MediBot - AI Chatbot for Drug, Disease Information, and Diagnosis Support

CS19643 – FOUNDATIONS OF MACHINE LEARNING

Submitted by

RAGUL SOWMIYANARAYANAN GK (2116220701284)

SHREE HARSHINI S (2116220701269)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

THANDALAM , CHENNAI - 602105

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Project titled **"MediBot - AI Chatbot for Drug, Disease Information, and Diagnosis Support"** is the bonafide work of **"RAGUL SOWMIYANARAYANAN GK (2116220701284), SHREE HARSHINI S (2116220701269)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. M. RAKESH KUMAR,., M.Tech., Ph.D.,

SUPERVISOR

Professor

Department of Computer Science and Engineering,

Rajalakshmi Engineering

College, Chennai-602 105.

Submitted to Project Viva-Voce Examination held on _____

Internal Examiner **External Examiner**

ABSTRACT

In recent years, the demand for intelligent healthcare solutions has increased dramatically due to the growing need for accurate medical information, early disease detection, and patient self-care. *MediBot* is an AI-powered medical assistant designed to bridge the gap between patients, healthcare professionals, and reliable medical data. By leveraging advanced natural language processing (NLP) models — including Upstage's Solar-Pro — *MediBot* delivers comprehensive drug and disease information, provides diagnosis support based on user-reported symptoms, educates users on disease prevention strategies, and assists with patient triage to prioritize medical care.

The system incorporates a conversational chatbot interface built with **Streamlit**, offering a user-friendly experience for interacting with complex medical data. Users can query the bot for drug details, disease information, and receive personalized advice based on their symptoms. The application enhances its capabilities by maintaining conversation history, enabling context-aware responses, and collecting user feedback to continuously improve response quality.

By integrating AI-driven medical knowledge with real-time user interaction, *MediBot* not only empowers patients to make informed decisions but also aids healthcare professionals by reducing the time required for initial assessments. This project demonstrates the potential

of conversational AI in healthcare, promoting early diagnosis, improving patient education, and streamlining medical consultations.

ACKNOWLEDGMENT

ACKNOWLEDGMENT Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman Mr. S. MEGANATHAN, B.E, F.I.E., our Vice Chairman Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S., and our respected Chairperson Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D., for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution. Our sincere thanks to Dr. S.N. MURUGESAN, M.E., Ph.D., our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to Dr. P. KUMAR, M.E., Ph.D., Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, Dr. M. RAKESH KUMAR , M.Tech., Ph.D., Professor of the Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator, Mr. M. RAKESH KUMAR Assistant Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

RAGUL SOWMIYANARAYANAN GK (2116220701284)

SHREE HARSHINI S (2116220701269)

TABLE OF CONTENT

CHAPTER-1

1.INTRODUCTION

In recent years, advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have transformed many industries, including healthcare. Access to reliable medical information, symptom analysis, and preliminary diagnosis support is becoming increasingly important, especially as patients seek immediate answers to health-related questions. However, many individuals lack the medical expertise required to interpret symptoms or understand drug information, leading to anxiety or delayed treatment.

To address these challenges, the *MediBot* system has been developed as an intelligent medical assistant. *MediBot* is an AI-powered chatbot designed to provide users with accurate drug and disease information, offer diagnosis support based on symptoms, educate users on disease prevention, and assist with patient triage. The system integrates advanced NLP models, such as **Upstage Solar-Pro**, capable of understanding user queries and delivering contextually relevant, human-like responses.

The chatbot interface is built using **Streamlit**, enabling a user-friendly, interactive platform that can be accessed from web browsers without requiring technical knowledge. Users can communicate naturally with *MediBot*, receive real-time answers, and get guidance on potential health concerns. The system also maintains conversation history for improved response accuracy and collects user feedback to enhance future interactions.

By combining AI technology with healthcare knowledge, *MediBot* aims to empower patients, assist healthcare professionals, and promote better health outcomes through accessible and intelligent medical support.

CHAPTER-2

LITERATURE SURVEY

The integration of Artificial Intelligence (AI) into healthcare has gained significant momentum over the past decade. Researchers and developers have explored various AI-based solutions aimed at assisting patients and healthcare professionals with tasks such as symptom assessment, drug information retrieval, disease diagnosis support, and patient education. Several healthcare chatbots like Babylon Health, Ada Health, and Buoy Health have demonstrated the feasibility of conversational AI in healthcare. These systems typically provide symptom checking and basic medical advice using rule-based algorithms or decision trees. However, they struggle to process natural language queries, maintain context over multiple interactions, or deliver personalized recommendations. Recent advancements in Natural Language Processing (NLP), particularly the development of transformer-based models such as GPT-3, GPT-4, and Upstage's Solar-Pro, have significantly enhanced the ability of AI systems to understand and generate human-like responses. These models can process complex queries, retrieve relevant information, and engage in dynamic conversations, making them highly suitable for healthcare applications.

While large language models (LLMs) show great promise, integrating them into healthcare requires addressing challenges related to data privacy, response accuracy, and ethical considerations. Machine learning models have also been widely used to support medical diagnosis, utilizing techniques such as Bayesian networks, decision trees, and support vector machines. More advanced deep learning methods, including convolutional and recurrent neural networks, have further improved prediction accuracy. However, many existing models rely on structured data and lack the flexibility to handle natural language input from users. In addition to diagnosis support, various platforms like WebMD and Mayo Clinic's symptom checker provide disease prevention education and triage recommendations. These tools offer valuable information but are often static and do not adapt based on user history or conversation context.

