AI RAG Chatbot – Architecture and Flow

Coding Round Task – AI-based Document QA System

Sree Gnyana Dhathri Paladi

GitHub: dhathri-paladi/ai-rag-chatbot: RAG-based document chatbot using Python, agents, and HuggingFace embeddings

Agent-Based Architecture with MCP Integration

```
[ User Input ]
      |
[ Streamlit UI ]
      |


[ Coordinator (main.py) ]
               |
┌─────┼──────────────┬──────────────┐

[IngestionAgent]    [RetrievalAgent]       [QueryAgent]
 (File Parsing)       (HF Embeddings)             |

                                              [LLMResponseAgent]
                                                (Mock LLM Answer)
```

System Flow and Message Passing

- User uploads document and asks a question via UI
- main.py coordinates all agent communication
- Ingestion Agent parses and chunks the file
- Retrieval Agent embeds it using Hugging Face
- Query Agent finds matching content
- LLMResponse Agent generates final answer
- All communication uses a message-passing structure with metadata like sender, receiver, trace ID

Tech Stack

Languages/Frameworks

•Python 3.10

•Streamlit (for UI)

•OOP Design with Modular Agents

Libraries

•sentence-transformers (HuggingFace)

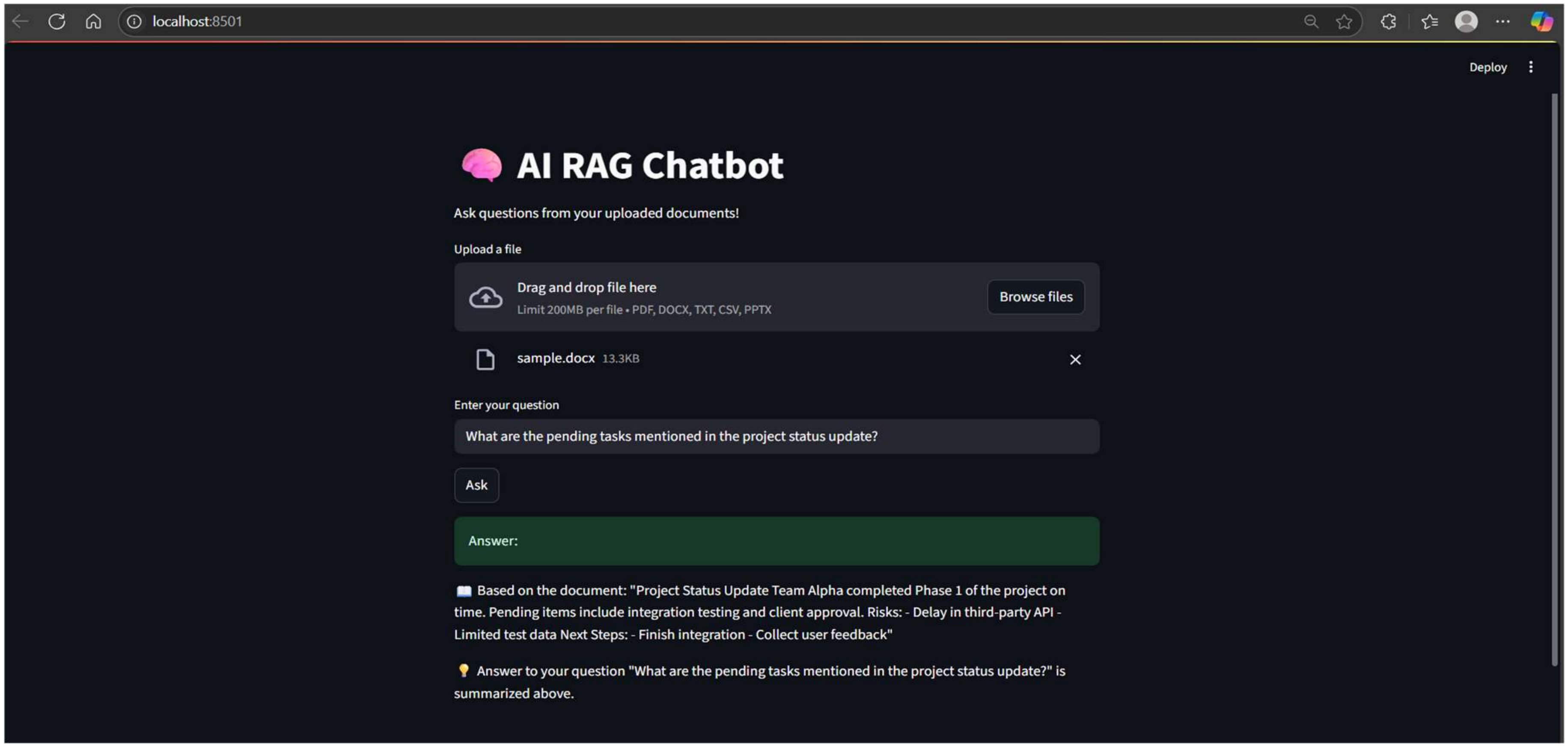
•PyMuPDF, python-docx, python-pptx, pandas, torch

Mock LLM:

•Simulates GPT-style output locally, no API required

Working Application – Screenshots

Input/Output Interaction"
User asks about pending tasks → Chatbot extracts from document and responds.



🧠 AI RAG Chatbot

Ask questions from your uploaded documents!

Upload a file

Drag and drop file here
Limit 200MB per file • PDF, DOCX, TXT, CSV, PPTX

sample.docx 13.3KB

Enter your question

What are the pending tasks mentioned in the project status update?

Answer:

📖 Based on the document: "Project Status Update Team Alpha completed Phase 1 of the project on time. Pending items include integration testing and client approval. Risks: - Delay in third-party API - Limited test data Next Steps: - Finish integration - Collect user feedback"

💡 Answer to your question "What are the pending tasks mentioned in the project status update?" is summarized above.

User → Upload File & Ask Question → [UI] → Get Answer ← Backend Agents

Challenges & Future Scope

- **Challenges:**
- Designing message-passing interface
- Handling multiple document formats
- Maintaining clean agent separation
- **Future Improvements:**
- Replace MockLLM with real OpenAI/GPT API
- Add PDF export of answers
- Deploy as a hosted web app (e.g., Streamlit Cloud)