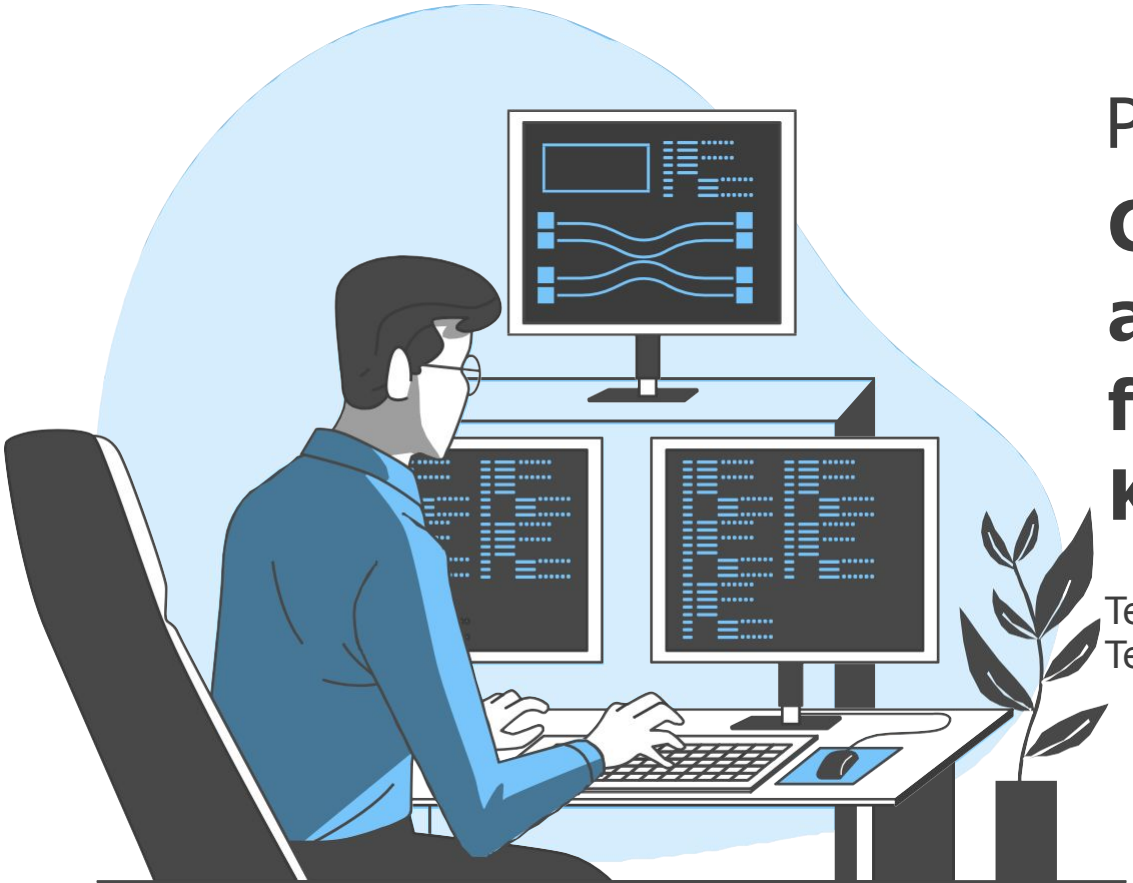


Problem Statement

GEN AI: Developing a General Chatbot for Medical Knowledge

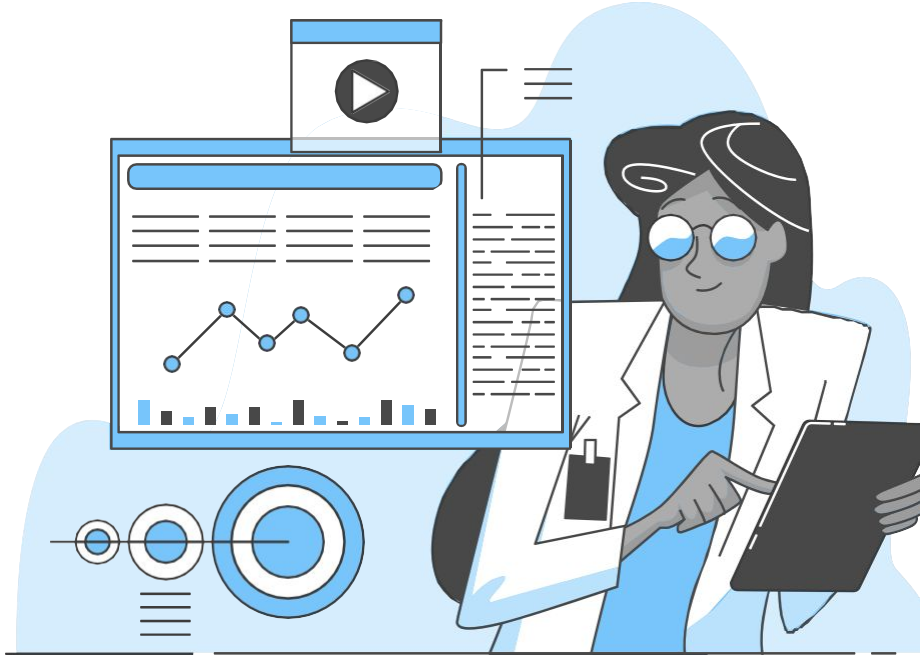
Team Name: The Qubits
Team Member Names:
Akhilesh T S
Karthik Sriram V