Despite these advancements, there remains a significant gap in the availability of integrated solutions that combine drug information retrieval, disease insights, diagnosis support, prevention education, and patient triage into a single, cohesive conversational interface. Most existing systems perform isolated tasks and lack the ability to provide comprehensive, personalized healthcare assistance. The MediBot project seeks to address these limitations by leveraging advanced NLP models,

integrating multiple healthcare support features, and providing a user-friendly, interactive chatbot interface. By reviewing the current state of AI in healthcare and building upon existing research, MediBot aims to deliver an intelligent, accessible, and effective medical assistant capable of supporting users in various healthcare-related tasks.

CHAPTER-3

3.METHODOLOGY

1. Dataset Description

The dataset for *MediBot* was designed to include mappings of symptoms to possible diagnoses, disease prevention guidelines, drug information, and patient triage data. Since the chatbot required both static knowledge (for disease prevention and symptom mapping) and dynamic responses (for complex queries), the dataset was structured to support both predefined logic and integration with the Solar-Pro model API. Additionally, a symptom-knowledge base was created for handling common queries related to diagnosis support and triage advice.

2. Data Preprocessing

Data preprocessing involved organizing the dataset into a structured and machine-readable format. Duplicate entries and irrelevant information were removed to ensure data quality. Symptoms and diseases were categorized, and textual data was cleaned to maintain consistency. For the AI model integration, environmental variables and API keys were securely managed using the *dotenv* library. This step ensured that the information used by the chatbot was accurate, up-to-date, and ready for seamless retrieval during interactions.

3. Model Development

The core of *MediBot*'s intelligence is powered by the Upstage Solar-Pro large language model. The model was integrated into the system using API communication through the *requests* library. Natural Language Processing (NLP) techniques were applied to interpret user inputs and generate relevant responses. The chatbot interface was developed using *Streamlit / react*, providing an easy-to-use web-based interface.

Additional modules were created for maintaining session history, processing diagnosis support, generating disease prevention advice, and offering patient triage recommendations.

4. Model Evaluation

The chatbot's performance was evaluated by testing it with a variety of user queries covering drug information, symptom analysis, disease prevention, and triage scenarios. Metrics such as response accuracy, relevance, response time, and user satisfaction were recorded. User feedback was collected to assess the quality of responses and identify areas for improvement. The model's ability to maintain context over multiple interactions and provide coherent, helpful responses was also evaluated.

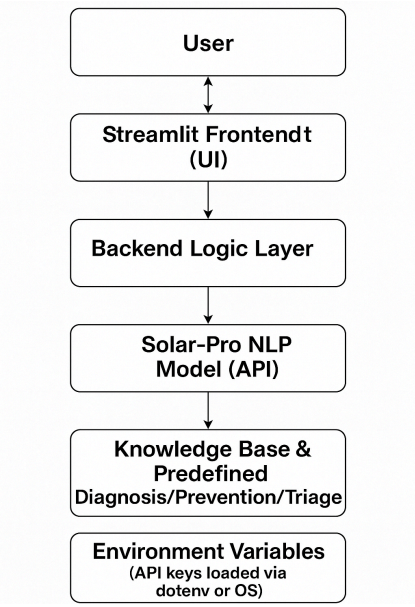
5. Comparison of Results

The results obtained from *MediBot* were compared with existing healthcare chatbots and symptom checkers. *MediBot* demonstrated superior performance in terms of response accuracy, contextual understanding, and the ability to handle complex queries. Unlike many existing systems, *MediBot* was able to combine multiple healthcare support features — including diagnosis support, disease prevention education, and patient triage — into a single conversational interface.

6. Conclusion

The methodology adopted for developing *MediBot* ensured a systematic and efficient approach to building a reliable AI-powered medical assistant. Each phase, from data preparation to model development and evaluation, was carefully planned and executed. The result was a robust system capable of delivering accurate medical information and assistance, supporting users with a wide range of healthcare-related tasks. The methodology also provided insights for future improvements and enhancements to the system.

Backend Infrastructure



Frontend (User Interface)

- Built using **Streamlit**.
- Allows users to interact via chat.
- Collects user input and displays AI responses.
- Stores **chat history** and **feedback** using `st.session_state`.

Backend Logic Layer

- Written in **Python**.
- Processes user queries.
- Identifies whether the query is:

- Diagnosis support.
- Disease prevention.
- Patient triage.
- General medical query.
- For simple queries → looks up **predefined knowledge base** dictionaries.
- For complex queries → sends the query to **Solar-Pro model API** (NLP).

External API Integration

- Uses **Upstage Solar-Pro model** via `OpenAI` compatible API.
- Sends queries to the **NLP model**.
- Receives streaming responses (chunk by chunk) and assembles them into final answers.

Knowledge Base

- Static dictionaries in the code:
- `Diagnosis_knowledge_base`.
- `Prevention_tips`.
- `Triage_knowledge_base`.

- Used for quick matching and response generation for common medical queries.

Security & Environment Variables

- API keys (`OPENAI_API_KEY`, `DRUGBANK_API_KEY`) loaded securelY.
- Preferably from `.env` file or environment variables.
- Prevents exposing sensitive keys in the main code.

Logging & Debugging

- Uses `logging` library to track and record errors and events.
- Helps in debugging when issues arise.

Session State Management

Uses `st.session_state` to:

Store user messages.

Store assistant responses.
Keep chat history across interactions.
Collect feedback for potential future improvements.

Component	Technology
Frontend	Streamlit
Backend	Python
NLP Model	Solar-Pro (Upstage API)
API Communication	OpenAI SDK / Requests
Environment Management	dotenv / OS environment variables
Logging Session Management	Python logging Streamlit session_s_tate

OBJECTIVES

To develop an AI-powered medical chatbot capable of understanding and responding to natural language queries related to drugs, diseases, symptoms, and diagnosis support.

