MedBot : A GenAI based Chatbot for Healthcare

Pratham Agarwal, Yash Agrawal, Shreeya Agarwal, Sangya Medhavi Shree Goyal, Merin Meleet

Department of Information Science and Engineering, RV College of Engineering, Bangalore, India

***Abstract*—Generative Artificial intelligence (GenAI) is transforming the healthcare industry by providing innovative solutions for patient care and information retrieval. MedBot is an innovative GenAI-driven chatbot designed to improve healthcare services by providing accurate and timely medical information. Utilizing advanced generative AI models, MedBot can respond to text, image, and audio queries, making it a versatile tool for diverse healthcare needs. The chatbot offers functionalities such as document summarization and insight extraction, aiding users in comprehending complex medical data. MedBot aims to enhance patient care by ensuring efficient and accessible interactions between users and healthcare information. This work signifies a substantial advancement in the integration of AI technologies within the healthcare sector, aiming to improve the overall efficiency and accessibility of medical support and information.**

Keywords—Generative Artificial Intelligence, Large Language Models, ChatBots, Healthcare

I. INTRODUCTION

The integration of Artificial Intelligence (AI) into healthcare systems has revolutionized patient care, providing innovative solutions to complex medical challenges. One of the most notable advancements in this field is the development of Generative AI (GenAI) driven chatbots, which offer personalized and efficient healthcare assistance. These tools have the potential to transform how patients access and interact with healthcare information, improving both the quality and efficiency of care. MedBot represents a significant step in this evolution, based on GenAI technologies to create a comprehensive healthcare chatbot designed to streamline healthcare interactions and make it easier for users to manage their health.

By offering features such as document summarization and insights extraction from medical reports, MedBot enhances the accessibility and comprehensibility of medical information, empowering users to make informed decisions about their health. MedBot's ability to process images and respond to audio queries, utilizing pre-trained Large Language Models (LLMs), ensures a seamless and intuitive user experience. This multi-modal functionality allows MedBot to cater to a wide range of user needs, providing personalized assistance in various contexts.

The significance of a GenAI-based chatbot like MedBot lies in its potential to bridge gaps in healthcare provision, particularly in areas where access to medical professionals may be limited. Based on advanced AI technologies, MedBot can offer timely and concise medical information, reducing the burden on healthcare systems and improving patient outcomes. Furthermore, the use of GenAI in healthcare chatbots represents a crucial advancement in the personalization of healthcare, allowing for tailored interactions that consider the specific needs and contexts of individual users. This level of personalization can lead to better patient engagement, adherence to medical advice, and overall health outcomes.

This paper explores the development of MedBot, detailing its underlying technology and design principles. It also examines the potential impact of AI-driven chatbots on healthcare provision, highlighting how MedBot can enhance accessibility, efficiency, and engagement in the healthcare domain. By demonstrating the capabilities of GenAI in healthcare, MedBot sets a new standard for patient care and information dissemination.

II. LITERATURE REVIEW

Tom B. Brown, et al [1] demonstrates that scaling up language models significantly enhances their few-shot learning capabilities, sometimes even rivaling state-of-the-art fine-tuning methods. GPT-3, a language model with 175 billion parameters—10 times larger than any previous non-sparse model—was trained and evaluated on various NLP tasks, including translation, question-answering, and close tasks, without any fine-tuning. It performed well across many datasets, though it faced challenges on some due to training on large web corpora. The findings suggest that larger models improve in-context learning, where models learn from few examples during inference rather than through extensive training.

Sarah Sandmann et al.[2] evaluates the clinical accuracy of GPT-3.5 and GPT-4 in suggesting diagnoses, examination steps, and treatments for 110 medical cases across diverse clinical disciplines. The study also assesses two configurations of Llama 2 open-source models in a sub-study and benchmarks the diagnostic task using a Google search for comparison. GPT-4 outperformed GPT-3.5 and Google in diagnosis and examination, showing better performance on frequent diseases. However, all models struggled with rare diseases. The findings indicate that commercial LLMs show potential for medical question answering but also highlight the need for robust, regulated AI models in healthcare. Open-source LLMs can be viable options for specific needs concerning data privacy and transparency.

