MS Thesis I Proposal

Patient Medical History Generation using AI



Student Name: Mairaj Uddin

Roll Number: 23k-8078

Supervisor Name: Syed Farooq Zaidi

National University of Computer and Emerging Sciences

October 10, 2024

**Abstract**

The rise of artificial intelligence (AI) is transforming healthcare significantly, which revolutionizes healthcare professionals' interactions with patients and documenting medical records. In this thesis project, I aim to develop an AI-based medical record summarization system for patients that will generate summarization by checking their past medical history, generally known as Electronic Health Records (EHR), which are patients' medical records saved in a digital form. The medical record summarization system known as “Patient Medical History Generation using AI (PMHG-AI) combines natural language processing and machine learning algorithms to generate short summaries that will be accurate by using the takeover quality metrics to reliably summarize patients' medical histories, which will help healthcare professionals save time and attend more patients. The PMHG-AI will be able to store patient medical history without relying on the clinician provider to insert the information manually and allow them hours of time to interact in the clinical process, focussing on communications including diagnosis, treatment, or patient suggestions, which our system inherently promotes the future need for patients to become more engaged in the healthcare delivery process.

PMHG-AI's abilities will extend further than EHR summarization, empowering the system to create new patient records by asking targeted questions, reducing the review time and effort required for healthcare professionals to gather essential medical information. This method will support the productivity of healthcare providers, speed up decision making and improve the quality of patient care by validating that critical medical records are correctly captured and improve accessibility.

**Introduction**

Over the last few years, a wide range of AI applications in healthcare have been created to increase the quality of healthcare, save time, and improve patient outcomes, using AI-powered systems for diagnostics, treatment planning, patient monitoring, etc. AI-powered chatbots for personalized healthcare advice and AI-powered imaging analysis for disease detection like cancer and Alzheimer's. The motivation for this research comes from the inherent challenges faced when creating manual EHRs and analyzing them can be time consuming and can have numerous errors and inaccuracies, the aim of this research is to develop a system powered by artificial intelligence (AI) for accurate electronic health records (EHRs) which enables healthcare professionals to summarize patient medical histories and store all EHRs at one place to support healthcare professionals in reducing time taken by traditional EHRs and making them paper less, accurate and reliable information, we will study the usage of generative AI in medical practices, how to diagnose and treat also what are its limitations that comes when implementing the system [1]. The system will use natural language processing (NLP) techniques like Tokenization, Named Entity Recognition, Part of speech Tagging, Dependency Parsing and Deep Learning models like Pegasus, GPT 4, and BART to analyze patient data and create an error free and well organized summary of their medical history, and if a patient has no medical record previously, it can create it by asking the targeting questions to save time and efforts.

The proposed research project is used to enhance the effectiveness and efficiency of EHR generation, thereby enabling physicians to focus on complex, high-value activities. EHRs contains a patients detailed medical summary record in a digital format that helps healthcare workers to make decisions about patient’s health and give complete medical assistance. In this study we will observe how NLP techniques help patient and physicians to share information seamlessly and potential challenges that rises in advancement of healthcare environment [2]. We aim to collect and use our own dataset form hospital for training and testing of AI models, we will also be using the publicly available datasets for training and testing of the model, our aim is to helps healthcare professionals to reduce review time and improve overall accuracy while focussing on patient centred care. The current EHR system relies on manual data entry, which can be time-consuming and increases the chances of errors or inaccuracies. Physicians spend most of the time creating, reviewing, and updating medical records to understand patient diseases. However, the present EHR system is not intelligent enough and is not able to summarize the medical records by analyzing past medical history, give recommendations to healthcare professionals accelerate decision making, and reduce the risk of human error.

**Literature Review**

AI has altered the approach of electronic health records quite much to the clinical settings efficiency. In a move to retrieve relevant information from databases and summarize some certain clinical information from the electronic health records, one study expressed into investing the knowledge of retrieval augmented generation (RAG) method in combination with generative AI methods to target only particular section of information that is asked and search only that division [3]. This research clearly describes that the EHRs and generative AI in clinical documentation or notes are still up to date with current issues; thus, forming an intelligent system by using synthetic data of clinical information to check model accuracy. It utilizes both structured and unstructured data and can save administrative work, time, and result in flawless documentation for healthcare professionals [4]. To understand the intensity of patient-centric clinical decision support system based on patient profile, medical records or lifestyle, provides recommendation using AI, the model is trained with EHRs, clinical guidelines, and genomic data using both supervised and unsupervised algorithms. It also collects feedback from patients to make a robust decision in respect of improving the precision of clinical recommendation making a robust decision process better and more reliable data protection measures implemented in clinical workflows [5].

