PROJECT REPORT

Adaptive Emotion-Guided Healthcare Chatbot

BY-

GOUTHAMI CHELLURI (48345377)

NIHARIKA SHILAMAMIDI (48182540)

Table of Contents

1. Introduction

As global healthcare systems evolve, the integration of artificial intelligence (AI) technologies into patient communication and support services is becoming increasingly critical. Healthcare chatbots have gained prominence for their ability to provide timely information, triage symptoms, and support health education, all while alleviating pressure on medical personnel. This project presents a fully functional, adaptive emotion-aware chatbot that leverages Natural Language Processing (NLP) and retrieval-augmented generation (RAG) to offer real-time, contextually relevant health information.

This chatbot operates across three primary modes—Symptom Checker, Prescription Explainer, and Health Literacy Tutor, each designed to handle specific user needs. The underlying architecture is composed of a Streamlit-based user interface, an intermediate processing pipeline for contextual retrieval and prompt formation, and a generative backend powered by LLaMA 3 served via Ollama. The inclusion of emotion-guided responses allows the chatbot to personalize interactions, enhancing user comfort and comprehension.

2. Project Objectives

The project was designed with the following objectives in mind:

- Build a healthcare chatbot capable of handling diverse user needs such as checking symptoms, explaining prescriptions, and enhancing health literacy.
- Integrate user emotions into the dialogue generation process to produce emotionally intelligent and context-sensitive responses.
- Employ retrieval-augmented generation (RAG) to ground chatbot outputs in factual, retrievable information.
- Utilize the LLaMA 3 model through the Ollama backend for local, efficient large language model inference.
- Design an intuitive user interface using Streamlit to facilitate accessibility and interaction.
- Ensure modularity and scalability for future improvements and deployment.

3. System Architecture

The architecture of the chatbot system is composed of three integrated layers:

1. Frontend (User Interface): Developed using Streamlit, the interface provides options for selecting the chatbot mode and emotional tone, entering queries, and viewing both the retrieved context and the generated responses.
2. Processing Pipeline: This layer handles keyword extraction, contextual document retrieval, prompt generation, and emotion-based message templating.
3. Response Generation Layer: The backend integrates with the LLaMA 3 model via Ollama, sending crafted prompts and returning user-specific, informative responses.

These components are loosely coupled, enabling independent updates and easier debugging.

4. Chatbot Modes and Functional Overview

4.1 Symptom Checker

In this mode, users can describe their symptoms in natural language. The system identifies key health terms using basic keyword-matching techniques and retrieves relevant context from stored health data. This context, along with the user's query and emotional state, is used to form a detailed prompt that is fed to the LLaMA model. The result is an emotionally guided, medically plausible response that helps users understand potential health conditions without providing formal diagnoses.

4.2 Prescription Explainer

Users input the name of a drug, or a phrase related to medication. The system retrieves textual context from known drug databases such as DrugBank and RxNorm. This includes side effects, usage guidelines, contraindications, and drug interactions. The chatbot then produces a simplified, patient-friendly explanation. Emotionally anxious users receive gentle, reassuring explanations, while confused users get more detailed breakdowns.

4.3 Health Literacy Tutor

In this educational mode, users can input complex medical terminology or concepts. The system provides layperson-friendly explanations of medical concepts. This feature is useful for patients trying to understand lab reports, diagnoses, or technical literature. Sources such as MedQuad and PubMedQA are intended for contextual grounding.

5. Emotion-Aware Interaction

A core innovation of this chatbot is its ability to adjust tone and explanation depth based on the user's selected emotional state. The user selects their current mood—neutral, anxious, or confused—via a dropdown. This selection maps to a system prompt that sets the chatbot's tone.

For example, a user in an anxious state might receive the prompt: "You are a comforting medical assistant who gently explains things to anxious users." This is combined with the retrieved context and the user's question to form a comprehensive prompt for LLaMA.

This approach provides a flexible and empathetic framework for dialogue generation, improving user trust and engagement.

6. Retrieval-Augmented Generation (RAG) and Prompt Construction

To support fact-based responses, the chatbot retrieves text snippets related to the user's query using keyword-based matching. While basic in this prototype, this retrieval mimics a real semantic search pipeline and can be upgraded with FAISS or Haystack.

The retrieved context is then used to construct a prompt in the following format:

"""

You are a medical chatbot. Based on the context and the user question, provide an informative answer.