Sydney Nguyen et al.[3] presents an innovative system that integrates computer vision and natural language processing to interpret medical images accurately. This system, Pathological-Llama, uses a generative task approach, distinguishing it from traditional classification-based VQA systems, to provide detailed, contextually relevant answers to complex medical questions. Developed and fine-tuned using the PathVQA dataset, the system emphasizes explainability through methods like Integrated Gradients and leverages GPT-4 for in-depth analysis. The system demonstrated high effectiveness, achieving a BERT score of 0.591 and an F1 score of 0.419, confirming its robust generalization capabilities. Pathological-Llama sets a benchmark for reliable, transparent AI in healthcare, aiming to enhance patient care and diagnostics by offering precise and explainable medical solutions.

Yang et al.[4] explores the evolution, potential applications, and challenges of large language models (LLMs) in healthcare. They highlight the remarkable capabilities of models like ChatGPT and categorize LLMs into biomedical and clinical domains based on their training data. The development of models such as BioBERT, SCIBERT, and ClinicalBERT, trained on specialized datasets, is emphasized for their enhanced performance in medical tasks. The applications of LLMs are extensive, including pre consultation, diagnosis, management, medical education, and medical writing. Despite their potential, significant challenges like data privacy, information credibility, data bias, and interpretability must be addressed for successful clinical implementation. The authors stress that overcoming these challenges is crucial to fully realize the transformative potential of LLMs in healthcare.
Yiqiu Shen et al. [5] discuss the dual nature of large language models (LLMs) like ChatGPT in healthcare. While LLMs can enhance medical applications such as diagnosis, treatment, and education, they also pose significant challenges, including the potential for generating inaccurate or biased information, substantial computational resource requirements, and concerns about patient data privacy. The authors emphasize the need to balance the innovative capabilities of LLMs with necessary precautions to mitigate their risks.

Radford et al. [6] present an innovative approach that diverges from recent state-of-the-art models by avoiding unsupervised pre-training or self-teaching methods. Instead, Whisper achieves strong performance through extensive supervised training on a large and diverse dataset. This approach emphasizes zero-shot transfer, enhancing the system's robustness across various speech recognition tasks. The authors highlight the benefits of focusing on weakly supervised pre-training, which has been underestimated in previous research. They suggest future improvements could include increasing training data for underrepresented languages and exploring the impact of language models on system robustness. The paper concludes that Whisper's success showcases the potential of scaling supervised training for robust speech recognition without relying on self-supervised techniques.

F.Zhang et al. [7] Discusses how Generative Adversarial Networks can be used for tasks such as medical image generation, data augmentation, anomaly detection, disease prediction, and drug discovery. The findings from studies or experiments conducted to demonstrate the effectiveness of Generative Adversarial Networks in healthcare applications supports the implementation feasibility.

P.Weber et al. [8] explores user perceptions and motivations behind engaging with chatbots and voice assistants. They conducted a study involving 104 participants to assess various factors influencing user interaction. The findings highlight that convenience, efficiency, and curiosity are primary motivators for using these technologies. Additionally, the study reveals that users have distinct preferences for chatbots and voice assistants based on the context of use, with chatbots being favored for customer service and voice assistants for personal tasks. The authors suggest that understanding these motivations can inform the design of more user-centric conversational agents, enhancing user satisfaction and engagement. The study also emphasizes the importance of addressing user concerns about privacy and security to increase trust in these technologies.

III. METHODOLOGY

A. Exploration of different LLM Models

The development of MedBot began with an extensive exploration of different LLMs to determine the most suitable options for providing robust healthcare support. Various LLMs are evaluated based on their capabilities in natural language understanding, context awareness, and response generation. The primary focus was on models that could handle complex medical terminology and provide accurate, contextually relevant answers to user queries. Among the models considered were GPT-3, GPT-4, BERT, Google Vision Pro, Large Language and Visual Assistant (Llava), Whisper model, Meta Llama 3-7b and Meta Llama 3-13b. Each model was assessed for its performance in engaging in meaningful conversations, understanding intricate medical queries, and generating precise, helpful responses. This exploration phase was crucial in identifying the strengths and limitations of each model, ultimately guiding the selection of the most effective LLM for MedBot’s needs.

