

RAGenius

Where Questions meet Intelligent Reasoning 

-Mohammed Muzammil

Project Layout

Phase1: Setup UI with Streamlit

- Setting-up Upload PDF functionality
- Chatbot Skeleton (Question & Answer)

Phase2: Setup Vector Database (Memory for LLM)

- Upload & Load raw PDF(s)
- Create Chunks
- Setup Embeddings Model (**Use DeepSeek R1 with Ollama**)
- Index Documents **Store embeddings in FAISS (vector store)

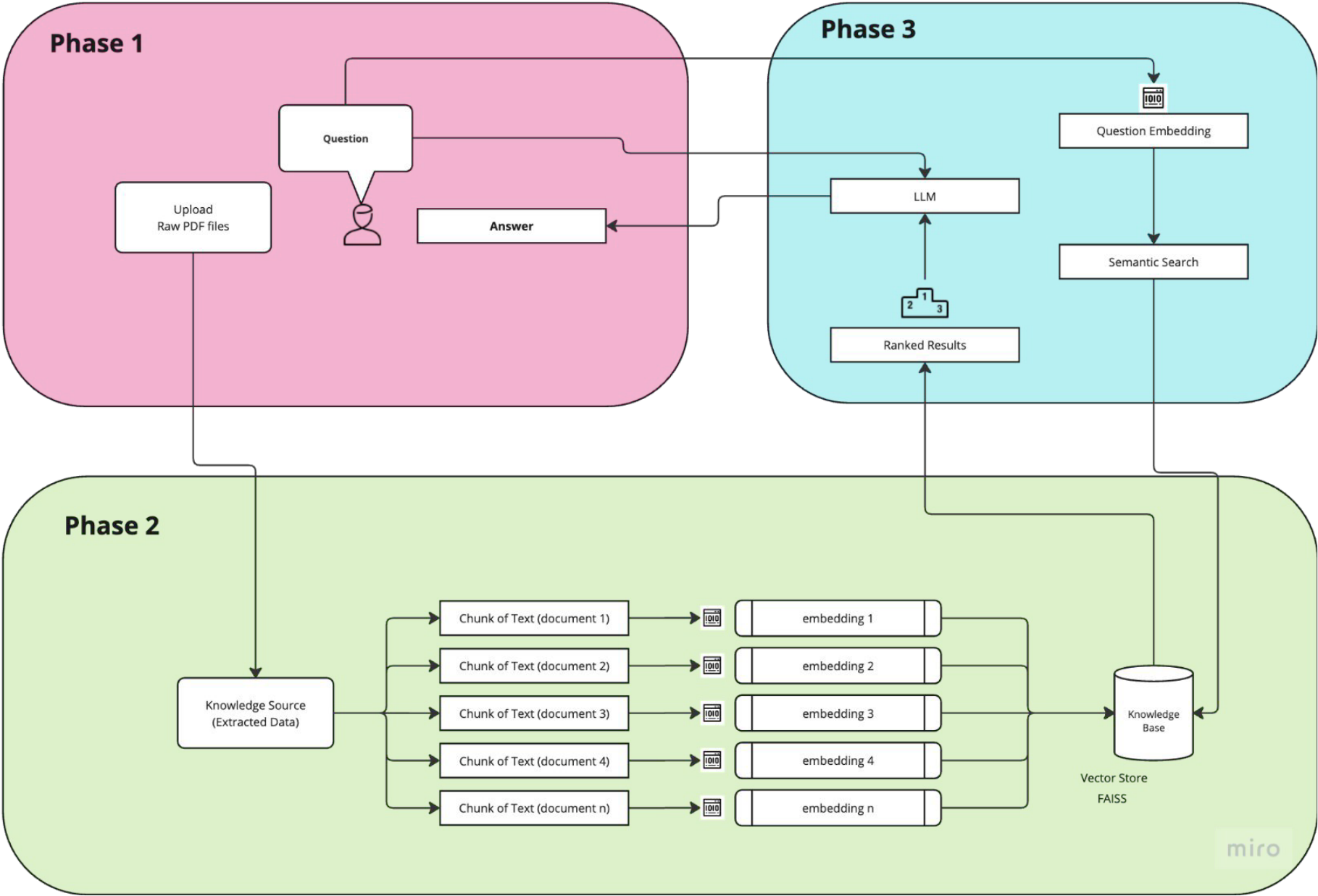
Phase3: RAG Pipeline (Connect Memory with LLM)

- Setup LLM (**Using DeepSeek R1 with Groq**)
- Retrieve Docs
- Answer Question

Phase4: Connect Everything Together

Technical Architecture

Streamlit



Tools and Technologies

- Deepseek R1 (Best Free CoT Reasoning Model)
 - With Ollama
 - With GROQ
- Ollama (Local LLM hosting/managing platform)
- Langchain (AI Framework for LLM applications)
- Streamlit (For Chatbot UI)
- FAISS (Vector Store)
- Pdfplumber and langchain loaders