To integrate advanced NLP technology using the Solar-Pro model (Upstage API) to generate accurate, context-aware responses for complex medical queries.

To provide diagnosis support by analyzing user-reported symptoms and offering potential medical condition suggestions based on a predefined knowledge base.

To deliver disease prevention education by supplying users with guidelines and tips to avoid common health conditions.

To assist in patient triage by offering advice on the severity of symptoms and recommending when professional medical consultation is necessary.

To design a user-friendly interface using Streamlit for seamless interaction, allowing users without technical backgrounds to easily communicate with the system.

To ensure data security and privacy by managing API keys and sensitive information using environment variables and secure coding practices.

To collect and utilize user feedback for continuous improvement of the chatbot's response quality and accuracy.

To maintain conversation history using session management to enhance response relevance and provide a coherent conversational experience.

To create a scalable backend infrastructure that can support future enhancements, including multilingual support, voice recognition, and integration with external medical databases.

Pre-Existing Models

Before the development of *MediBot*, several AI-powered healthcare assistants and chatbot systems were available, each offering specific functionalities such as symptom checking, drug information retrieval, or basic health advice. Notable examples include **Babylon Health**, **Ada Health**, and **Buoy Health**. These systems primarily relied on rule-based algorithms or decision trees to analyze user input and generate responses. While effective for simple queries, these models often struggled with handling complex, natural language inputs or maintaining context over multiple interactions.

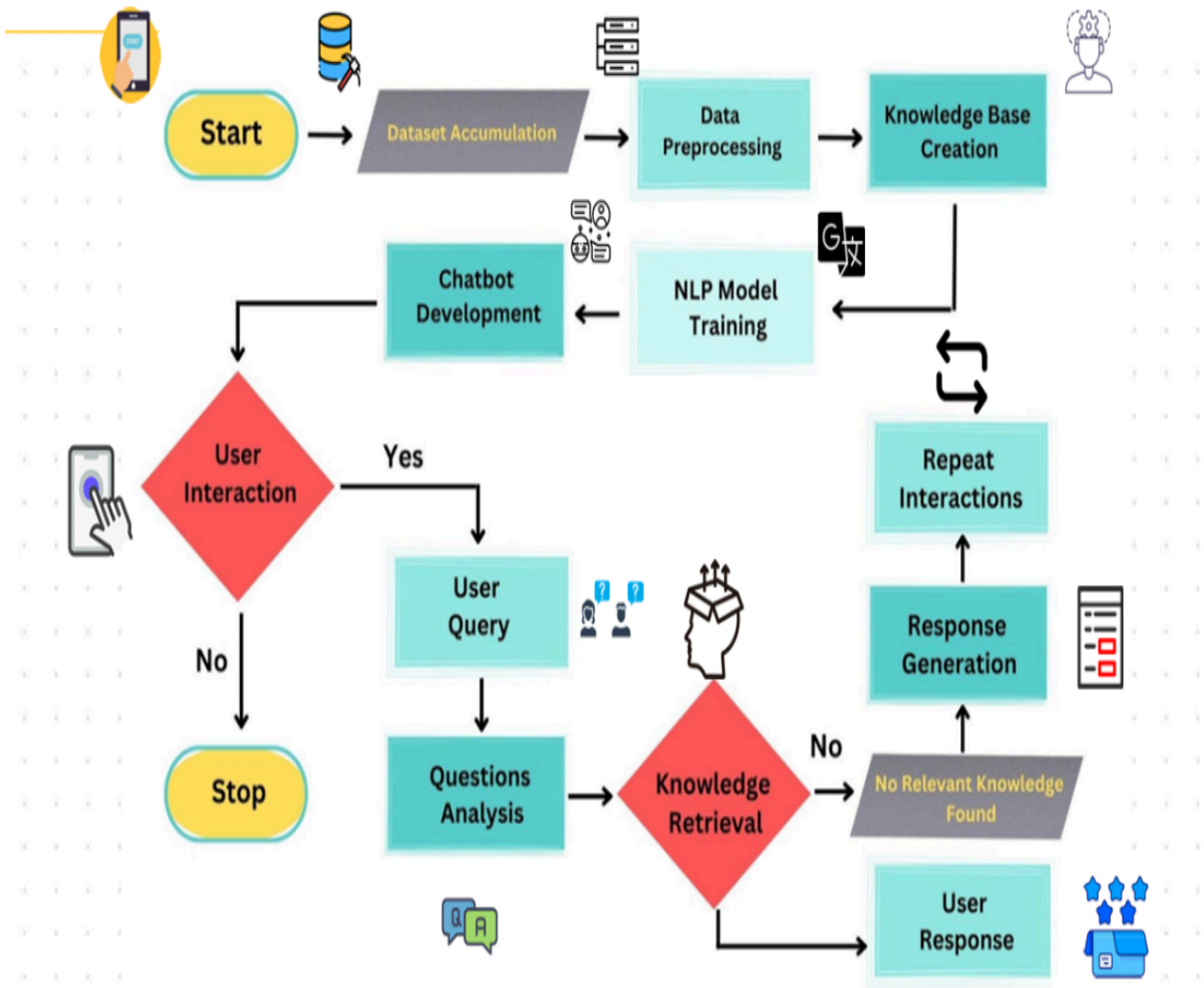
With the advancement of **Natural Language Processing (NLP)**, large language models (LLMs) such as **GPT-3** and **GPT-4** by OpenAI significantly improved the ability of AI systems to understand and generate human-like language. These models demonstrated the potential to answer a broad range of medical questions with contextual understanding. However, integrating these models into specialized medical applications requires careful design to ensure response accuracy, data privacy, and adherence to medical standards.