B. Model Selection

MedBot, a GenAI-based chatbot for healthcare, integrates multiple advanced AI models to provide a comprehensive and user-friendly interface. For handling text queries, the Llama 13B model was selected due to its superior performance in healthcare-related responses, demonstrating an accuracy of 80.5% in specialized medical queries and an overall accuracy of 76.9% .[9]

For image analysis, the Llava multi-modal model was chosen. Trained on 158K unique language-image instructions, this model achieved an impressive accuracy of 90.92%, significantly outperforming the GPT-4 model, which has an accuracy of 82.69% .[10] This capability allows MedBot to effectively interpret and analyze medical images, providing valuable insights to users.

In handling audio queries, the medium version of a speech-to-text model was employed. This model, consisting of 769 million parameters and trained on 680,000 hours of audio data and corresponding transcripts, achieved an accuracy of 80.3% .[11] This ensures that MedBot can accurately transcribe and understand spoken language, facilitating smooth and efficient communication through voice commands.

Additionally, for text-to-audio responses, the Python gTTS (Google Text-to-Speech) library is utilized. This allows MedBot to convert text responses into natural-sounding speech, enhancing the accessibility and user experience for individuals who prefer auditory information.

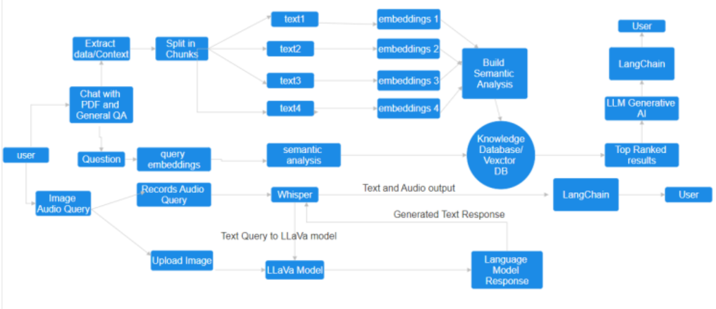


Figure 1: Embedding of Data from external files

C. Data Embedding

For relevant information retrieval MedBot, embedding relevant medical data is essential for providing accurate and comprehensive responses. The embedded data includes medical books and information from prominent medical websites. This data is first converted into PDFs and text, which are then transformed into clusters of vectors and stored in a vector database, as illustrated in Figure 1.

The process begins with extracting data from PDFs. Once the text is extracted, it is split into smaller, manageable chunks using LangChain's tokenization techniques. This segmentation ensures that each chunk is contextually coherent and suitable for embedding.

Each of these individual chunks is then converted into embeddings, which are numerical representations of the text. These embeddings capture the semantic meaning of the text and are generated using advanced language models. The embeddings are stored in a vector database, creating a semantic knowledge base.

When a user query is received, MedBot generates an embedding for the query and searches the vector database for similar embeddings. This process leverages the semantic relationships captured in the embeddings to retrieve the most relevant information. The retrieved information is then used to generate a response to the user's query, ensuring that the response is both accurate and contextually appropriate.

D. Integration of VectorDB

In ensuring efficient retrieval of accurate and pertinent data for the LLM model, the integration of a vector database is crucial. In this work, ChromaDB was seamlessly integrated to support the storage and retrieval of embeddings derived from medical text data.

ChromaDB, developed by Meta, stood out for its robust capabilities in managing vector data. Its integration into MedBot's infrastructure ensures that the generated embeddings are stored efficiently and readily accessible for query processing. This integration enables MedBot to quickly and accurately retrieve medical information, thereby facilitating timely and relevant responses to user queries.

