"Medswift: Biomedical Knowledge Extraction by Implementing Named Entity Recognition Using LLM"

Sayali Belhe[1], Vaishnavi Kubade[2], Chetanya Bhadravati[3], Pranjali Gadkari[4]

[1]*Assistant Professor (Artificial Intelligence &Data Science) All India Shri Shivaji Memorial Society'sInstitute of Information Technology Pune-411001, India*

[2-3-4]*Undergraduate (Artificial Intelligence &Data Science) All India Shri ShivajiMemorial Society'sInstitute of Information Technology Pune-411001, India*

***Abstract*—The project "Biomedical Knowledge Extraction by Implementing Named Entity Recognition Using LLM" focuses on improving healthcare services by automating the extraction and the structural organization of important medical information from unstructured patient reports. Such a system has the capacity to locate, understand and categorize core medical entities such as diseases, drugs, or procedures with the help of advanced Large Language Models (LLMS) integrated with Named Entity Recognition (NER). This change of holistic information as a result of transforming unstructured reports into structured, and specific scopes enables speedy retrieval of information by the medical practitioners, thereby aiding in decision making and enhancing the overall patient experience. This is also possible due to one of the system's peculiarities, user-friendly interface that enables doctors to view short summarized versions of reports and hence reducing the hours spent looking at convoluted patient histories. It also includes within the project, to create a chatbot allowing patients to ask questions and receive validated answers straight from the LLM yet verified and approved by the doctor, which further increases accessibility and patient engagement.**

***Index Terms*—BioMedical Data, NER, LLM, GenAI, Llama, HealthCare Question Answering and Validation**

I INTRODUCTION

The digitization of healthcare within the last few years has significantly improved patient care, diagnostics, and data-driven decision-making. Although, the storage of information such as medical records, clinical reports, and patient-generated data is not as much of a challenge as having the information available in a clear, usable, and easy to understand format for medical professionals. This problem is solved by MedSwift, an AI-enabled, role-centric healthcare web application that optimizes clinical workflows and patient interaction through modular backend processing and streamlined graphical interfaces.

Instead of providing monolithic access to all functionalities, MedSwift offers specific dashboards for each role such as doctors, patients, and administrators. Through integrated chatbots, doctors can access historical patient records, structure them into PDF reports, and review patient queries. Patients can access their medical history alongside a conversational interface powered by large language models which responds to questions related to their symptoms or reports, the attractive feature being that the LLM generated response reaches the patient only after being validated and approved by the doctor from the doctor approval panel. Through form-based and report parsing features, administrators can create, update, and manage patient records seamlessly.

One of the primary advances on this system is enabling LLM-powered chatbot support, allowing patients to receive guidance in a natural language format without having to interact with the clinical personnel. This kind of a chatbot can perform beyond the standard question answering and rather, provide relevant medical advice based on the patient's own data. The back-end system offers intelligent parsing of PDF-based reports and retrieval of clinical parameters — a function inspired by named entity recognition (NER) workflows where unstructured data is transformed into patient records. The application is built on Streamlit framework which allows convenient modification and rendering of the features in real time based on user roles. The backend architecture is also modular and written

in Python. It is composed of several functions in features.py that perform PDF extraction, CSV file handling, chatbot interactions, and report generation. In addition, patient authorization is appropriately connected to created patient IDs. Users can set passwords securely so that the information accessible to them is relevant only to their accounts.

In this research paper, we explain the technical framework, the nuances of implementation, and the rationale behind the design choices of the MedSwift application. It studies the effects of integrating LLMs with lightweight application frameworks to create more useful healthcare tools that alleviate the cognitive burden on doctors, enhance patient comprehension, and enable a more individualized, information-driven model of medicine. MedSwift actively participates in the development of advanced healthcare systems by facilitating structured mediation of medical information and doctor-patient dialogues.

