AI PDF Analyzer Project Roadmap

This roadmap outlines the development process for building the AI PDF Analyzer project, including backend, frontend, integration, and scaling phases.

# Phase 1: Backend Foundation (Python, FastAPI)

* Set up project structure & virtual environment
* Build `/health` endpoint
* Implement `/upload-pdf` endpoint (store + parse PDFs)
* Add OCR fallback (for scanned PDFs)
* Implement text chunking
* Add embeddings + vector DB (ChromaDB local)
* Build `/ask` endpoint (retrieval only, no AI yet)
* Connect DeepSeek API for summarization & Q&A
* Test backend thoroughly with sample PDFs

# Phase 2: Frontend Development (React + Tailwind)

* Design minimal UI (file upload + text input)
* Connect frontend to backend endpoints
* Display JSON results (raw)
* Improve UI with styled components (cards, chat-like results)
* Add drag-and-drop upload, progress bar
* Add summary download feature

# Phase 3: Integration & Production

* Add user authentication
* Namespace vector DB by user
* Add 'delete document' option
* Deploy backend (Render/Railway/AWS) & frontend (Vercel/Netlify)
* Buy domain + set up HTTPS
* Test API performance with multiple files
* Monitor usage & costs with DeepSeek API

# Phase 4: Scaling & Advanced Features

* Support multiple files per user
* Long-document summarization (auto-split + merge)
* Fine-tune prompt templates
* Add billing/subscription model for users
* Deploy database & vector DB to cloud (e.g., PostgreSQL + Pinecone)