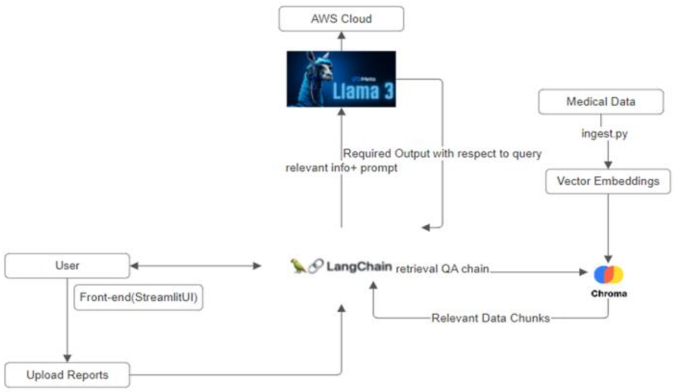


Figure 2: Architecture of Medbot

E. Semantic Search and Model Utilization

Upon uploading a PDF document, MedBot initiates the data embedding process, wherein the text content is segmented into manageable chunks. These chunks undergo embedding, facilitating efficient storage and retrieval within the vector database.

When a user query is received, MedBot employs the Llama 13B model, which uses its transformer architecture comprising encoders and decoders. The transformer encoder, a fundamental component of the Llama 13B model, plays a crucial role in understanding the semantic context of the input text. The transformer encoder operates by processing the input tokens sequentially through multiple self-attention layers. In each layer, attention mechanisms allow the model to focus on relevant parts of the input text while considering the relationships between different tokens.[12] This enables the encoder to capture intricate semantic nuances and dependencies within the text, facilitating a deeper understanding of the content.

Once the input text has been encoded by the transformer encoder, MedBot conducts a semantic search within the embedded data. Based on the semantic relationships captured in the embeddings, the Llama 13B model identifies and retrieves relevant information to generate contextually relevant responses to the user query as shown in figure 2.

Upon receiving an image-based query, MedBot employs the Llava model, a state-of-the-art neural network architecture designed specifically for multimodal tasks. The process begins with the image uploaded by the user, which is then processed by the Llava model's vision encoder. This encoder extracts visual features and representations from the image, transforming it into a high-dimensional vector embedding that captures the essential information within the image.

Once the image embedding is obtained, MedBot utilizes it to perform a semantic search within the vector database. The Llava model's semantic understanding capabilities allow it to identify and retrieve relevant information related to the user query. For audio queries, MedBot relies on the Whisper model, an advanced speech recognition system that converts spoken words into text. When a user submits an audio query, the Whisper model transcribes the audio input into text format, ensuring accurate and reliable conversion. The transcribed text is then passed through the Llava model along with any accompanying images.

The combined input of text and images undergoes semantic analysis by the Llava model, which leverages its multimodal capabilities to generate a comprehensive understanding of the user query as shown in figure 3. By integrating information from both the audio transcription and accompanying images, MedBot ensures that its responses are accurate, contextually relevant, and tailored to the user's needs.

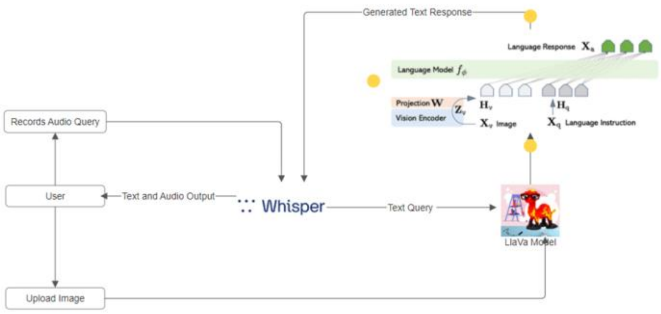
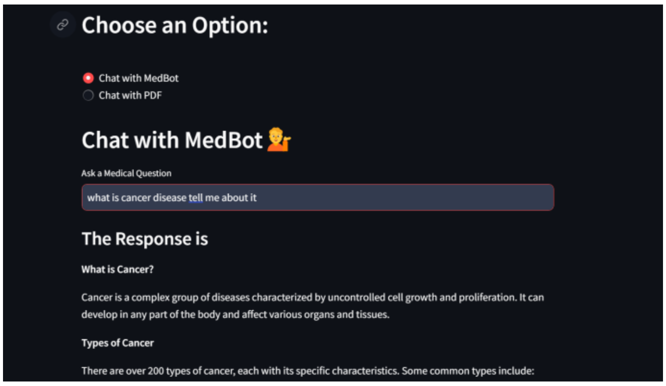