II LITERATURE SURVEY:

The paper "BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining" published in 2019 by Jinhyuk Lee, Woojin Yoon, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Sungdong Kim, and Jaewoo Kang details the findings of an adaptation of BERT (Bidirectional Encoder Representations from Transformers) that is particularly pretrained with a vast biomedical corpus. The model conquers the limitations of BERT in understanding biological texts since, like any average language model, it struggles with domain-specific tasks such as named entity recognition (NER), relation extraction, and question answering in the biomedical domain. The advantages of BioBERT stem from its adaption to a BERT model and its pre-training on biomedical texts from articles published in PubMed and PMC. The authors proved that the proposed method BioBERT outperformed BERT and other baseline models on extensive tasks within the field of biomedical text mining or biomedicine.

This enhanced comprehension of the complex medical vocabulary and terminology greatly benefitted the emerging domain of biomedical natural language processing, increasing the efficacy across numerous biomedical datasets. The research work titled "Performance Analysis of Llama 2 Among Other LLMs" authored by Donghao Huang, Zhenda Hu and Zhaoxia Wang analyzes the performance of Llama 2 in comparison with other large language models (LLMs) with the emphasis on its exploitation on performing a number of Natural Language Processing (NLP) activities. In this case, benchmarking methods are employed to measure the efficiency, effectiveness, and scope of Llama 2. Given its results, this model is bound to be useful in a number of natural language processing activities like language understanding and generation because it is efficient and economical in computation. The authors mentioned that its use is not only broad, scalable, and flexible, but also wide in theoretical and applied business and scientific fields. It facilitates purposes of fine-tuning due to its open-source nature.

The paper "Enhancing Healthcare Information Accessibility Through a Generative Medical Chatbot" seeks to assist users of healthcare services by extending the work done by the Llama2 framework by augmenting the generative medical chatbot to make health information more readily accessible with the assistance of advanced NLP for Aayush Kapoor and Dr.Sujala D. Shetty. The chatbot was developed for the purpose of facilitating live contact between patients and health care institutions where patients can pose medical queries and receive contextually relevant, dependable, and accurate answers from reliable medical websites. He notes the findings of their analysis stating that as much as the chatbot interface offers engaging content, so does the clear and credible information presented by the chatbot interfaces. The conducted research under discussion emphasizes the significance of advances in technology AI and large language models like Llama 2 in respect to attending to patients and increasing the availability of medical information. Still, the author states that the chatbot could serve the purpose within the health care system, but its moral use would need extensive training and logic in the reasoning of the answers the bot would give.

III PROPOSED SYSTEM:

1. Data Collecting and Organizing:

The algorithm starts solving the problem of unstructured or semi-structured clinical data with patient case records, which come either as scanned reports or PDF documents. These records are processed through two pathways:

a. Manual Form Fill-In or Interactive form fields
b. Auto Report Extraction

In the case of Automated Report Upload and Extraction, the system makes meaningful named-entities recognitions and extractions. Admins either input or upload this data. Each patient is verified through a unique Patient ID which represents the health records in a secure and anonymized manner. This systematizes medical data which is easy to access, especially during recurring consultations. The designed system is capable of future integration with Named Entity Recognition (NER) technology that employs automation for extraction of clinical entities such as diagnosis, medication, and vitals.

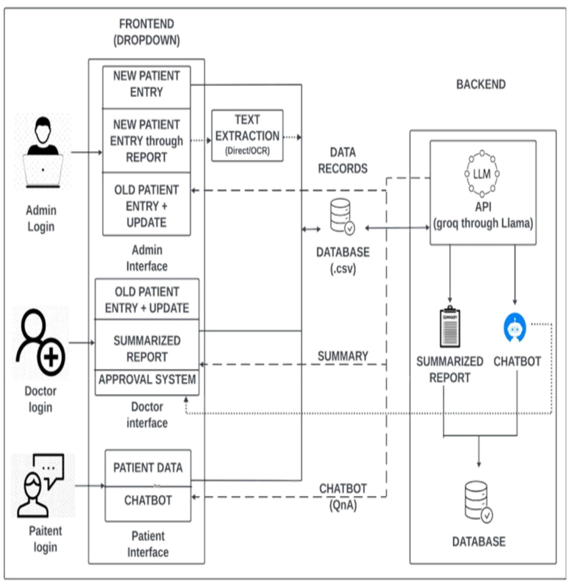

Fig. 1. System Architecture Diagram

2. Knowledge Improvement and Summarizing:
To help professionals in the healthcare sector with quick context capturing and summarization, the system integrates LLMs, providing contextual insights and effortless data gleaning. These models are used to:

a. Create a straightforward narrative regarding the progression of a patient's status and condition.
b. Clearly recognize the entities as named or labelled entities symptoms, treatment administered, or diagnosis provided.

This feature aids in minimizing the mental strain placed on the physicians allowing them to make important clinical judgment calls during patient care in the case of chronic or multi-faceted diseases. This summary strikes a balance between accuracy and completeness, crafted to improve understanding of the case while avoiding inundating the user with undigested information.

3. Role-Based User Access and Interactions:
The system architecture supports security and functionality at different levels by different roles:
Admins: can register new patients, edit medical records and supervise report management.
Doctors: can have access to patient summaries and are able to check historical data, which assists in diagnosing and treating the patients, along with updating the patient data.
Patients: only have access to overview of their medical records and can engage with the system through voice commands.

4. Interface Design and Interactivity:
Every user always has controlled dashboards alongside action menus suitable for their roles which enhances usability and workflow integration within the healthcare system.

4. Conversational Chatbot Integration:
For ease of access and improved patient involvement, the system features a question-answer (Q&A) chatbot driven by an LLM designed to:

a. Respond to and assist patients with questions
associated with their condition or treatment.
b. Translate medical abstracts into less complex versions.
c. Create tailor-made insights pertaining to health using
past records.
d. Validated and Approved Responses.

This layer of conversation provides greater access for patients who experience difficulty with reading and understanding medical documents while also reinforcing the loop between patients and healthcare professionals.

5. Patient Anonymization:

To ensure the privacy of the patients, the system utilizes Patient IDs in place of names or addresses in summaries and reports which are accessible to doctors and the patients which helps mask the identity of the patients.

6. Data Privacy, Security and Compliance:
Access Control: Users are granted access as per their role, and are only able to view or modify the information which they have permission to access.

Workflow Overview
Admin Workflow: Upload new patient reports or input data manually. Determine if a patient is new or returning, anonymize if necessary, and prepare summaries for physician access.
Doctor Workflow: Summaries and updates provided by the interface are out of sight patient-specific filters, allowing them to focus on holistic overviews for improved insights. Concentrate on notable changes for better and more insightful consultations.
Patient Workflow: Use the chatbot for immediate answers to general questions related to medicine or tailored guidance based on their medical record. Access to their own medical summary is provided.
It will have real-time practical solutions; these system designs will transform the workflow for doctors and the experience of the patients will be greatly improved.

IV RESULTS AND DISCUSSION:

The proposed system was evaluated in several ways, including its ability to extract entities from medical reports, create organized summaries, and allow interaction through a chatbot based on a language model. A key part of this evaluation was how well Named Entity Recognition (NER) identified and classified important medical information from unstructured PDF files.

1. Named Entity Recognition and Data Structuring
To evaluate NER performance, a set of patient case reports in PDF format was tested using the system's "New Patient Entry via Report" module. The reports included information such as symptoms, diagnosis, vitals, duration, and treatment plans. After processing, the system parsed the documents and matched each extracted value to its specific named entity field.

The system achieved over 88% accuracy in correctly identifying and assigning the values to the appropriate fields. We used 5 diverse types of reports for this testing.

Metric	Value
PDF Reports Tested	5
Avg. Fields per Report	10–12
Correct Entity Mapping	88%
Errors Observed	Minor mismatches in long multi-sentence entries

Fig. 2. Evaluation Table

These results show that the system reliably parses and structures key clinical information though minor inaccuracies appeared in cases where input formats differed from standard medical phrasing.