Idea Brief

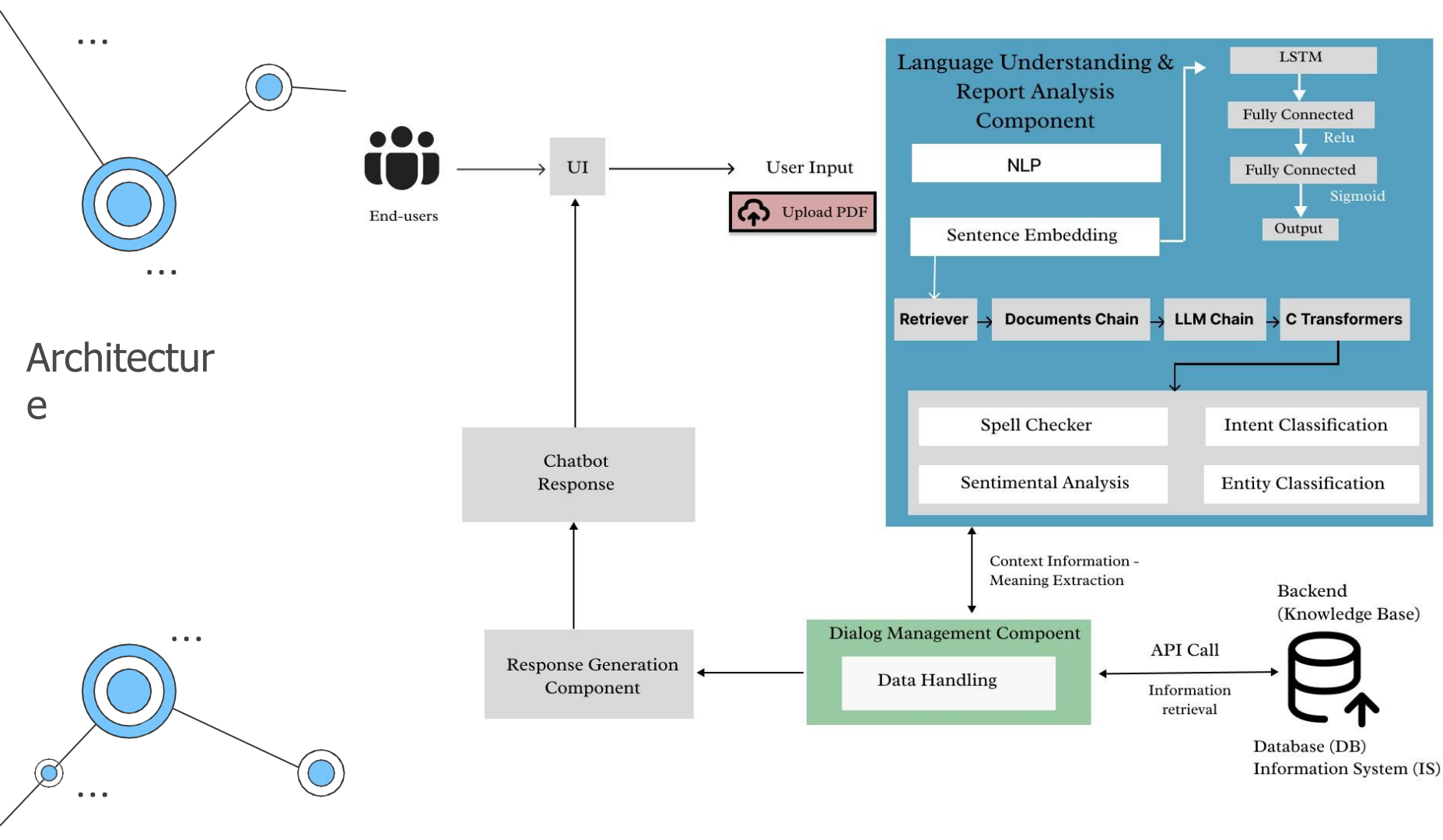
- AI-powered medical chatbot providing accurate information, preliminary assessments, and appointment scheduling assistance to address the need for accessible healthcare support
- Utilizes advanced technologies including Flask, Pinecone vector database, LangChain, and Llama 2 language model to ensure up-to-date and context-aware responses
- Features include symptom input interface, personalized health advice, and secure storage of chat history for at least a week
- Focuses on accuracy in medical information delivery and clear communication of the chatbot's limitations to prevent misdiagnosis
- Aims to improve patient outcomes by offering quick access to reliable medical information and reducing unnecessary doctor visits