Context: [Retrieved Context]

Emotion: [Retrieved Emotion]

Question: [User Query]

Answer:

"""

This prompt is sent to the LLaMA 3 model running locally via Ollama, ensuring data privacy and low latency. Responses are parsed and presented to the user with the retrieved context shown alongside for transparency.

7. Technologies and Libraries

- Python (backend logic and integration)
- Streamlit (frontend web interface)
- Ollama (lightweight local model inference)
- LLaMA 3 (generative language model)
- PubMedQA / MedQuad (data sources)
- Prompt engineering for emotion and context integration

8. Testing and Output Examples

To validate functionality, the chatbot was tested across all three operational modes using a range of emotional contexts. Key observations include:

- Symptom Checker responses were medically plausible and adjusted in tone based on emotional input.
- Prescription explanations-maintained accuracy and readability, especially under the 'confused' emotion setting.
- Health Literacy explanations were simplified effectively, aiding users in understanding complex medical concepts.

9. Evaluation and Limitations

To objectively assess the quality of the chatbot's generated responses, we implemented an automatic evaluation mechanism using two widely accepted Natural Language Generation (NLG) metrics: ROUGE and BLEU. This function computes:

ROUGE (Recall-Oriented Understudy for Gisting Evaluation): A set of metrics for evaluating automatic summarization and machine translation. It measures the overlap between the generated (prediction) and ground truth (reference) text in terms of n-grams, word sequences, and word pairs.

BLEU (Bilingual Evaluation Understudy): A precision-based metric that evaluates how many n-gram sequences in the predicted output match the reference output. Originally designed for machine translation, it’s also used in chatbot response assessment.

Both metrics provide a quantifiable way to compare a predicted answer to a reference (ideal) answer and are helpful during model tuning and benchmarking.

```
example output:
{
   "rouge": {
      "rouge1": {"precision": 0.625, "recall": 0.714, "fmeasure": 0.666},
      ...
   },
   "bleu": {
      "bleu": 0.487
   }
}
```

The system demonstrates strong performance in generating informative, empathetic responses. However, several limitations remain:

- Document retrieval is not semantic or learned—reliant on keyword matches.
- Emotion must be manually selected and is not inferred from user text.
- No persistent memory or multi-turn dialogue handling.
- Only supports English and assumes general medical literacy.
- Real-time updates (e.g., from APIs) are not yet implemented.

10. Future Enhancements

- Replace keyword retrieval with Haystack vector search using sentence embeddings.
- Integrate sentiment analysis (e.g., using BERT or RoBERTa) to auto-detect user emotion.
- Expand to support multi-turn conversations with context tracking.
- Deploy on cloud (e.g., Streamlit Community Cloud or HuggingFace Spaces) with secure access.
- Add support for multilingual queries and responses.