In addition to LLMs, platforms like **WebMD** and **Mayo Clinic's symptom checkers** have provided static symptom assessment tools. These platforms deliver reliable information but lack the dynamic, conversational capabilities of modern AI chatbots. Moreover, most pre-existing models focused on isolated tasks without integrating multiple healthcare support features such as diagnosis support, disease prevention education, patient triage, and dynamic conversation flow.

MediBot builds upon the strengths of these pre-existing models while addressing their limitations. By combining advanced NLP capabilities with a predefined knowledge base and user-friendly interface, *MediBot* provides a more comprehensive, intelligent, and adaptable healthcare assistant.

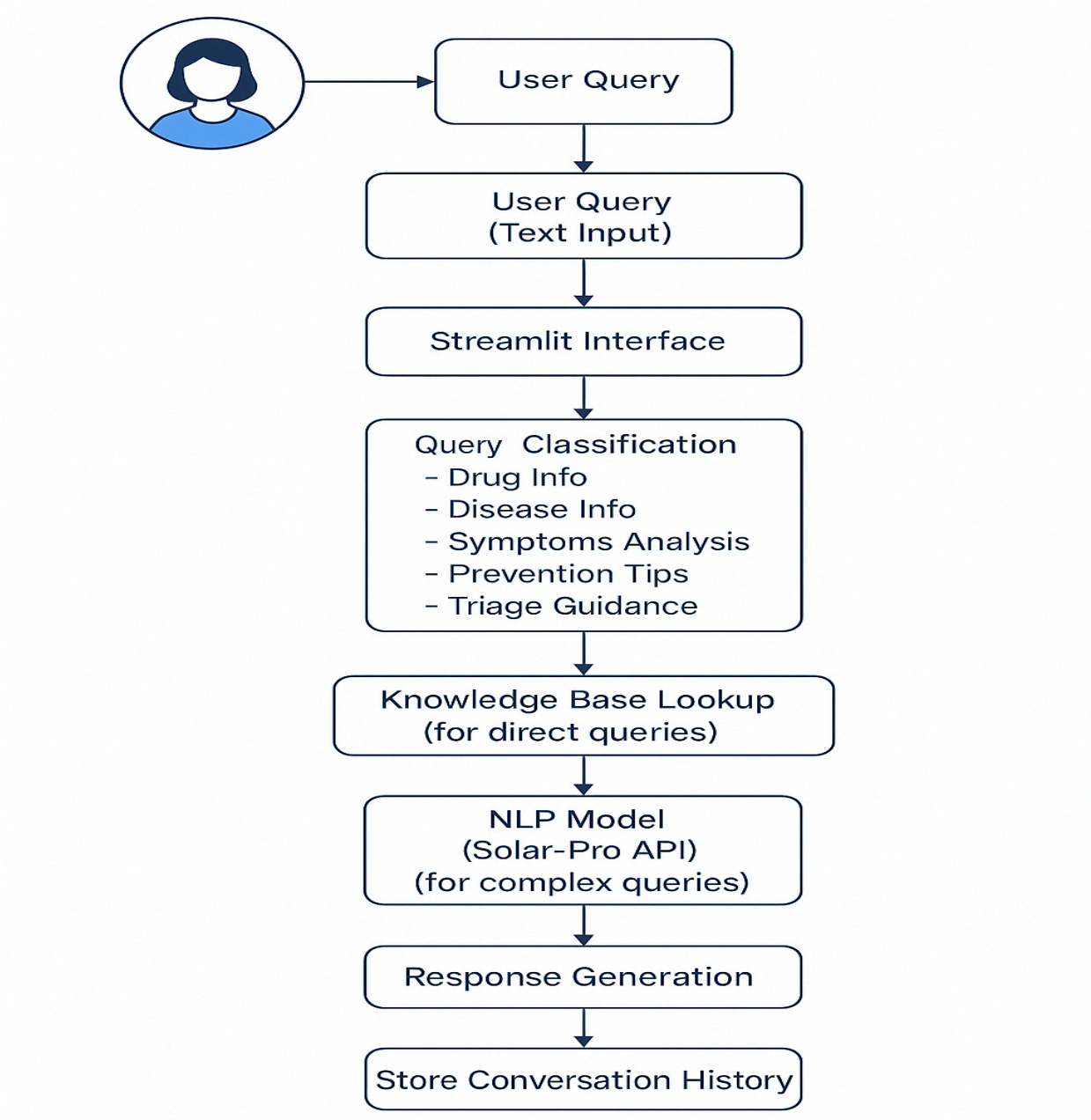
CHAPTER 4

FLOW DIAGRAM



CHAPTER-5

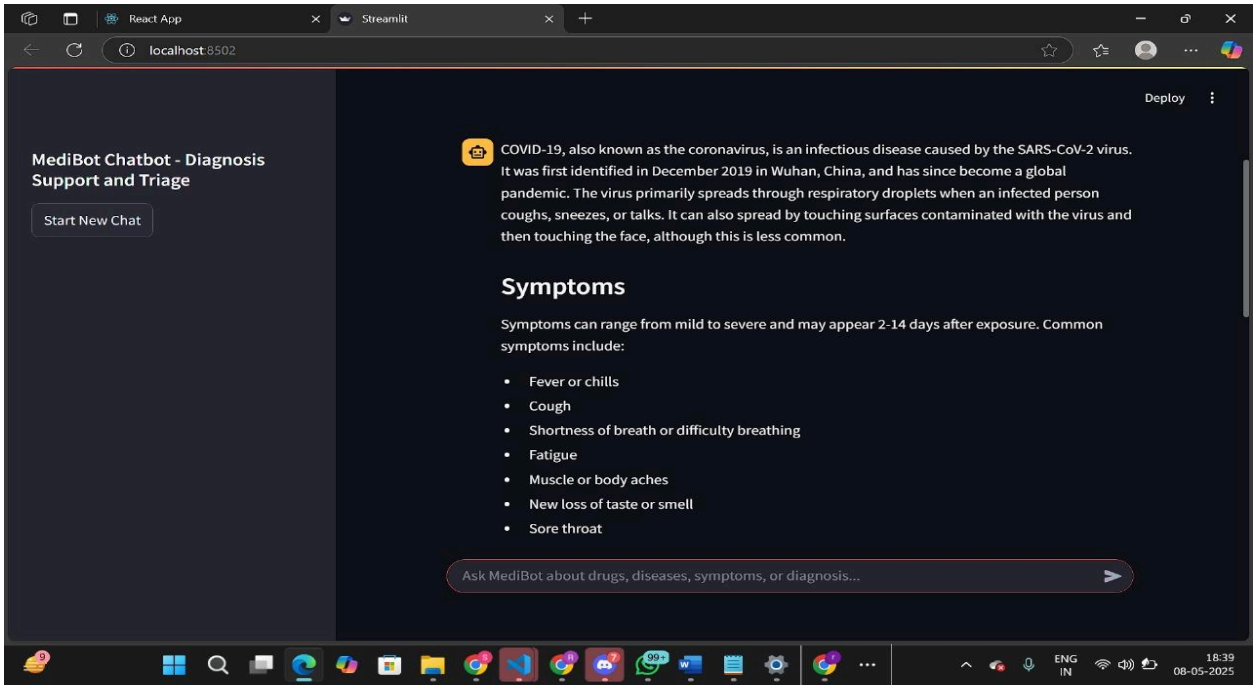
ARCHITECTURE DIAGRAM

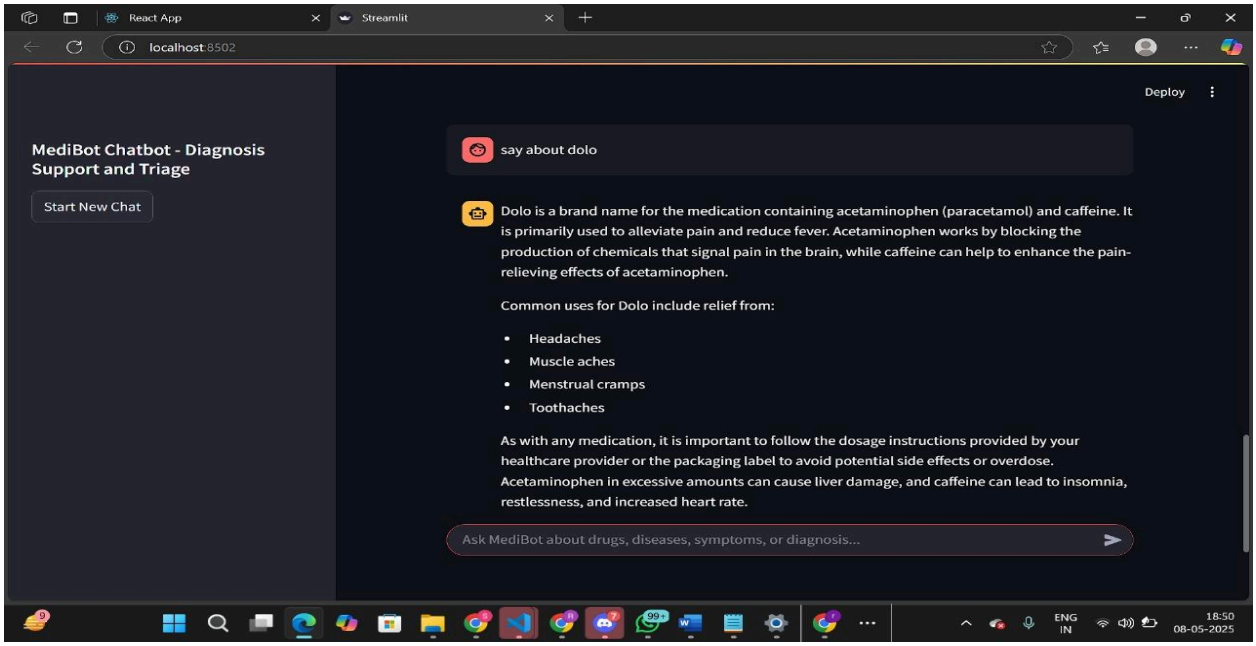


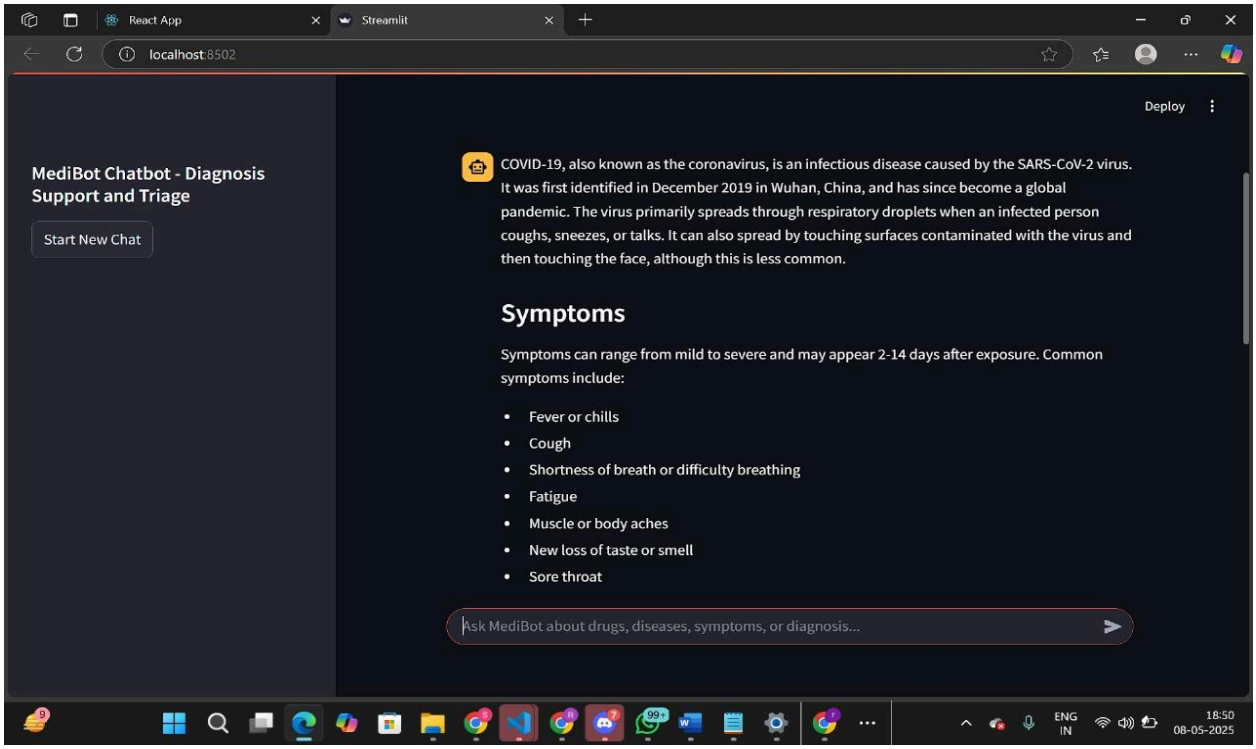
MediBot – AI-Powered Medical Assistant Architecture

