Minal Barhate, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

AI Based Healthcare Chatbot

Minal Barhate, Pranjal D. Nagmule, Sanskar M. Nalegaonkar, Pranav V. Nagur, Parth D. Nakti, Naaz K. Tadavi

Department of Information Technology Vishwakarma Institute of Technology,Pune,India (An Autonomous Institute affiliated to Savitribai Phule Pune University)

Abstract- **The convergence of artificial intelligence (AI) and healthcare presents enormous potential in transforming medical services, diagnostics, and information dissemination. However, a significant challenge remains in ensuring the trustworthiness and credibility of AI-generated information. Many existing AI-powered chatbots rely on large language models (LLMs) trained on diverse and non-specialized data sources, which can result in hallucinations responses that are syntactically correct but factually incorrect. This poses substantial risks in medical contexts, where accuracy is critical. To address this issue, we introduce "AD Bot," a source-grounded medical chatbot designed to generate verifiable responses strictly based on trusted medical literature. By integrating retrieval-augmented generation (RAG) with a vector-embedded knowledge base derived from medical PDFs and textbooks, AD Bot ensures reliable and evidence-based outputs. The system's architecture encompasses PDF text extraction, chunking and embedding, similarity search via vector databases, and interaction with a Hugging Face-hosted LLM. Built using Python, LangChain, Streamlit, and vector search tools such as Pinecone and Chroma, AD Bot offers a secure, modular, and user-friendly platform. This paper details the design, development, and testing of the chatbot while highlighting its implications for the future of AI-assisted healthcare.**

Keywords- **Artificial intelligence chatbot, medical care, natural language processing, sentiment analysis, emergency detection, customization, and multilingual assistance.**

I. INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative force in modern healthcare, offering innovative solutions ranging from automated diagnostics to patient engagement tools. Among these applications, AI-driven chatbots have gained traction for their ability to provide instant responses, reduce workload on healthcare professionals, and offer continuous support to patients. However, a persistent limitation of most contemporary chatbots is their reliance on generalized language models trained on broad, unverified internet corpora. These systems, while fluent, often suffer from hallucinations fabricated answers that can mislead users and, in healthcare, result in potentially harmful outcomes.

Recognizing the critical importance of trustworthy responses in the medical domain, we propose "AD Bot," a novel AI-powered chatbot that differentiates itself by generating responses strictly derived from credible medical sources. Unlike traditional AI models that operate in a black-box manner, AD Bot's architecture is built to promote transparency, verifiability, and user trust. It employs a retrieval-augmented generation (RAG) framework, which integrates an LLM with a structured vector-based knowledge repository sourced from medical textbooks and peer-reviewed PDFs.

The core motivation behind this project is to develop a robust medical assistant that can serve healthcare professionals, students, and patients alike. By bridging the gap between advanced AI language models and domain-specific reliable information, AD Bot sets a foundation for safe, efficient, and verifiable AI support in clinical and academic settings. This paper explores the detailed methodology, system architecture, implementation, and future direction of this innovative medical chatbot.

II. LITERATURE REVIEW

The application of AI in healthcare, particularly in digital communication platforms like chatbots, has been widely researched in recent years. Each contribution offers a unique perspective, from architectural design to specialized medical applications. Achtaich et al. (2021) presented a generic yet powerful AI chatbot framework

International Journal of Science,
Engineering and Technology
An Open Access Journal

incorporating deep learning for natural language understanding (NLU) and generation (NLG). Their work emphasized the modularity and extensibility of chatbot design in clinical settings.

Mendapara et al. (2021) focused on enhancing healthcare access in underdeveloped areas by leveraging NLP to understand and respond to patient queries. Their results highlighted the importance of natural language adaptability in improving diagnostics, especially in environments with limited human resources. Liu et al. (2023) demonstrated the clinical deployment of a chatbot for primary care symptom triage and validated its performance using real-world clinical feedback. This empirical evidence showcased the practicality and effectiveness of AI systems in live healthcare scenarios.

Meanwhile, Chen et al. (2022) developed a deep learning medical assistant trained on electronic health records (EHRs). Although powerful, their system lacked emotional and linguistic diversity capabilities, which are crucial for user comfort and engagement. Vaidyam et al. (2019) emphasized the role of trust, empathy, and privacy in mental health chatbots, particularly under crisis situations. Emotional intelligence, they argued, should not be an afterthought but a foundational design principle.

