Project Plan-Version2

Note for Team Collaboration:

If any team member encounters challenges or is stuck on any task, they are encouraged to seek assistance from the team immediately. Any team member with relevant knowledge or expertise will provide support and guidance to help resolve the issue efficiently. Our goal is to work collaboratively and ensure that everyone progresses smoothly toward completing the project.

Week	Dates	Tasks
Week 1	Oct 19 – Oct 25, 2024	**Backend Setup and Initial OCR Development**
		- **Project Manager & Backend Developer**: Set up Django backend for file upload and text extraction.
		- 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**
		- **Project Manager & Backend Developer**: Refine OCR and chunk extracted PDF text into coherent sections.
		Backend Developer: Use OpenAI embeddings to vectorize text chunks and store them in FAISS or Pinecone.
		- Test OCR quality and adjust chunking to ensure content relevance.
Week 3	Nov 2 – Nov 5, 2024	**Frontend Development with Next.js**
		| - **Frontend Developer**: Build a Next.js interface for PDF upload and user interaction.
		- **Project Manager**: Support testing of the frontend OCR output display.

		- Add question-answering functionality where users submit questions and view results.
Week 4	**Nov 6 – Nov 8, 2024**	**Integration of RetrievalQA Pipeline**
		- **Project Manager & Backend Developer**: Implement LangChain's RetrievalQA pipeline for question answering.
		Backend Developer: Connect RetrievalQA with FAISS/Pinecone for chunk retrieval and answer generation
		- Integrate frontend and backend to support smooth interaction for uploads and answers.
Week 5	**Nov 9 – Nov 12, 2024**	**Summarization and System Optimization**
		- **Project Manager**: Lead implementation of BART summarization for document summarization.
		- **Project Manager & Backend Developer**: Implement dynamic chunking to improve context preservation.
		- Refine summarization quality and test integration with the question-answering pipeline for efficiency.
		- Add user experience improvements like loading indicators and optimized API calls.
Week 6	**Nov 13 – Nov 15, 2024**	**Testing, Evaluation, and Deployment**
		- **Project Manager**: Oversee final testing of all functionalities including document upload, QA, and summarization.
		- **Deployment Manager & Project Manager**: Deploy the frontend on Vercel or Netlify and backend on AWS/DigitalOcean.
		- Conduct performance evaluations (accuracy, response time, and user feedback).
		- Document the architecture, final workflows, and prepare the project report.

Roles and Responsibilities

- **Backend Developer (Oct 19 – Nov 8, 2024):** Set up Django backend, implement OCR, vector storage, LangChain RetrievalQA, and coordinate with the project manager for OCR and vectorization.

- **Frontend Developer (Nov 2 – Nov 8, 2024):** Develop the Next.js frontend, handle user interactions, and integrate with the backend.

- **Project Manager (Oct 19 – Nov 15, 2024):** Oversee project timeline, assist in OCR implementation, lead BART summarization, support frontend testing, and ensure quality completion of project milestones.

- **Deployment Manager (Nov 13 – Nov 15, 2024):** Deploy and configure the application across cloud services and verify accessibility.