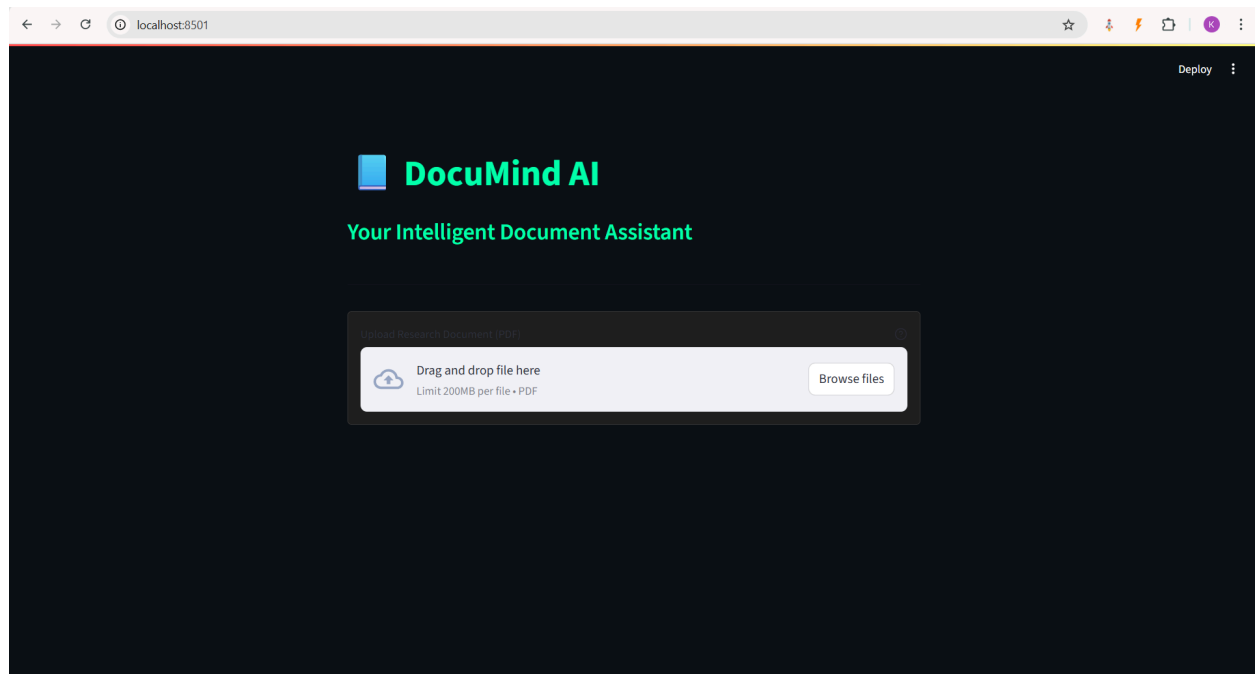


DocuMind AI – RAG-Based Document Assistant







Github repo: https://github.com/KhawajaAbdullah2000/llm_rag_deep_seek

1. Project Overview

DocuMind AI is an intelligent document assistant built using Retrieval-Augmented Generation (RAG) architecture. It enables users to upload PDFs, analyze content, and ask context-based questions powered by an LLM and vector search. The goal is to simplify information retrieval from large documents through natural language queries.



2. Key Features

-  Upload and analyze PDF documents
-  LLM-powered question answering (DeepSeek-R1)
-  Document chunking with recursive text splitting
-  Semantic similarity search using in-memory vector DB
-  Chat-like UI with styled user/assistant messages
-  Dark theme custom CSS for better UX