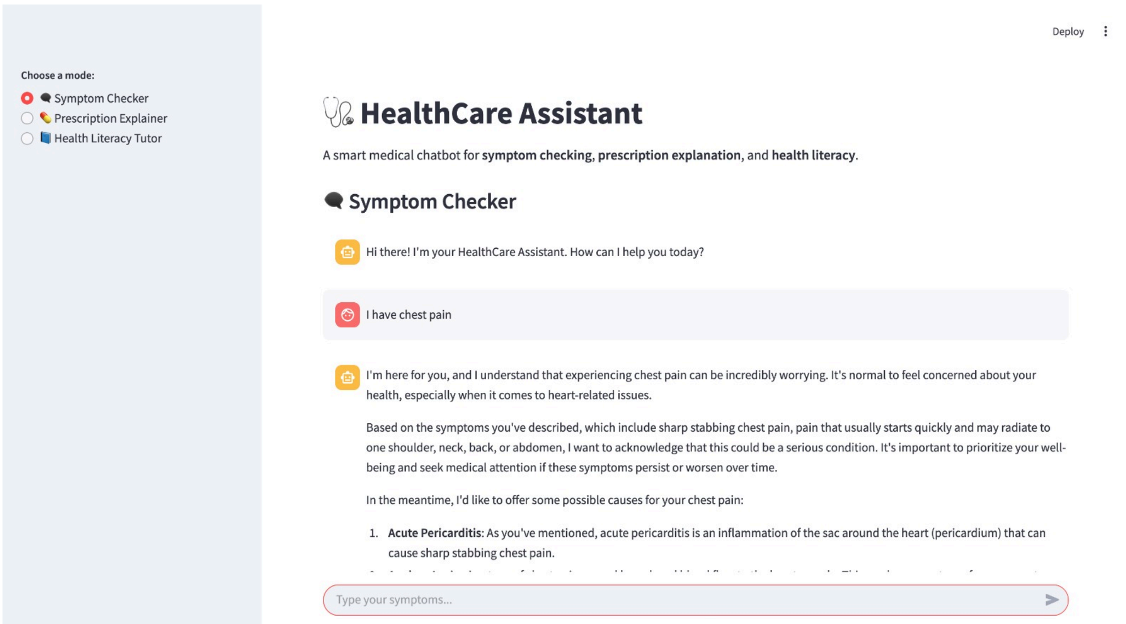
11. Conclusion

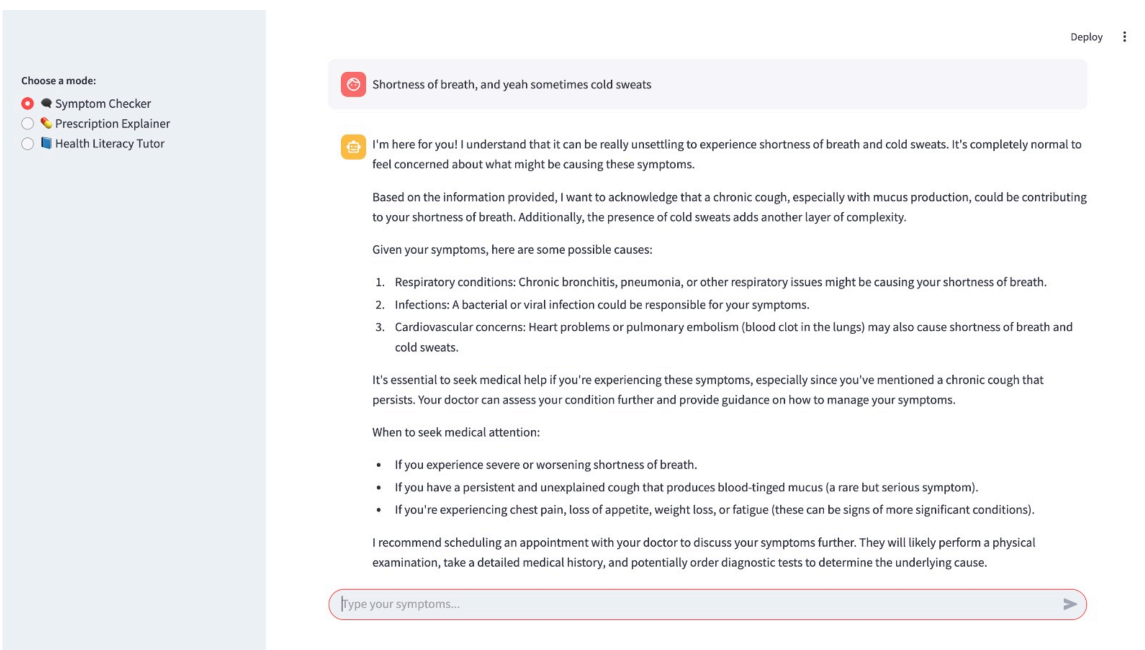
This project presents a comprehensive, modular healthcare chatbot prototype capable of interpreting user inputs, adapting responses based on emotional context, and delivering context-rich, understandable answers. With its flexible design and integration of modern NLP techniques, this chatbot serves as a strong foundation for future medical AI tools aimed at improving health communication and education.