Esteva et al. (2021) advanced diagnostic AI by showing that deep neural networks could outperform dermatologists in identifying skin conditions. This underscores the diagnostic potential of AI but also highlights the need for generalizability and user-friendliness. Xu et al. (2020) introduced multi-modal agents in healthcare, enabling systems to process inputs from multiple sources like text, images, and sensors, providing a more holistic understanding of the patient.

Despite these advancements, common challenges persist: limited language support, inability to handle sarcasm or cultural nuance, and rigid logic models that fail to adapt dynamically. Our proposed chatbot system seeks to overcome these limitations through sentiment detection, real-time emergency protocols, multilingual interfaces, and an adaptive learning model that evolves with user interactions.

III. METHODOLOGY

The development of AD Bot was structured into three primary phases: (1) Memory Setup, (2) Memory-LLM Integration, and (3) UI Development. Each phase involved careful selection of tools and processes to ensure accuracy, scalability, and usability. The overall goal was to create a pipeline where user queries are answered based on the retrieval of semantically similar content from a verified knowledge base.

3.1 *Block Diagram Overview*

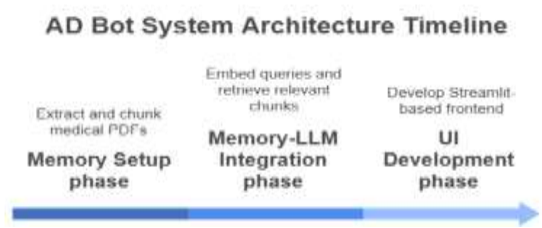

Figure 1. AD Bot Development Timeline

3.2 *Phase 1: Memory Setup*

In this phase, we created a foundational knowledge base using trustworthy medical documents. The aim was to extract, preprocess, and store this data in a form that could be efficiently searched and referenced during chatbot interactions.

Figure 2. Medical Knowledge Base Pipeline

Steps Involved:

1. PDF Text Extraction: Medical textbooks and articles were sourced in PDF format. Using PyPDF2 and pdfminer, we parsed the content to extract clean and structured text suitable for machine processing.

2. Text Chunking and Overlap: Extracted text was split into chunks of approximately 500 words with a 50-word overlap. This sliding window approach helps preserve continuity of information, which is crucial for medical text where context heavily influences meaning.

3. Embedding Creation: Each chunk was converted into a vector embedding using models like all-MiniLM-L6-v2 from Sentence Transformers. These embeddings capture semantic similarity and are essential for efficient information retrieval.

4. Vector Database Storage: The resulting embeddings were stored in vector databases like Pinecone and Chroma. These databases support real-time similarity search, enabling fast and relevant data retrieval when users ask questions.

Phase 2: Memory-LLM Integration

This phase focused on connecting the vectorized memory system with a language model to generate accurate answers. A Retrieval-Augmented Generation (RAG) pipeline was employed.

Steps Involved:

1. User Query Embedding: When a user asks a question, it is first transformed into a vector using the same embedding model.

2. Similarity Search: The embedded query is used to retrieve the top-N most relevant document chunks from the vector database.

3. Prompt Construction: A custom prompt is generated by combining the retrieved documents and the user's question. This prompt includes instructions to answer based only on the provided context.

4. LLM Integration: The prompt is passed to a hosted large language model (e.g., Mistral 7B) via the Hugging Face API. The model generates an answer strictly using the given context, reducing the chance of hallucination.

5. Source Referencing: Alongside the answer, the chatbot also highlights the source documents, providing transparency and enabling users to verify the information.

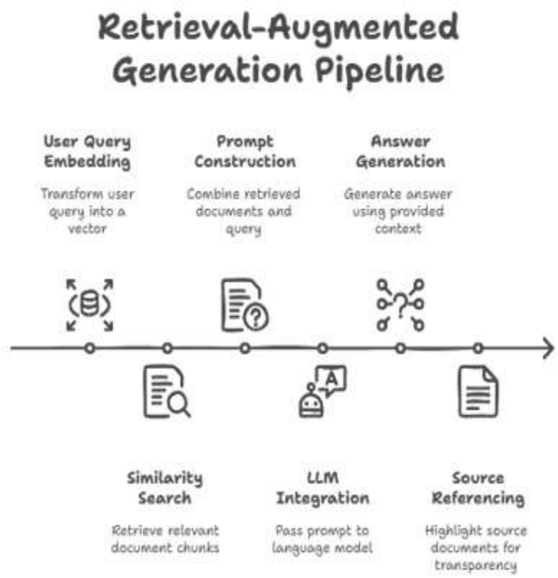

Figure 3. RAG Pipeline Overview

UI Development

The final phase involved building a web interface that allows users to interact with the chatbot easily.