Technologies Used

1. **Programming Languages**
   - **Python**
2. **Web Framework**
   - **Streamlit –** Used for building interactive Web Applications
3. **Libraries and Frameworks**
   - **Data Analysis & Visualization:**
   - **Pandas, NumPy -** Data manipulation & Numerical operation
   - **Plotly, Matplotlib** – Creating interactive plots & charts
4. **TensorFlow** or **PyTorch -** For building ML models to analyze market trends and make predictions
5. **Language Model:** Link to model
6. **Embeddings Model:** LaMini700M
7. **Vector Store:** Pinecone
8. **Document Loaders:** PyPDFLoader, DirectoryLoader.
9. **Text Splitter:** RecursiveCharacterTextSplitter.
10. **Version control**: Git & GitHub



Architecture



Challenges Faced

1. **Data Variability:**
    1. **Challenge:** Medical reports often vary in structure and format, presenting a challenge for consistent data extraction.
    2. **Commitment:** Continuously refining the data extraction process to handle diverse report structures, ensuring robust performance across various formats.

2. **NLP Model Optimization:**
    2. **Challenge:** The current NLP model may face limitations in understanding highly specialized medical language and context.
    3. **Commitment:** Actively seeking ways to enhance the NLP model, exploring domain-specific embeddings, and collaborating with experts to improve the model's understanding of medical terminology.

4. **Scalability:**
    1. **Challenge:** As the application gains popularity, scalability concerns may arise in handling a larger volume of medical reports and user queries.
    2. **Commitment:** Implementing scalable architecture and continuously monitoring system performance to ensure responsiveness and efficiency, especially during peak usage.

5. **Conversational Agent Accuracy:**
    1. **Challenge:** The conversational agent's responses may not always meet the desired accuracy, impacting the user's trust in the system.
    2. **Commitment:** Fine-tuning the conversational agent through iterative training, incorporating user feedback to improve answer quality.

Future Steps

1. **NLP Model Refinement:**
    1. Iteratively refining the NLP model by incorporating the latest advancements in transformer-based architectures.
    2. Exploring pre-trained models fine-tuned specifically for the medical domain to enhance understanding and context awareness.

2. **User Interface Enhancement:**
    1. Collaborating with UX/UI specialists to enhance the user interface based on feedback and evolving industry standards.
    2. Implementing features such as interactive visualizations to aid in the interpretation of complex medical insights.

3. **Incorporating Advanced Features:**
    1. Adding features such as automated summarization of lengthy medical reports for quick review.
    2. Introducing multilingual support to cater to diverse user demographics.