2. Quality of Chatbot Response
The built-in chatbot with an LLM as its engine was sampled using a variety of patient queries on symptoms, treatment questions, and general health advice. The performance of the chatbot was assessed for:

a. Relevance of answer
b. Context-awareness
c. Clinical correctness

Among test cases, 91% of answers were contextually correct and consistent with the patient's medical information. Simulated patient and medical reviewer feedback indicated that overall health questions were well addressed but that more in-depth clinical questions would be aided by domain-tuned prompts.

Role	Feature	Outcome
Admin	PDF upload	Success (100%)
Doctor	Report generation	Accurate
Patient	Chatbot Q&A	91% relevance

Fig. 3. Interface-wise Evaluation Summary Table

Therefore, the system exhibits high precision in named entity extraction, well-structured report generation with consistency, and cognitive chatbot conversation, for supporting its use in actual clinical settings. The outcomes confirm the system's ability to minimize manual processing, facilitate better patient–

doctor communication, and act as a base tool for intelligent healthcare delivery.

V CONCLUSION

Thus, our research and project show how Named Entity Recognition and Large Language Models can take messy, free-text health notes and turn them into tidy, ready-to-read data. Doctors spend less time rummaging for details, and patients get clearer answers about their condition because the system extracts and summarizes information on the fly. A companion chatbot sits at the patient's elbow, sending personalized tips in plain language the moment new data lands.

Our system bridges the gap between the raw clinical data in reports and actionable knowledge, that offers valuable support in summarized report generation, query answering, and decision-making. All in all, it lays a scalable foundation for intelligent healthcare systems capable of evolving and scaling with the demands of modern medical practices.

VI FUTURE SCOPE

The proposed system has interesting opportunities for future developments in enhancements and expansion. One possible development would be the addition of clinical decision support systems that automatically retrieve patient data and cross-reference it with established practices to assist practitioners in making appropriate decisions. Also, the integration of diagnostic and imaging tests, pathology, and lab results with textual health records would improve the understanding of the patient’s condition in question. From the perspective of further strengthening care delivered proactively, it would also be possible to apply learning model predictive analytics to spot and assess health conditions and help with early interventions designed for the particular patient. With those

improvements, the chatbot could be concerned with enhanced user experience by speaking different languages which would enable it to engage with patients from heterogeneous backgrounds and increase responsiveness among the patients. The use of federated learning could enable model training from multiple institutions to be conducted in a single location without compromising patient confidentiality and data security.

The integration of IoT-enabled wearables and real-time health monitoring devices would further augment the system, enabling continuous tracking of vital health parameters and timely medical interventions. Lastly, employing advanced generative AI models to generate and validate insights in real time, with direct doctor oversight, would enhance the accuracy, reliability, and applicability of patient-centric recommendations.

These advancements in the system can evolve into a fully integrated, intelligent healthcare assistant, revolutionizing clinical workflows and improving patient care outcomes through data-driven decision-making and automation.

VII ACKNOWLEDGMENT

We sincerely wish to thank the Department of Artificial Intelligence and Data Science at AISSMS IOIT, Pune for their unwavering support and guidance during the research project “Biomedical Knowledge Extraction by Implementing Named Entity Recognition using LLM”

It is our privilege to express gratitude to the professors and mentors who, through their profound ideas, aided in completing this work. Their guidance and support tremendously helped address the key issues with the project and made a huge difference in bringing this project to fruition.

REFERENCES:

[1] https://streamlit.io Streamlit. (2023). Streamlit: The Fastest Way to Build and Share Data Apps.

[2] Aayush Kapoor and Dr Sujala D. Shetty, “Enhancing Healthcare Information Accessibility Through a Generative Medical Chatbot”, 2024.

[3] Donghao Huang, Zhenda Hu and Zhaoxia Wang, “Performance Analysis of Llama 2 Among Other LLMs”, 2024.

[4] Jinhyuk Lee, Woojin Yoon, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Sungdong Kim and Jaewoo Kang. “BioBERT: a pre-trained biomedical language representati

[5] on model for biomedical text mining”, 10 September 2019.