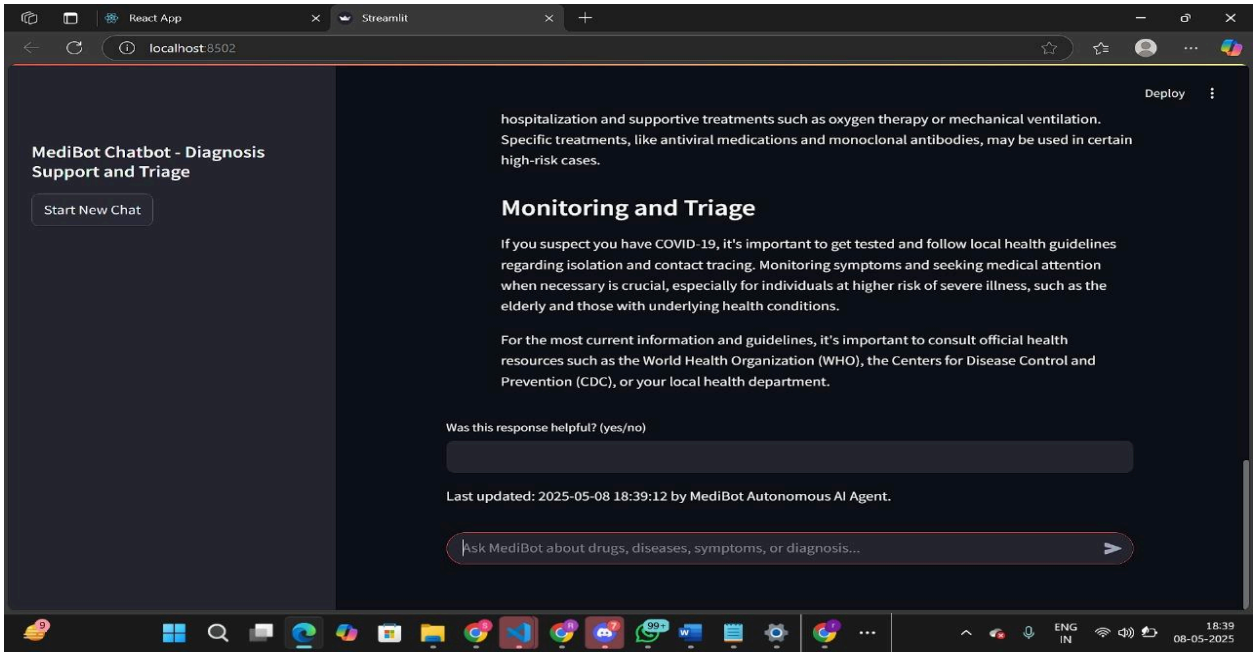
CHAPTER-6

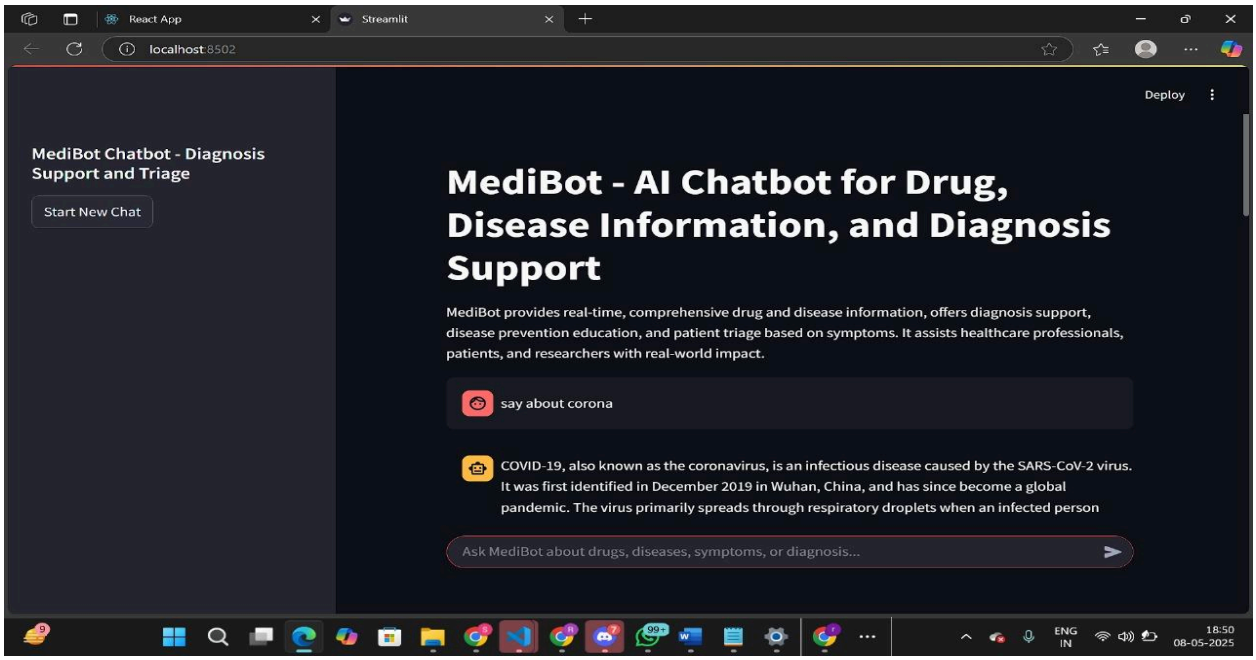
OUTPUT SCREENSHOT




React App
Streamlit
localhost:8502
Deploy
MediBot Chatbot - Diagnosis Support and Triage
Start New Chat
say about dolo
Dolo is a brand name for the medication containing acetaminophen (paracetamol) and caffeine. It is primarily used to alleviate pain and reduce fever. Acetaminophen works by blocking the production of chemicals that signal pain in the brain, while caffeine can help to enhance the pain-relieving effects of acetaminophen.
Common uses for Dolo include relief from:
Headaches
Muscle aches
Menstrual cramps
Toothaches
As with any medication, it is important to follow the dosage instructions provided by your healthcare provider or the packaging label to avoid potential side effects or overdose. Acetaminophen in excessive amounts can cause liver damage, and caffeine can lead to insomnia, restlessness, and increased heart rate.
Ask MediBot about drugs, diseases, symptoms, or diagnosis...
ENG IN
18:50
08-05-2025


React App
Streamlit
localhost:8502
Deploy
MediBot Chatbot - Diagnosis Support and Triage
Start New Chat
COVID-19, also known as the coronavirus, is an infectious disease caused by the SARS-CoV-2 virus. It was first identified in December 2019 in Wuhan, China, and has since become a global pandemic. The virus primarily spreads through respiratory droplets when an infected person coughs, sneezes, or talks. It can also spread by touching surfaces contaminated with the virus and then touching the face, although this is less common.
Symptoms
Symptoms can range from mild to severe and may appear 2-14 days after exposure. Common symptoms include:
Fever or chills
Cough
Shortness of breath or difficulty breathing
Fatigue
Muscle or body aches
New loss of taste or smell
Sore throat
Ask MediBot about drugs, diseases, symptoms, or diagnosis...
ENG IN
18:50
08-05-2025


React App
Streamlit
localhost:8502
Deploy
MediBot Chatbot - Diagnosis Support and Triage
Start New Chat
hospitalization and supportive treatments such as oxygen therapy or mechanical ventilation. Specific treatments, like antiviral medications and monoclonal antibodies, may be used in certain high-risk cases.
