RAG Knowledge retrieval engine

Problem statement:

The RAG engine will ingest document and retrieve information from the document based on user query and create structured answers with citations

Project success criteria:

1. Ingest .pdf files
2. Accept user query and retrieve relevant information
3. Create a structured response with citations

Backend

- FastAPI

- PostgreSQL + pgvector

- SQLAlchemy

- Pydantic

AI

- OpenAI / Anthropic (abstract behind interface)

- Text-embedding model

Infra (local)

- Docker + docker-compose