AI-POWERED MEDICAL CHATBOT USING GAMMA LLMV2 AND BERT

Problem Statement:- Healthcare accessibility is challenged by geographic barriers, resource limitations, and high patient loads, necessitating an AI-powered medical chatbot to provide real-time, accurate, and context-aware medical advice and support.

Authors
Don't forget the names of the research authors and co-authors. Use full names and include any titles or honorifics the authors may have, as well as the university or research institution they are representing.

Affiliations
Researches are often under or on behalf of a university, an organisation or academic/ research institutions. When available, include their logos with the names.

INTRODUCTION

This project focuses on developing an AI-powered medical chatbot designed to provide accurate, context-aware medical advice on symptoms, medications, and dietary recommendations. Utilizing advanced language models like Gamma LLMv2 and BERT, along with vector-based retrieval algorithms, the chatbot processes a vast medical knowledge base to deliver precise answers. The system integrates Pinecone for efficient semantic search, Streamlit for a user-friendly interface, and Flask for backend functionality. By offering real-time, accessible medical assistance, the chatbot aims to improve patient engagement and address challenges faced in healthcare, such as limited access to specialists and overwhelming patient loads.

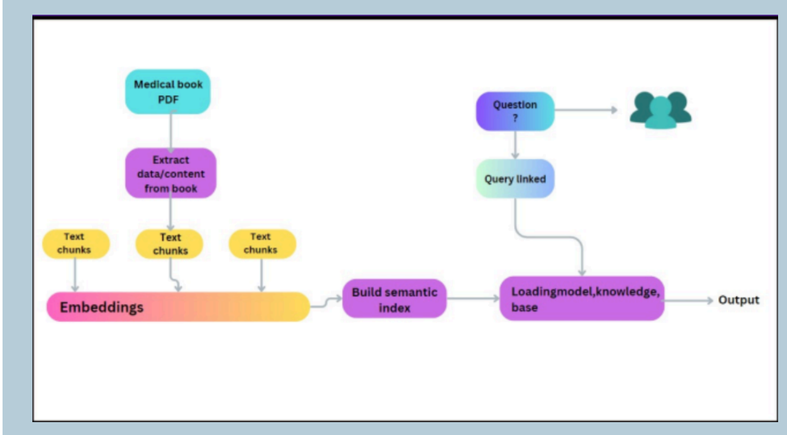
OBJECTIVE

The project aims to develop an AI-powered medical chatbot using Gamma LLMv2 and BERT models to provide accurate, real-time medical advice. It leverages Pinecone for efficient semantic search, enhancing healthcare accessibility and user engagement.

METHODOLOGY

The project utilizes Gamma LLMv2 and BERT models for generating context-aware medical advice based on a large medical dataset. Pinecone is employed for efficient semantic search by converting text data into vector embeddings. The frontend is built using Streamlit for a user-friendly interface, while Flask handles backend operations. The system processes user queries, performs vector searches, and generates accurate responses to enhance healthcare accessibility.

ARCHITECTURE DIAGRAM



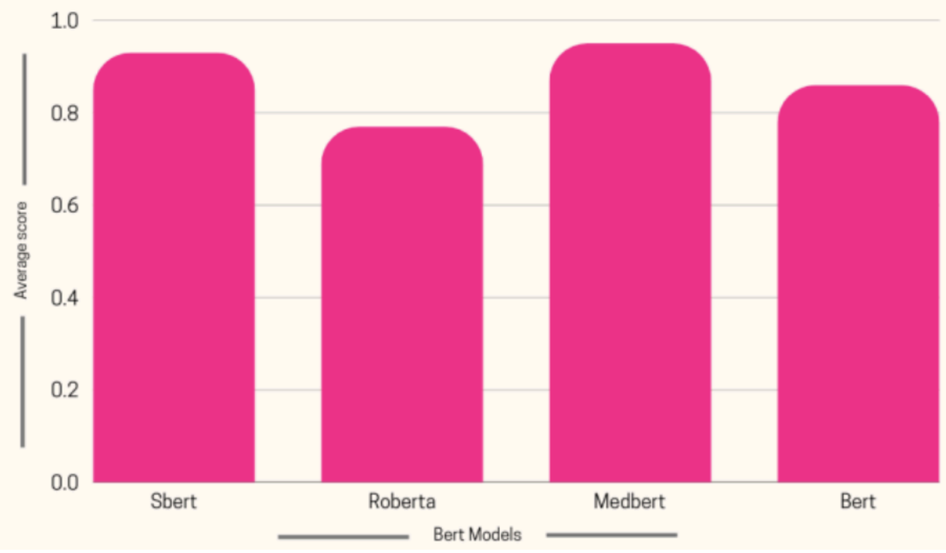
The diagram illustrates the process of extracting content from a medical PDF book, converting it into text chunks, and embedding them to build a semantic index. When a user poses a query, the system links it to the knowledge base and provides an accurate output by loading the model and relevant data.

ANALYSIS

The performance evaluation of various BERT models, including SBERT, RoBERTa, MedBERT, and BERT, was conducted to assess their effectiveness in handling medical queries within the chatbot system. From the results, SBERT emerged as the top performer, achieving the highest average score, which highlights its exceptional ability to understand and generate contextually accurate medical responses. This is particularly important for the project, as SBERT's strong semantic understanding aids in providing reliable advice on symptoms, medications, and treatment options.

On the other hand, RoBERTa, MedBERT, and BERT also performed well, though their average scores were slightly lower than SBERT. These models, while still capable of offering valuable responses, were not as finely tuned for the medical domain compared to SBERT, which is specifically optimized for such tasks. The comparison of these models emphasizes the importance of selecting the most suitable pre-trained model for a medical chatbot, as it directly impacts the chatbot's performance in terms of response accuracy and user satisfaction. This analysis contributed to refining the chatbot's performance and ensuring that it could deliver precise and useful medical advice to users.

Performance of Bert models wrt average score



RESULT

- Performance Evaluation: The chatbot demonstrated high accuracy and fast response times in providing medical advice based on user queries.
- Model Comparison: The chatbot’s performance was compared to other models like SBERT, RoBERTa, MedBERT, and BERT, with Gamma LLMv2 showing superior results in medical context understanding.
- User Satisfaction: The chatbot received positive feedback from users in terms of response relevance and clarity, contributing to better user engagement in healthcare interactions.
- Semantic Search Efficiency: Pinecone significantly improved the chatbot's ability to quickly retrieve relevant medical information through vector-based search.
- Accuracy Scores: The system achieved a high accuracy rate in diagnosing symptoms and offering treatment recommendations, enhancing its reliability as a healthcare assistant.

CONCLUSION

The project successfully developed an AI-powered medical chatbot that leverages advanced models like Gamma LLMv2 and BERT variants to deliver accurate and context-aware medical advice. Through the integration of Pinecone for efficient semantic search and a well-structured pipeline for processing medical data, the chatbot demonstrated significant potential in enhancing healthcare accessibility. The evaluation results showed that SBERT outperformed other models in understanding and responding to medical queries, ensuring reliability and precision. Combining a user-friendly frontend using Streamlit and a robust backend with Flask ensured seamless interaction for users. By addressing critical challenges such as lack of accessibility to medical professionals and the need for instant support, this project highlights the transformative role of AI in healthcare. It lays a strong foundation for future advancements, including further model fine-tuning, integration of additional datasets, and real-world deployment to serve diverse user needs effectively.

GROUP MEMBERS

202101070103 :- Girish Amrutkar

202101070094 :- Omkar Awari

PROJECT ADVISER

Prof. Deepti Ghusse
Department Of Computer Engineering