Figure 3: Architecture of Medbot with audio and image input

F. User Interface

The user interface (UI) is developed using Gradio and Streamlit, offering an intuitive platform for users to interact with the chatbot. Users can easily upload reports or PDF documents, soliciting personalized insights from the chatbot. Additionally, the UI supports image and audio inputs, processed by the Llava and Whisper models respectively. Llava generates descriptive insights for uploaded images, while Whisper transcribes audio inputs, enabling voice interactions. The interface seamlessly delivers responses in both text and audio formats, ensuring accessibility and user convenience. Through its user-friendly design and integration of advanced AI models, MedBot's UI enhances the overall user experience, facilitating effective communication and information exchange between users and the chatbot.

IV. RESULTS AND DISCUSSION



Choose an Option:

● Chat with MedBot
○ Chat with PDF

Chat with MedBot 💁

Ask a Medical Question

what is cancer disease tell me about it

The Response is

What is Cancer?

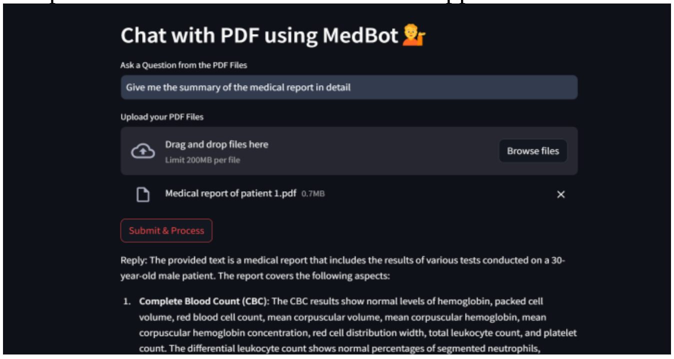
Cancer is a complex group of diseases characterized by uncontrolled cell growth and proliferation. It can develop in any part of the body and affect various organs and tissues.

Types of Cancer

There are over 200 types of cancer, each with its specific characteristics. Some common types include:

Figure 4: Interface of Text based Query Page

The user interface for MedBot, built with Streamlit, offers two options: "Chat with MedBot" and "Chat with PDF" as shown in figure 4. In "Chat with MedBot," users can ask general medical questions, which are answered using knowledge from MedBot's vector database. The "Chat with PDF" option allows users to upload documents and receive personalized insights. This dual-mode interface provides comprehensive and accessible medical support to users.



Chat with PDF using MedBot 💁

Ask a Question from the PDF Files

Give me the summary of the medical report in detail

Upload your PDF Files

Drag and drop files here
Limit 200MB per file

Browse files

Medical report of patient 1.pdf 0.7MB ✕

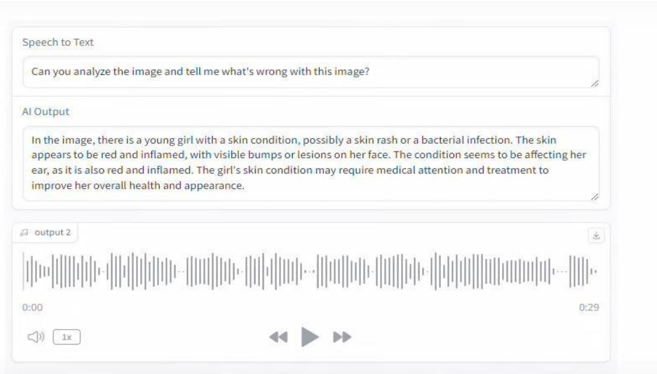
Submit & Process

Reply: The provided text is a medical report that includes the results of various tests conducted on a 30-year-old male patient. The report covers the following aspects:

1. **Complete Blood Count (CBC):** The CBC results show normal levels of hemoglobin, packed cell volume, red blood cell count, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, red cell distribution width, total leukocyte count, and platelet count. The differential leukocyte count shows normal percentages of segmented neutrophils,

Figure 5: Interface of Query using uploaded reports

In the "Chat with PDF" mode, users can upload their PDFs and receive detailed insights or answers to specific questions about the content. Users submit text queries to MedBot, which then provides accurate and relevant responses based on the document's content as shown in figure 5. This functionality allows for an interactive exploration of uploaded documents, enhancing user understanding and engagement.