3. Tech Stack

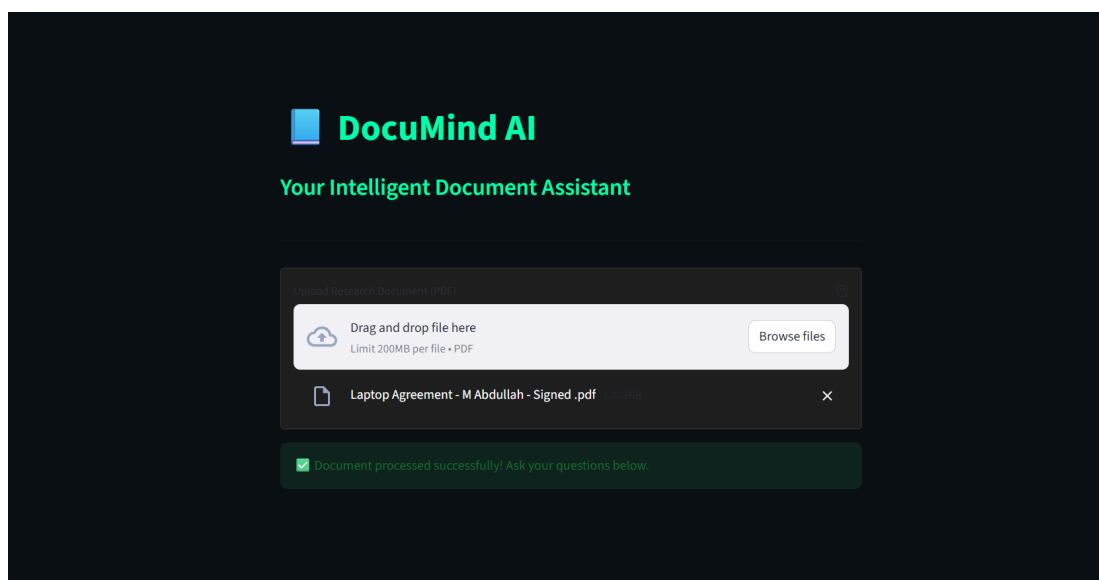
Component	Technology Used
Frontend UI	Streamlit
LLM	Ollama - DeepSeek-R1 (1.5B)
Embeddings	OllamaEmbeddings
Document Loader	LangChain PDFPlumberLoader
Text Splitting	RecursiveCharacterTextSplitter
Vector Store	InMemoryVectorStore (LangChain)
Prompting	ChatPromptTemplate (LangChain)
Styling	Custom HTML/CSS injected via <code>st.markdown()</code>

5. How It Works – Step by Step

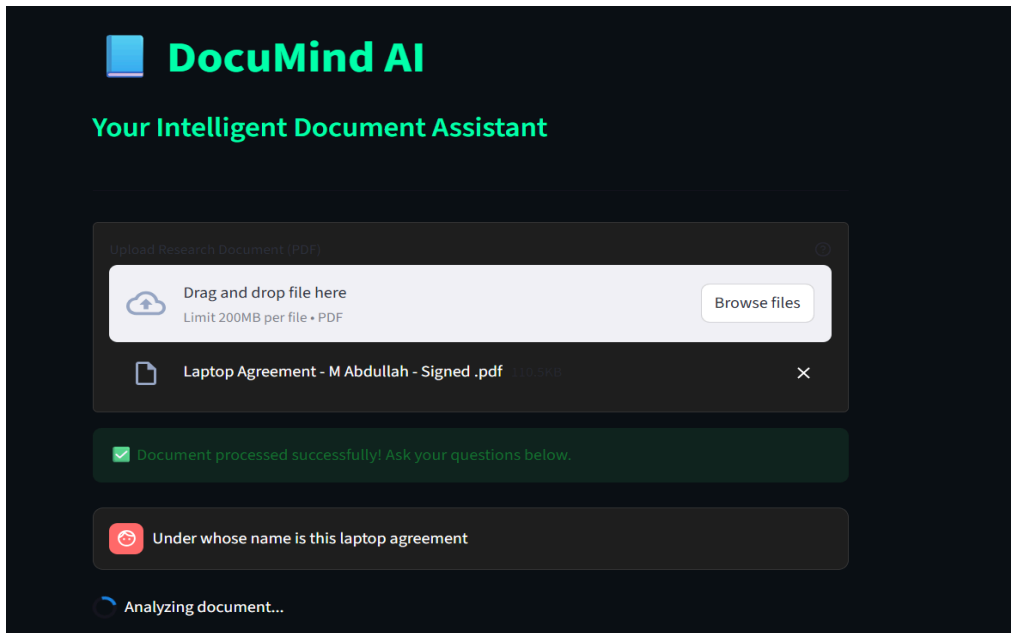
1. **PDF Upload:** Users upload a PDF, which is saved to a local storage directory.
2. **Document Parsing:** PDF is parsed using `PDFPlumberLoader`.
3. **Chunking:** Parsed text is split into overlapping chunks using `RecursiveCharacterTextSplitter`.
4. **Vector Embedding:** Chunks are converted into vector embeddings and stored in an `InMemoryVectorStore`.
5. **Query Input:** User enters a question via chat input.
6. **Context Retrieval:** Semantically similar chunks are retrieved using similarity search.
7. **Answer Generation:** Retrieved context is passed with the query into a structured prompt, processed by the Ollama LLM.
8. **Output:** The assistant returns a concise, factual answer within the chat UI.