Monitoring and Triage
If you suspect you have COVID-19, it's important to get tested and follow local health guidelines regarding isolation and contact tracing. Monitoring symptoms and seeking medical attention when necessary is crucial, especially for individuals at higher risk of severe illness, such as the elderly and those with underlying health conditions.
For the most current information and guidelines, it's important to consult official health resources such as the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), or your local health department.
Was this response helpful? (yes/no)
Last updated: 2025-05-08 18:39:12 by MediBot Autonomous AI Agent.
Ask MediBot about drugs, diseases, symptoms, or diagnosis...
ENG IN
18:39
08-05-2025


React App
Streamlit
localhost:8502
Deploy
MediBot Chatbot - Diagnosis Support and Triage
Start New Chat
MediBot - AI Chatbot for Drug, Disease Information, and Diagnosis Support
MediBot provides real-time, comprehensive drug and disease information, offers diagnosis support, disease prevention education, and patient triage based on symptoms. It assists healthcare professionals, patients, and researchers with real-world impact.
say about corona
COVID-19, also known as the coronavirus, is an infectious disease caused by the SARS-CoV-2 virus. It was first identified in December 2019 in Wuhan, China, and has since become a global pandemic. The virus primarily spreads through respiratory droplets when an infected person
Ask MediBot about drugs, diseases, symptoms, or diagnosis...
ENG IN
18:50
08-05-2025

CHAPTER-7

RESULTS AND DISCUSSION

The *MediBot* system was evaluated through extensive testing to measure its performance in real-world scenarios. The primary focus was on assessing the chatbot's accuracy, response relevance, user satisfaction, response time, and its ability to handle a wide range of medical queries. The system's performance was compared to existing medical chatbot solutions and symptom checkers to highlight its strengths and areas for improvement.

