

Project Plan-Version1

Week	Dates	Tasks
Week 1	Oct 19 – Oct 25, 2024	Backend Setup and PDF Upload
		- Set up Django backend for managing file uploads and extracted text storage.
		- Implement OCR using PyMuPDF for extracting text from uploaded PDFs.
		- Set up storage folders (uploads/, extracted_texts/) for raw PDFs and extracted text files.
Week 2	Oct 26 – Nov 1, 2024	Text Processing and Vectorization
		- Chunk extracted PDF text into semantically meaningful pieces.
		- Use OpenAI embeddings to convert text chunks into vector representations.
		- Set up FAISS or Pinecone to store vectors and facilitate fast similarity search.
Week 3	Nov 2 – Nov 5, 2024	Frontend Development with Next.js
		- Build a responsive frontend interface in Next.js to enable PDF upload and user interaction.
		- Add a question-answering interface where users can ask questions and view generated responses.
Week 4	Nov 6 – Nov 8, 2024	Integration of RetrievalQA Pipeline
		- Implement LangChain's RetrievalQA pipeline for question answering.
		- Connect RetrievalQA with FAISS or Pinecone for retrieving document chunks and generating answers.
		- Integrate the backend and frontend, ensuring smooth file upload and question-answering functionality.
Week 5	Nov 9 – Nov 12, 2024	Summarization and System Optimization
		- Implement dynamic chunking to maintain context during text processing.
		- Use BART to summarize extracted document text before passing it to the question-answering module.
		- Refine user experience with features like loading indicators and response optimization.
Week 6	Nov 13 – Nov 15, 2024	Testing, Evaluation, and Deployment
		- Conduct testing on all functionalities: document upload, question answering, and summarization.
		- Evaluate system performance metrics: response accuracy, response time, and user feedback.
		- Deploy the frontend on Vercel or Netlify and the backend on AWS or DigitalOcean.
		Document the system architecture and prepare a final project report.

Roles and Responsibilities

- **Backend Developer** (Oct 19 – Nov 8, 2024): Set up Django, handle OCR extraction, manage vector storage, and integrate LangChain RetrievalQA.
- **Frontend Developer** (Nov 2 – Nov 8, 2024): Develop Next.js frontend, handle user interactions, and connect with backend for data exchange.
- **Deployment Manager** (Nov 13 – Nov 15, 2024): Deploy and configure the website and databases on the cloud for scalable access.
- **Project Manager** (Oct 19 – Nov 15, 2024): This position oversees the timeline, coordinates task completion, and ensures project goals are met on schedule.