

## **Project Title:**

### **Medical Chartbot for Heart Health**

## **Objective:**

To create a chatbot powered by BioMistral LLM that serves as a medical assistant specifically focused on heart health. The bot leverages natural language processing (NLP) to interpret and respond to user queries about heart disease risk factors, preventive measures, and general heart health information.

## **Components and Workflow:**

### **1. Dataset Loading:**

- The project uses a heart health PDF file loaded through Google Drive, containing information on heart disease, risk factors, and prevention.
- Each page in the PDF is treated as a separate document for processing.

### **2. Installation of Required Libraries:**

- Essential libraries installed include LangChain, Sentence Transformers, ChromaDB, and Llama-cpp.
- Libraries handle text processing, embedding creation, and similarity search for question answering.

### **3. Document Splitting and Embedding Creation:**

- The PDF text is split into chunks of 300 characters, with 50 characters overlapping to ensure context retention.
- The **SentenceTransformer** model (NeuML/pubmedbert-base-embeddings) is used to create embeddings, facilitating efficient similarity searches for user queries.

### **4. Vector Store Creation:**

- The Chroma vector store is built using the generated embeddings, allowing similarity search functionality to retrieve relevant sections of the document in response to user questions.

### **5. LLM Model Loading and Query Processing:**

- **BioMistral-7B** model is loaded using **LlamaCpp** with specific configurations, including parameters such as max\_tokens and temperature.
- A template is defined for generating responses, setting the bot's tone as a helpful medical assistant.

### **6. Query and Response Mechanism:**

- The chatbot is set up to take user queries, perform similarity searches against document chunks, and generate responses through the LLM.

- Queries about heart disease risks, prevention, and health-related conditions return relevant information from the heart health data, structured in a conversational format.

### **Example Queries and Responses:**

#### **1. "Who is at risk of heart disease?"**

- The chatbot returns a summary of risk factors, such as age, family history, lifestyle habits, and health conditions (e.g., high blood pressure, diabetes).

#### **2. "What are the diseases that affect heart health?"**

- Response includes common heart-related conditions, like hypertension, high cholesterol, coronary artery disease, and congestive heart failure.

#### **3. "How does high cholesterol affect heart health?"**

- The chatbot explains that high cholesterol increases the risk of coronary artery disease, which can lead to serious heart conditions.

### **Challenges and Solutions:**

#### **1. Dependency Conflicts:**

- During installation, dependency issues with packages like protobuf were noted, causing compatibility warnings. Solutions could involve setting specific versions for compatibility.

#### **2. Bot Exit Mechanism:**

- The code includes a while loop for continuous query processing, with an 'exit' command to end the session. A warning suggests using exit, quit, or Ctrl-D for smoother termination.

### **Future Improvements:**

#### **• Enhanced Risk Factor Assessment:**

- Integrate real-time data sources or user-specific risk assessments.

#### **• Broader Health Topics:**

- Extend the chatbot's knowledge base to include other health domains or chronic diseases for a more comprehensive medical assistant.