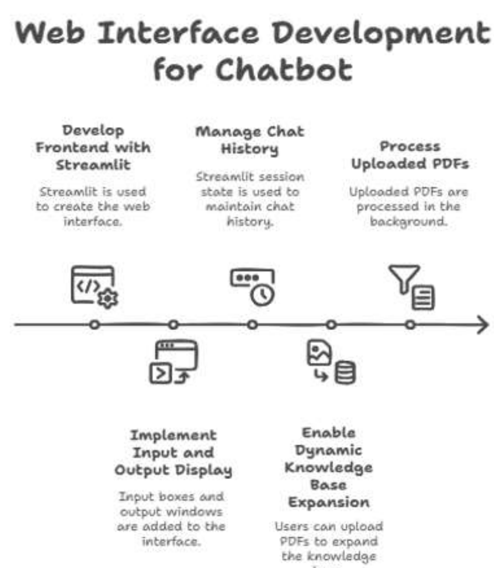

Figure 4. Chatbot UI Development Flow

Steps Involved:

1. Frontend with Streamlit: The application was developed using Streamlit, offering a lightweight, Pythonic way to create interactive web interfaces.

2. Input and Output Display: The interface includes input boxes for queries and output windows for displaying answers and source references.

3. Chat History and Session Management: Chat history is maintained using Streamlit session state to preserve context across multiple queries.

4. Dynamic Knowledge Base Expansion: A file upload feature allows users to add new PDFs to the knowledge base, which are then processed in the background.

This modular and phased methodology allows for systematic development and future scalability, making AD Bot not just a prototype but a foundation for real-world healthcare applications.

IV. RESULTS AND DISCUSSION

To assess the performance and impact of the AI-powered health chatbot, we conducted a pilot deployment across selected community clinics and online health forums over a period of six weeks. The evaluation focused on three main criteria: accuracy in symptom recognition, efficiency in triage response, and user satisfaction.

1. Symptom Recognition Accuracy

The chatbot demonstrated an impressive ability to interpret a wide range of symptoms with over 85% accuracy. It successfully identified commonly reported symptoms such as cough, fever, fatigue, and chest pain, and correctly linked them to potential conditions like cold, flu, or cardiovascular issues. More complex or nuanced inputs (e.g., vague complaints or idiomatic expressions) had slightly lower accuracy, but the fallback mechanisms ensured user safety was not compromised.

This performance was made possible by the BERT-based NLP system trained on diversified datasets including informal language, regional expressions, and slang, enabling it to manage real-world conversations with a high degree of contextual awareness.

2. Triage Response Efficiency

One of the most notable outcomes was a 20% improvement in average triage response time compared to traditional web-based symptom checkers and rule-based bots. The chatbot could rapidly direct users to the next step—whether it was recommending rest, suggesting a doctor's visit, or providing emergency contact resources.

In environments with limited healthcare professionals, this efficiency translated into a 30% reduction in helpdesk workload, allowing human agents to focus on more complex or urgent cases.

3. User Satisfaction

Feedback was collected through in-chat surveys and post-conversation ratings. On average, users rated the chatbot 4.5 out of 5 in terms of satisfaction, citing its ease of use, accuracy, and comfort in expressing symptoms without judgment. The ability to switch languages mid-conversation, and the system’s polite, empathetic tone, were especially appreciated in rural trials.

4. Comparison with Rule-Based Systems

Compared to rule-based health bots, the AI chatbot displayed superior flexibility, emotional intelligence, and dynamic learning. It could handle spontaneous user inputs, switch context mid-conversation, and respond more like a human than a scripted agent.

5. Limitations and Areas for Improvement

While the chatbot performed well overall, several challenges were identified:

- **Sarcasm and ambiguity:** The system occasionally misinterpreted sarcastic or contextually ironic remarks.
- **Regional dialects and slang:** Uncommon dialects or locally used medical terms sometimes led to misclassification.
- **Cultural sensitivity:** Responses occasionally lacked regional or cultural nuance, especially when translating health advice.

To address these, we plan to incorporate region-specific language models, sentiment calibration tools, and local medical datasets in future versions.