Improving with the day-to-day advancement in AI in healthcare, automating discharge summaries saving time and supporting clear communication among healthcare professionals, focusing on seam less workflow, this model was trained on a dataset of existing discharge summaries from multiple clinical parameters with a wide range of clinical context and uses NLP techniques for achieving that but paying heed to the clarity and completeness of clinical guidelines [6] and meeting or perhaps exceeding quality produced by physicians to help reduce the amount of documentation for physicians. Further, presents a novel approach for LLMs to develop complex models based both on structured and unstructured data that may acquire useful information from EHRs. The author utilizes a personalized framework called EHR agent, which has been trained to retrieve useful information from EHRs, for instance, personalized treatment plans and answers to any query regarding reports. This illustrates that AI can be applied in the processing and analytics of complex data structures in medical records for high-order tabular reasoning over EHRs [7]. Despite the mentioned research, many questions still hang around on how AI will be implemented to construct a comprehensive patient's medical history in real-time for a particular hospital, particularly in patient-centric, and current studies mostly focus on automating clinician's documentation and reducing paperwork for saving time.

**Research Gap and Research Question**

The literature reflects a broad gap in comprehensive use of AI for generation of medical histories for patients. While the completed studies were focused on the automation of clinical documentation and hence better administrative efficiencies, they often missed aspects of real-time patient data and synthesis of structured and unstructured data into coherent medical histories also their entities according to region. For instance, even though RAG-based methods are studied for summarizing clinical data, few works link them to the real-time update aspects concerning patients. Moreover, though patients can provide their feedback, the facility does not prove to serve with even the finest adjustments for individual medical histories. Finally, the quality and completeness of AI-generated patient medical records summaries are still underexplored some of them are under development. Their experimental effect on the related clinical choices or decisions is inadequate.

To fill those gaps, the following research questions have been formulated for this study: how AI techniques may be integrated towards generating real-time comprehensive patient medical histories from both structured and unstructured data available in EHRs, and what frameworks may be developed to ensure that patient feedback has been considered properly in the updating of these histories. How does using AI-generated medical histories affect clinical decision-making and patient outcomes? Moreover, how can AI make patient medical histories clearer more comprehensive in a way that upholds data protection and confidentiality? All these will contribute significantly to the field in terms of adding more value to patient-centered care and automating clinical workflows.

**Problem Statement**

Considering that my country has yet to develop systems of comprehensive creation of summaries using patient histories using AI, it leads to fragmentation and incompleteness in medical records. EHR systems have tremendous regional variations. I wish to create an AI-driven solution that utilizes the dataset of medical records collected from hospitals including structured and unstructured data of patients for the generation of organized patient summaries, thus providing quality care for patients, reducing paperwork saving time and supporting clinical decision-making.

**Project Timeline**

**MS Thesis**

**2024**

**Sep**

**Oct**

**Nov**

**Feb**

**Mar**

**Apr**

% complete

100

% complete

100

% complete

50

% complete

30

0

% complete

0

% complete

0

% complete

**2025**

**Step 1 Thesis I**

Proposal Writing

Proposal Plagiarism Checking

*Proposal Submission*

**Literature Review**

Literature Search Model Architecture

**Data Collection**

Pre-Processing and Feature Engineering

Analyzing Multiple Model for best Performance

**Thesis I Submission**

**Step 2 Thesis 2**

Model Training and Evolution

NLP Techniques for AI Summarization

**Thesis II Writing**

UI/UX design and Model Deployment

% complete

000

Report Presentation

0

% complete

Plagiarism Checking

0

% complete

*Thesis II Submission*

**Figure 2.** Gantt Chart for MS Thesis I.

**Block Diagram**

Failure

Success

Model Deployment

Performance Evaluation

Fine Tuning Model Performance

Documents Layout Understanding

Text Splitter, Text Extraction Data Transformation

Generated Structure Summary of Sections and AI Advice

Four Training Models

BART

GPT 4

Lama 3

Pegasus

Current Symptoms

Medical History

**References**

1. Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. (2024). Generative AI for transformative healthcare: A comprehensive study of emerging models, applications, case studies and limitations. *IEEE Access*.
2. Sarella, P. N. K., & Mangam, V. T. (2024). AI-driven natural language processing in healthcare: Transforming patient-provider communication. *Indian Journal of Pharmacy Practice*, 17(1).
3. Alkhalaf, M., Yu, P., Yin, M., & Deng, C. (2024). Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. *Journal of Biomedical Informatics*, 104662.
4. Biswas, A., & Talukdar, W. (2024). Intelligent clinical documentation: Harnessing generative AI for patient-centric clinical note generation. *arXiv preprint arXiv:2405.18346*.
5. Sanchez, R. (2024). AI-powered clinical decision support systems for precision medicine: Developing AI-powered systems to provide clinicians with decision support tailored to individual patient characteristics. *Journal of Bioinformatics and Artificial Intelligence*, 4(2), 70-79.
6. Clough, R. A. J., Sparkes, W. A., Clough, O. T., Sykes, J. T., Steventon, A. T., & King, K. (2024). Transforming healthcare documentation: Harnessing the potential of AI to generate discharge summaries. *BJGP Open*, 8(1).
7. Shi, W., Xu, R., Zhuang, Y., Yu, Y., Zhang, J., Wu, H., Zhu, Y., Ho, J., Yang, C., & Wang, M. D. (2024). Ehragent: Code empowers large language models for complex tabular reasoning on electronic health records. *arXiv preprint arXiv:2401.07128*.