WORKING DEMO

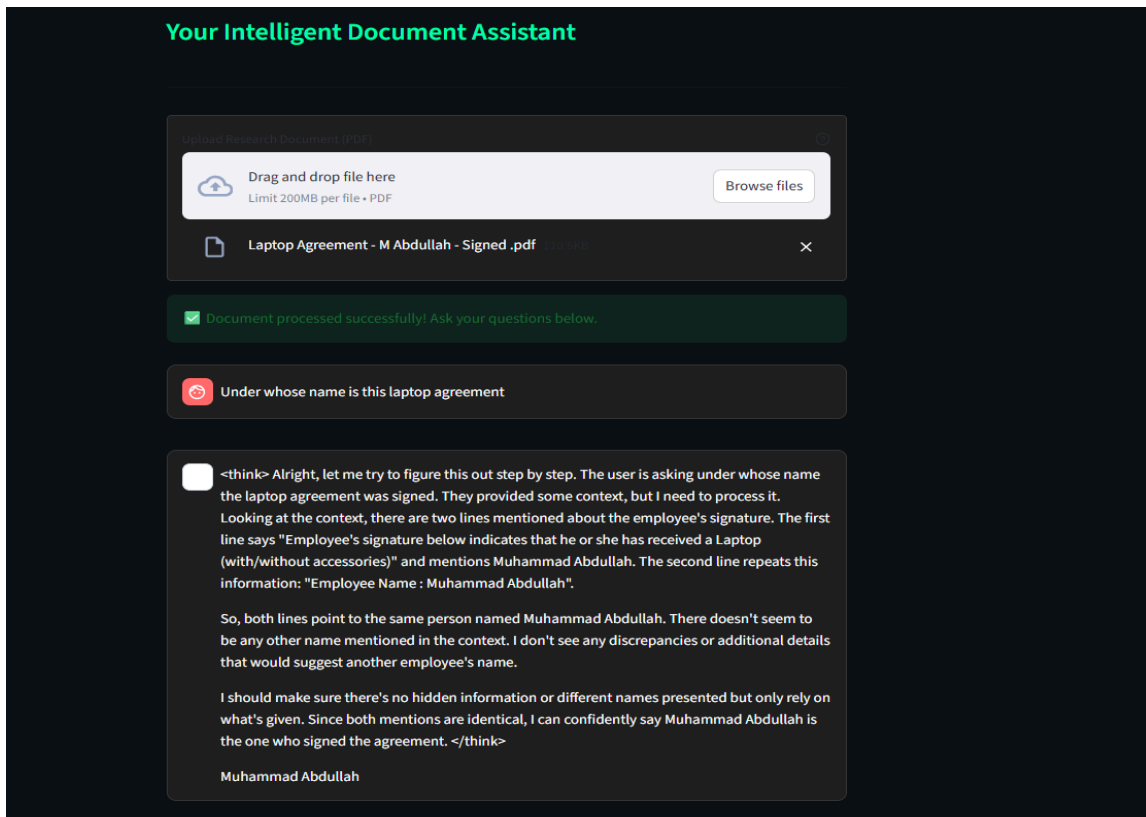
1- File Upload



2. Ask Question related to the uploaded file



Under whose name is this laptop agreement



Under whose name is this laptop agreement