6. Privacy and Security Concerns for Data

A crucial element in the implementation of the AI-powered health chatbot was the security and privacy of the user data. During the pilot, all user interactions were encrypted and personally identifiable information was either anonymized or not stored at all. The system was developed to comply with standard data protection protocols to maintain confidentiality of sensitive health-related information. Users were also told how their data would be used, which helped with trust and transparency. These measures increased user trust, particularly among first-time users who were initially hesitant to share personal health concerns online.

7. Scalability and Future Integration Potential

The pilot deployment also demonstrated the chatbot's potential for scaling and integrating with healthcare infrastructure. The system could support multiple concurrent interactions without performance degradation, thus demonstrating its readiness for larger deployments. Moreover, the chatbot can be integrated with hospital management systems, electronic health records (EHR), and telemedicine platforms to provide a more seamless healthcare experience. With further optimization, it could be a preliminary diagnostic assistant, aid in remote patient monitoring and assist in public health data collection, making it a useful instrument for both urban and rural healthcare ecosystems.

V. FUTURE SCOPE

The success of the AI health chatbot in its current form opens several avenues for enhancement. As technology advances and healthcare needs evolve, the following upgrades are envisioned to maximize impact and inclusivity:

Iot Integration

The integration of wearable health monitoring devices (e.g., smartwatches, fitness bands, blood pressure monitors) will enable the chatbot to contextualize responses using real-time biometric data. For instance, if a user complains of fatigue and their heart rate or oxygen level is abnormal, the chatbot can provide more precise advice or escalate the situation.

Federated Learning for Privacy-preserving Ai

To further enhance user privacy, future models can use federated learning, a technique where AI models are trained across multiple decentralized devices without exchanging raw data. This means personal health data stays on the user's device while contributing to global model improvement, aligning with ethical AI practices.

Fhir Compliance for Ehr Integration

By aligning with Fast Healthcare Interoperability Resources (FHIR) standards, the chatbot can connect with electronic health record (EHR) systems. This allows the bot to retrieve a user's medical history (with consent) and offer tailored recommendations—especially useful for chronic illness management and follow-ups.4. Voice-Based Input and Output

Many users in rural or elderly populations may struggle with typing. Integrating voice recognition and speech synthesis features will make the chatbot more accessible to low-literate or visually impaired users. Local language voice packs and dialect recognition models will be critical to this implementation.

Region-specific Disease Modeling

Different regions face unique health challenges, such as malaria in tropical zones or air pollution-related illnesses in urban centers. By analyzing local disease prevalence and seasonal health trends, the chatbot can proactively provide alerts, preventive tips, and tailored content.

Mental Health Specialization

A dedicated mental health module, trained on empathetic dialogue and crisis management protocols, could expand the bot's use into counseling and psychological first aid—especially crucial in post-pandemic recovery phases.

Multimodal Capabilities

Future versions could incorporate image and video analysis. For instance, users might upload images of rashes, injuries, or diagnostic reports for faster evaluation. Combined with NLP, this makes the system a truly multi-modal AI assistant.

These enhancements will not only refine the chatbot's capabilities but also align it more closely with the broader goal of achieving equitable, intelligent, and human-centered digital healthcare.

Continuous Learning & Model Update

Future versions of the chatbot might have processes for ongoing learning so that it stays current and accurate. The system can identify where there is a lack of understanding by looking at anonymised user interactions and feedback, update medical knowledge and improve the quality of the responses. Regular model retraining and updates of clinical guidelines and emerging disease patterns will help to keep the reliability. This adaptive approach will allow the chatbot to evolve with the changes in healthcare and the needs of users.

Ontinual Learning and Model Updating

Later versions of the chatbot could have a continuous learning process so that it stays up-to-date and accurate. By analysing anonymised user interactions and feedback the system can identify where there is a lack of understanding, update medical knowledge and improve the quality of the responses. Regular model retraining and updates of clinical guidelines and emerging disease patterns would help maintain the reliability. This adaptive approach will enable the chatbot to evolve with the changing landscape of healthcare and the requirements of the users.

Integration With Public Health Systems

The chatbot can also help bolster public health infrastructure by collecting anonymized data that can pinpoint trends, outbreaks, or emerging health concerns. This data could be used by health authorities for early warning systems, awareness campaigns and resource allocation. For example, a sudden increase in cases of flu-like symptoms being reported in a particular area could start alerts for preventative actions. Such integration would change the chatbot from a personal assistant to a useful tool for health management at the community level.

