Jossey-Bass.

Wang, G., Yang, G., Du, Z., Fan, L., & Li, X. (2023). *ClinicalGPT: Large Language Models Finetuned with Diverse Medical Data and Comprehensive Evaluation.* arXiv preprint arXiv:2306.09968.

Wong, M., Lim, Z. W., Pushpanathan, K., Cheung, C. Y., Wang, Y. X., Chen, D., & Tham, Y. C. (2023). Review of emerging trends and projection of future developments in large language models research in ophthalmology. British Journal of Ophthalmology. https://doi.org/10.1136/bjo-2023-324734

Zhang, D., Yin, C., Zeng, J., Yuan, X., & Zhang, P. (2020). Combining structured and unstructured data for predictive models: a deep learning approach. BMC Medical Informatics and Decision Making, 20 (280).

REFERENCES (Cont.)

Zhang, N., Yao, Y., Tian, B., Wang, P., Deng, S., Xi, Z., Mao, S., Zhang, J., Ni, Y., Cheng, S., Xu, Z., Xu, X., Gu, J.-C., Jiang, Y., Xie, P., Huang, F., Liang, L., Zhang, Z., Zhu, X., & Zhou, J. (2024). *A Comprehensive Study of Knowledge Editing for Large Language Models*. arXiv. https://arxiv.org/abs/2401.01286

APPENDIX

Appendix A: Evolution of Hospital Information Systems with LLM Integration

This appendix presents a comparative view of traditional Hospital Information Systems (HIS) and a conceptual framework for HIS enhanced by Large Language Models (LLMs).

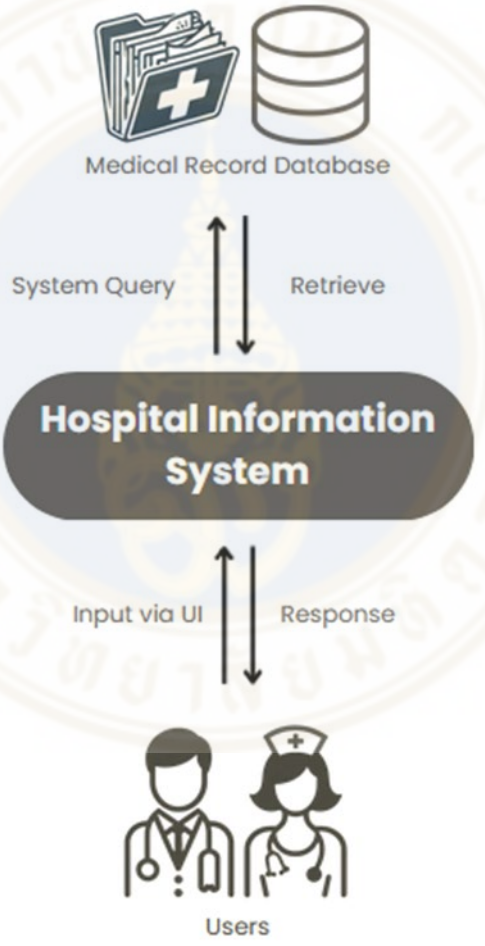

Figure A1 Traditional Hospital Information System

Figure A1 illustrates the basic structure and flow of a traditional Hospital Information System:

1. Medical Record Database: Represented by a medical folder icon and a database cylinder, storing structured or unstructured patient information.

2. Hospital Information System: The central component that manages data flow and user interactions.

3. Users: Healthcare professionals (doctors and nurses) who interact with the system.

4. Data Flow:

- System Query and Retrieve: Bi-directional data exchange between the HIS and the database.
- Input via UI and Response: Direct interaction between users and the HIS.

This traditional model, while functional, faces challenges in data retrieval, interpretation, and decision support, as identified in our research.

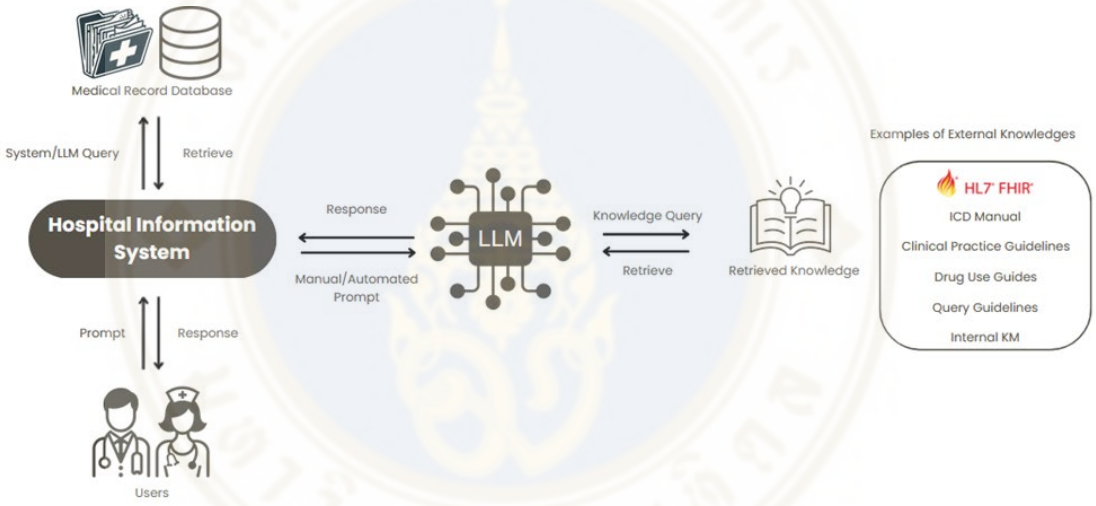

Figure A2 Conceptual Framework for LLM-Enhanced Hospital Information System

Figure A2 presents a conceptual framework for integrating LLMs into HIS, addressing challenges identified in our study:

The framework illustrated in Figure A2 envisions how LLMs could potentially interact with various components of the healthcare information ecosystem. Key elements include:

1. Medical Record Database: Represents the storage of patient data, addressing data management challenges in HIS.

2. Hospital Information System: Acts as the core interface for healthcare professionals, integrating LLM capabilities.

3. Users: Depicts healthcare professionals interacting with the HIS.

4. LLM: Centrally positioned, illustrating its potential role in facilitating various processes within the HIS.

5. External Knowledge Sources: Includes important medical references and standards, addressing the need for integrated knowledge bases.

Appendix B: Simplified Data Flow Visualization for Proof of Concept

One of the key applications of LLMs in HIS identified in our proof of concept is the organization and curation of medical records. Figure B1 illustrates this process.

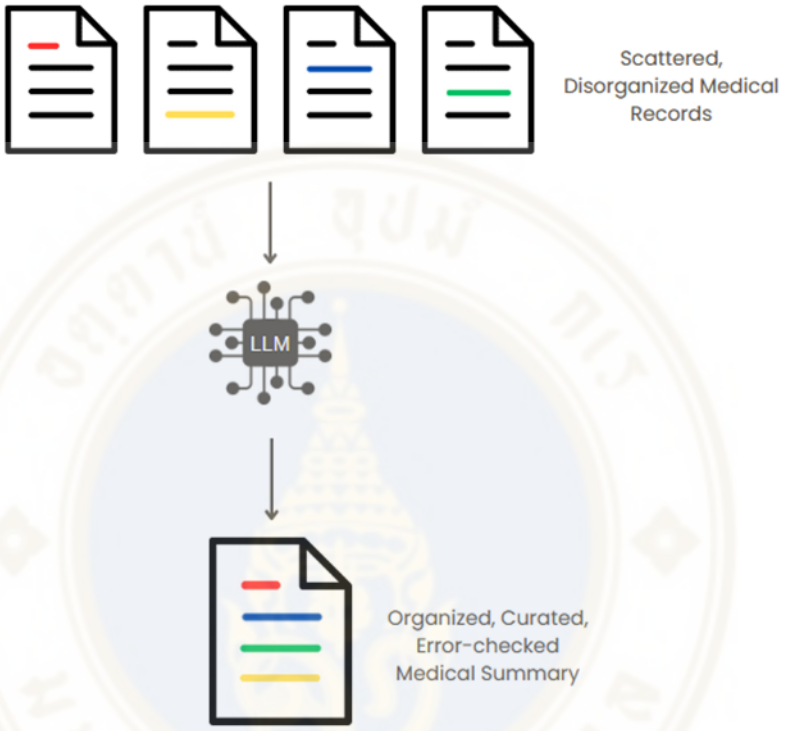

Figure B1 LLM-Assisted Organization of Medical Records

As shown in Figure B1, LLMs have the potential to transform scattered, disorganized medical records into a single, organized, curated and error-checked document.

Another significant application of LLMs demonstrated in our proof of concept is automated medical coding. Figure B2 depicts this process.

Figure B2 LLM-Assisted Medical Coding

Figure B2 illustrates how an LLM can potentially convert natural language medical records into standardized reimbursement codes. This directly addresses the pain point highlighted by Doctor A regarding the difficulty and time-consuming nature of manual ICD-10 coding.

Figure B3 LLM-Assisted Standardization of Medical Records

Figure B3 demonstrates a crucial application of Large Language Models (LLMs) in standardizing medical records:

1. Input: Natural Language Medical Records This represents the typical unstructured or semi-structured medical records that healthcare providers often work with. These records may include free-text notes, varied formats, and non-standardized terminologies.

2. Process: LLM The Large Language Model serves as the central processing unit, capable of understanding and interpreting the nuances and context of natural language medical information.

3. Output: Structured Medical Record (e.g., HL7 FHIR) The LLM transforms the input into a structured format, specifically mentioning HL7 FHIR (Fast Healthcare Interoperability Resources) as an example. FHIR is a standard for exchanging healthcare information electronically, promoting interoperability between different healthcare systems.

Figure B4 LLM-Assisted Error Correction in Medical Data

As illustrated in Figure B4, LLMs show potential in identifying and correcting errors in medical data:

1. Input: Wrong Medical Data containing medication errors
2. Process: LLM analysis
3. Output: Corrected Medical Data with suggestions for error correction