Speech to Text

Can you analyze the image and tell me what's wrong with this image?

AI Output

In the image, there is a young girl with a skin condition, possibly a skin rash or a bacterial infection. The skin appears to be red and inflamed, with visible bumps or lesions on her face. The condition seems to be affecting her ear, as it is also red and inflamed. The girl's skin condition may require medical attention and treatment to improve her overall health and appearance.

Figure 6: Response to audio query

In the audio input feature, users can submit their queries through voice. The system then provides a text answer to the query, which can also be converted to audio, allowing users to receive responses in both text and spoken formats as shown in figure 6.

V. CONCLUSION

Future developments for MedBot could include advanced symptom analysis through integration with medical knowledge graphs, personalized interactions based on user data, and multilingual support. Additionally, secure connections with Electronic Health Records (EHRs) and integration with wearable health monitors could enable real-time health monitoring and early detection of potential issues, further enhancing MedBot's capabilities and impact on healthcare provision.

MedBot, with its foundation in Llama3, Llava, and Whisper, presents a future where AI chatbots play a transformative role in healthcare. By empowering patients, streamlining processes, and fostering a more inclusive environment, MedBot paves the way for a future where patients are active participants in their well-being, supported by a next-generation healthcare companion that is based on the power of GenAI to revolutionize patient engagement.

VI. REFERENCES

[1] T. B. Brown et al., “Language Models are Few-Shot Learners,” Neural Information Processing Systems, vol. 33, pp. 1877–1901, May 2020, [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[2] S. Sandmann, S. Riepenhausen, L. Plagwitz, and J. Varghese, “Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks,” Nature Communications, vol. 15, no. 1, p. 2050, Mar. 2024, doi: https://doi.org/10.1038/s41467-024-46411-8.

[3] S. Nguyen, “Pathological-Llama: an Explainable Medical Visual Question Answering System,” Available:https://www.zhaw.ch/storage/engineering/institutezentren/cai/studentische_arbeiten/Herbst_2023/MSE__Master_Thesis_23_bogo_PathologicalLlama.pdf

[4] R. Yang, T. F. Tan, W. Lu, A. J. Thirunavukarasu, D. S. W. Ting, and N. Liu, “Large language models in health care: Development, applications, and challenges,” HealthCare Science, vol. 2, no. 4, pp. 255–263, Jul. 2023, doi: 10.1002/hcs2.61.

[5] Y. Shen et al., “ChatGPT and other large language models are double-edged swords,” Radiology, vol. 307, no. 2, Apr. 2023, doi: 10.1148/radiol.230163.

[6] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust speech recognition via Large-Scale Weak Supervision,” arXiv.org, Dec. 06, 2022. https://arxiv.org/abs/2212.04356

[7] F. Zhang, L. Wang, J. Zhao, and X. Zhang, “Medical applications of generative adversarial network: a visualization analysis,” Acta Radiologica, vol. 64, no. 10, pp. 2757–2767, Aug. 2023, doi: 10.1177/02841851231189035.

[8] P. Weber and T. Ludwig, ‘‘(Non-) interacting with conversational agents: Perceptions and motivations of using chatbots and voice assistants,’’ in Proc. Conf. Mensch Comput., vol. 1, Sep. 2020, pp. 321–331

[9] “What LLM is The Most Accurate? – Originality.AI.” https://originality.ai/blog/what-llm-is-the-most-accurate

[10] H. Liu, C. Li, Q. Wu, and Y. J. Lee, “Visual instruction tuning,” arXiv.org, Apr. 17, 2023. https://arxiv.org/abs/2304.08485

[11] X. Yi, E. Walia, and P. S. Babyn, “Generative adversarial network in medical imaging: A review,” Medical Image Analysis, vol. 58, p. 101552, Dec. 2019, doi: 10.1016/j.media.2019.101552.

[12] A. Vaswani et al., “Attention is All you Need,” arXiv (Cornell University), vol. 30, pp. 5998–6008, Jun. 2017, [Online]. Available: https://arxiv.org/pdf/1706.03762v5