AI Medical Chatbot using RAG (MediBot)

PROJECT LAYOUT

Phase 1–Setup Memory for LLM (Vector Database)

- Load raw PDF(s)
- Create Chunks
- Create Vector Embeddings
- Store embeddings in FAISS

Phase 2–Connect Memory with LLM

- Setup LLM (Mistral with HuggingFace)
- Connect LLM with FAISS
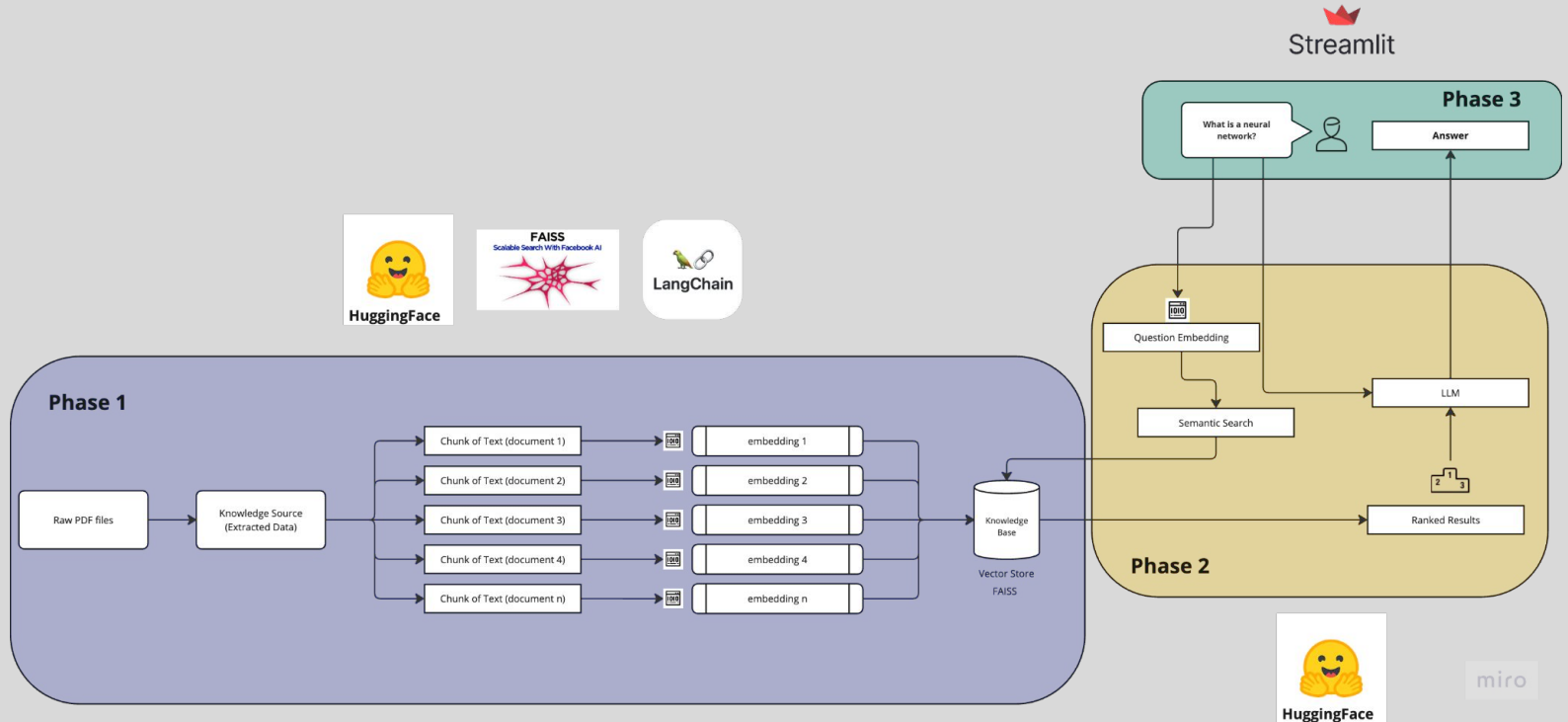- Create chain

Phase 3–Setup UI for the Chatbot

- Chatbot with Streamlit
- Load Vector store (FAISS) in cache
- Retrieval Augmented Generation–RAG

Tools & Technologies

- Langchain (AI Framework for LLM applications)
- HuggingFace (ML/AI Hub)
- Mistral (LLM Model)
- FAISS (Vector Database)
- Steamlit (For Chatbot UI)
- Python (Programming Language)
- VS Code (IDE)

Technical Architecture

Improvement Potential/Next Steps

- Add authentication in the UI
- Make use of self-upload document functionality
- Add multiple documents and embed them together
- Add Unit testing of RAG applications

Summary

- Modern AI Chatbot for Documents
- Modular 3 phased Chatbot project
- Talked about
  - Streamlit | Langchain | HuggingFace
  - RAG
  - Vector Embeddings
  - End to End RAG pipeline
- Future Possibilities
- Everything (Code and all dependencies) will be shared in the comments section
- Any sort of Feedback will be highly appreciated