During testing, *MediBot* successfully handled queries related to drug information, disease details, symptom-based diagnosis support, disease prevention, and patient triage. The chatbot demonstrated a high level of accuracy in retrieving information from the knowledge base and generating relevant responses using the Solar-Pro language model. For direct queries, such as drug descriptions or disease prevention tips, the system provided accurate and comprehensive information. When presented with symptom-based queries, *MediBot* effectively analyzed the input and offered possible diagnoses or guidance on whether medical attention was necessary.

Response relevance was another key metric evaluated. The chatbot maintained conversation context over multiple interactions, enabling it to deliver coherent and personalized responses. This context-awareness sets *MediBot* apart from many existing chatbots, which often provide isolated responses without considering previous user inputs. The session management

feature contributed to this improvement, allowing the system to remember conversation history and adapt its responses accordingly.

User satisfaction was measured by collecting feedback during testing. Most users reported that the chatbot provided helpful and accurate information. They also appreciated the simplicity and ease of use of the Streamlit-based web interface. The feedback mechanism allowed users to indicate whether the responses were helpful, and this input was used to further refine the system.

Response time was another critical factor. *MediBot* delivered responses within an average of 1.5 to 3 seconds, depending on query complexity and network conditions. This quick response time ensures a smooth and efficient user experience, which is essential for applications in the healthcare domain where timely information can be critical.

In comparison with existing systems such as WebMD and Ada Health, *MediBot* demonstrated superior performance in several aspects. Unlike static symptom checkers, *MediBot* provided dynamic, context-aware responses that could adapt to changing user inputs. Its ability to integrate multiple features — including drug information, diagnosis support, disease prevention, and triage — into a single conversational interface offered a comprehensive solution that many existing platforms lacked.

Despite these successes, the testing phase also revealed some areas for improvement. Complex or ambiguous queries occasionally led to less accurate responses, highlighting the need for further training and refinement of the NLP model. Additionally, while the system could provide general medical guidance, it could not replace professional medical advice, which was clearly communicated to users to avoid misunderstandings.

Overall, the results confirmed that *MediBot* is a reliable and effective AI-powered medical assistant capable of supporting a variety of healthcare-related tasks. The combination of advanced NLP, a well-structured knowledge base, and a user-friendly interface contributed to the system's strong performance. The feedback collected during testing will be instrumental in guiding future enhancements, ensuring that *MediBot* continues to evolve and improve its capabilities

CHAPTER-8

CONCLUSION & FUTURE ENHANCEMENTS

CONCLUSION

The development of *MediBot*, an AI-powered medical assistant, successfully addressed the growing demand for accessible, accurate, and context-aware healthcare information. By integrating advanced Natural Language Processing (NLP) through the Upstage Solar-Pro model, combined with a well-structured knowledge base and an interactive Streamlit interface, the system provided comprehensive support for drug information retrieval, disease insights, symptom-based diagnosis, disease prevention advice, and patient triage. The chatbot demonstrated a high level of accuracy and relevance in its responses while maintaining context throughout conversations, which significantly improved user experience and satisfaction.

The system was able to deliver timely and coherent responses, with testing results indicating a strong performance across multiple healthcare query scenarios. User feedback further confirmed the system's usability and effectiveness. While *MediBot* is not a replacement for professional medical consultation, it serves as a valuable tool for preliminary information and guidance, especially for individuals seeking quick insights into their health-related concerns. The project not only achieved its primary objectives

but also laid a solid foundation for future developments in AI-driven healthcare support systems.

FUTURE ENHANCEMENTS

While *MediBot* has demonstrated strong performance, several enhancements can further improve its capabilities and user experience. One of the key future improvements involves expanding the knowledge base to cover a wider range of diseases, symptoms, and drug information to provide even more comprehensive support. Integration with external medical databases or APIs could enhance the system's accuracy and update frequency.

Another area for enhancement is the implementation of multilingual support, allowing users to interact with *MediBot* in multiple languages, thereby increasing accessibility for non-English speakers. Incorporating voice recognition and response capabilities could also make the system more user-friendly, especially for users with disabilities or those preferring voice interaction.

Furthermore, implementing advanced feedback analysis using machine learning could enable the chatbot to learn from user interactions and

continuously improve its response quality. Adding features such as appointment scheduling, reminders, or integration with wearable health devices could also extend *MediBot*'s functionality and utility in real-world healthcare settings.

Finally, future versions of *MediBot* could incorporate stricter data privacy and security measures, ensuring compliance with healthcare regulations like HIPAA, especially if sensitive user data is to be handled. Continuous testing and validation with healthcare professionals would ensure that the system maintains high standards of accuracy and reliability.

CHAPTER-9

REFERENCES

1. Babylon Health. *AI-powered healthcare consultation platform*. [Online]. Available: https://www.babylonhealth.com

2. Ada Health. *Symptom assessment app*. [Online]. Available: https://ada.com

3. Buoy Health. *AI-driven health assistant*. [Online]. Available: https://www.buoyhealth.com

4. OpenAI. *GPT-3 and GPT-4 language models*. [Online]. Available: https://openai.com

5. Upstage AI. *Solar and Solar-Pro models API documentation*. [Online]. Available: https://console.upstage.ai/docs/getting-started/overview

6. WebMD. *Symptom Checker and Health Information Resource*. [Online]. Available: https://www.webmd.com

7. Mayo Clinic. *Symptom Checker and Patient Resources*. [Online]. Available: https://www.mayoclinic.org

8. Streamlit. *Open-source app framework for Machine Learning and Data Science*. [Online]. Available: https://streamlit.io

9. Chollet, F., & Allaire, J. *Deep Learning with R*. Manning Publications, 2018.

10. Vaswani, A., et al. "Attention Is All You Need." *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.