12. Appendix: Chatbot Logs and Screenshots

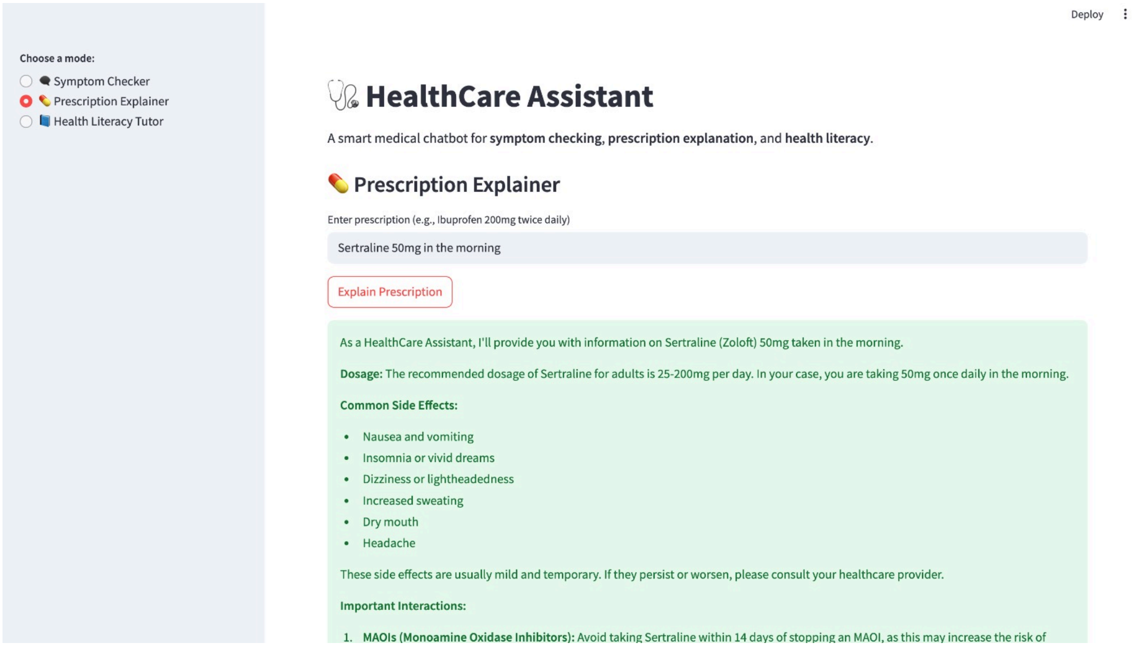
Attaching the test results, interaction transcripts, and output screenshots here.

1. Symptom Checker Chatbot

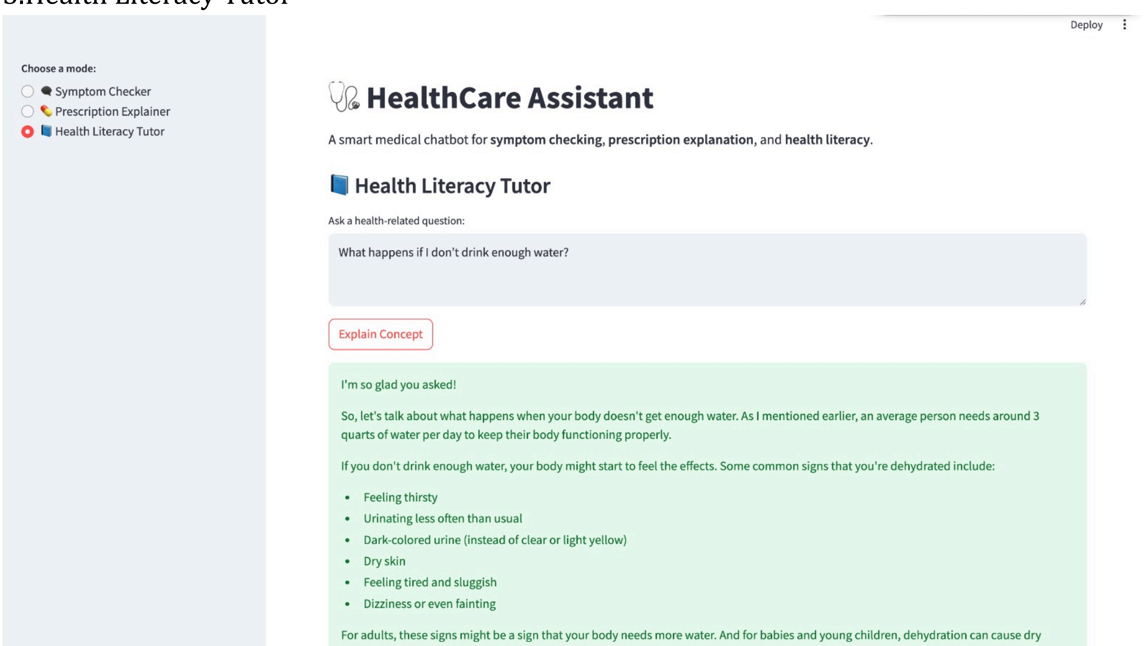




2. Prescription Explainer



3.Health Literacy Tutor



13. Team Contribution

Gouthami Chelluri implemented the Streamlit frontend, integrated the LLaMA 3 backend via Ollama, and developed the chatbot logic.

Niharika Shilamamidi worked on the emotion-aware prompt engineering, retrieval module, and contributed to system testing and evaluation.