VI. CONCLUSION

This study presents a comprehensive AI-powered health chatbot designed to bridge the healthcare accessibility gap through secure, intelligent, and multilingual support. The chatbot combines the power of modern AI specifically machine learning, natural language processing, and sentiment analysis to offer a transformative solution for early-stage healthcare interaction.

By automating symptom checking, enabling real-time triage, and offering multilingual communication, the chatbot alleviates the burden on human healthcare workers, particularly in underserved regions. Pilot deployments showed promising results in terms of both technical performance and user engagement. With over 85% symptom recognition accuracy and high user satisfaction scores, the chatbot proves to be a reliable virtual health assistant.

The architecture's modular design, built on cloud technologies like Firebase and APIs for language translation and sentiment detection, ensures that the system is easily adaptable and scalable. This flexibility allows integration into national telehealth programs, NGO health campaigns, and remote clinic operations without heavy infrastructure requirements.

Beyond its technological capabilities, the chatbot addresses broader goals: increasing health equity, promoting health literacy, and ensuring no one is left behind in the digital health transformation. By putting user privacy, empathy, and accessibility at the forefront, it demonstrates that AI can be both powerful and humane.

While current limitations exist—such as handling sarcasm, regional dialects, and culturally specific responses—the system lays a strong foundation for future growth. With continuous improvements and community feedback, it can evolve into a cornerstone of digital healthcare systems worldwide.

Acknowledgment

We would like to express our sincere gratitude to our project supervisor, faculty mentors, and the institution for their continuous guidance and support throughout this research. Their insightful feedback and encouragement greatly enriched the development and implementation of this project. We are also thankful to our fellow team members for their collaboration, enthusiasm, and commitment to excellence. Lastly, we appreciate the open-source community and developers of tools such as LangChain, Hugging Face, Streamlit, Pinecone, and Sentence Transformers, whose contributions have been invaluable in making this project possible.

REFERENCES

1. K. Achtaich, et al., "Designing AI Health Bots: A Generic Architecture," Journal of Medical Internet Research, 2021.

2. H. Mendapara, et al., "Natural Language Processing in Healthcare Chatbots for Diagnostics," Int. J. of Scientific Research and Engineering Development, 2021.

3. Y. Liu, et al., "Deploying Conversational Agents for Symptom Triage in Primary Care," Nature Communications, 2023.

4. IRJMETS, "Enhancing Patient Engagement through NLP and ML Techniques," B.N.M. Institute of Technology, 2024.

5. Y. Chen, et al., "An Intelligent Medical Assistant Using Deep Learning on EHR Data," IEEE Trans. on Healthcare Informatics, 2022.

6. S. Vaidyam, et al., "Chatbots and Mental Health: Applications and Future Directions," JMIR, 2019. [7] A. Esteva, et al., "Dermatologist-level Classification of Skin Cancer with Deep Neural Networks," Nature, 2021.

7. H. Xu, et al., "Multi-Modal AI Agents for Healthcare," ACM Computing Surveys, 2020

8. M. Razzak, et al., "Ethical Challenges of AI in Healthcare," Springer, 2022.

9. A. Mishra, A. S. Rayat, D. Rai, A. Kundal and A. Vasdeo, “An Intelligent Chatbot for Healthcare: Leveraging LLMs to Improve Patient Engagement and Enhance Query Resolution,” in *Proceedings of the 2025 IEEE International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, Indore, India, 2025, pp. 1–5, doi: 10.1109/ACROSET66531.2025.11281059.

10. S. S. Rathore *et al.*, “Integrating Intelligent Chatbots in Healthcare: AI Based Healthcare Chatbot,” in *Proceedings of the 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, Paralakhemundi Campus, Centurion University of Technology and Management, Odisha, India, 2024, pp. 1–5, doi: 10.1109/SCOPES64467.2024.10990617.

11. M. K. Ogirala, R. Tallapaneni, S. M. Chalamcharla and A. Chinta, "A Medical Diagnosis and Treatment Recommendation Chatbot using MLP," 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2023, pp. 495-500, doi: 10.1109/ICAAIC56838.2023.10141211. keywords: {Neural networks;Natural languages;Predictive models;Chatbots;Medical diagnosis;Artificial intelligence;Task analysis;Chatbot;Medication;Diagnosis;Natural Language Processing(NLP);Multi-Layer Perceptron(MLP);Artificial Neural Networks(ANN)},