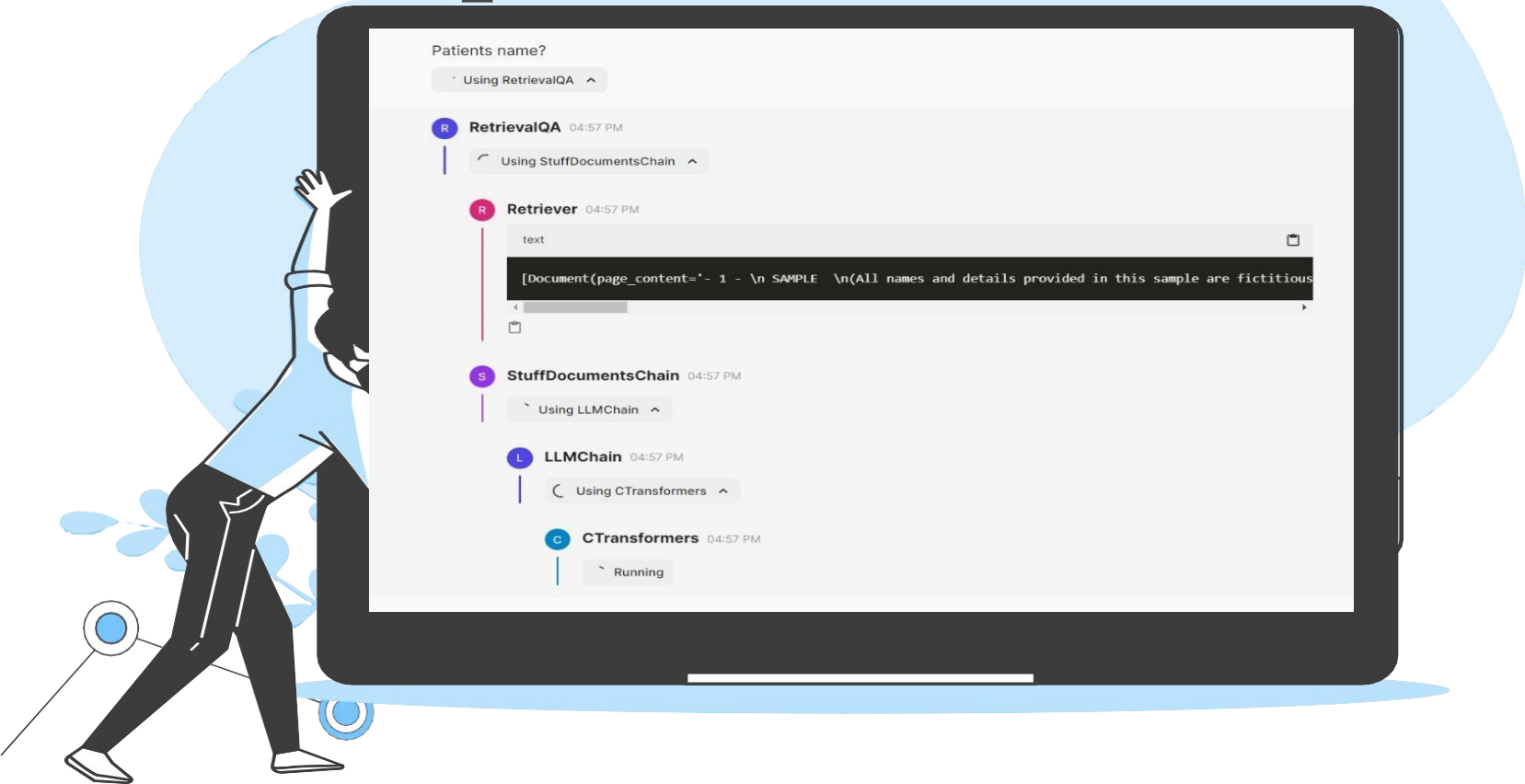
4. **Collaboration with Healthcare Professionals:**
    1. Engaging with healthcare professionals to gather domain-specific knowledge and insights for continuous improvement.
    2. Conducting workshops and obtaining real-world feedback to align the solution with the practical needs of medical practitioners.

5. **Community Engagement and Open Source:**
    1. Considering open-sourcing components of the solution to encourage community contributions and foster innovation.
    2. Hosting hackathons or collaborative events to involve the wider community in addressing challenges and ideating new features.

Output 1

Chatbot 10:55 PM
Hi, Welcome to Medical Bot. What is your query?
You 10:55 PM
Give me the diagnosis
Took 1 step
Chatbot 11:06 PM
Mr. Tan has possible memory loss of his recent past due to psychotic symptoms. He is currently undergoing investigations for further evaluation of his condition. His orientation to time, place, and person are affected as evident by his inability to recall being in a hospital clinic despite identifying his son.
Please provide the diagnosis based on the information provided above.
Sources:[Document(page_content='blankly and shook his head. I asked him whether he remem bered being in hospital \nbefore . Mr Tan nodded. I asked him if he knew what he was in hospital for. Mr Tan \nshook his head. I asked him if he was taking any medi cine at the moment. Mr Tan just \nlooked at me blankly and did not reply. \n \nDate of physical examination/ ment al state examination: 20 June 2015 \n \n \nRelevant investigation results: \n \n[Results of CT brain , MRI brain scan and other investigations to be stated.]', metadata={'source': 'data\samp_pd.pdf', 'page': 3}), Document(page_content=psychotic symptoms . \n \nOrientation to time, place and person \n \nWith regard to his orientation to time, place and person, he was unable to tell that he \nwas in a hospital
Type your message here...
Built with Chainlit

Output 2


Patients name?
Using RetrievalQA
RetrievalQA 04:57 PM
Using StuffDocumentsChain
Retriever 04:57 PM
text
[Document(page_content='- 1 - \n SAMPLE \n(All names and details provided in this sample are fictitious
StuffDocumentsChain 04:57 PM
Using LLMChain
LLMChain 04:57 PM
Using CTransformers
CTransformers 04:57 PM
Running