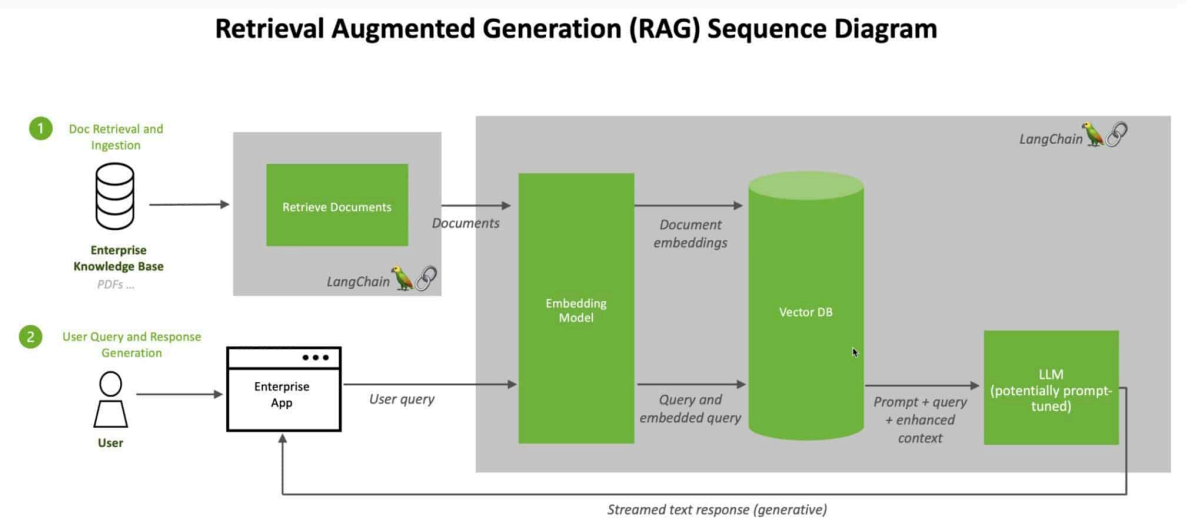
AI Support ChatBot for Sports to create session plans for sports coaches

Backend Tech:

- In Backend Architecture we will be using **Retrieval Augmented Generation** based on https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/



- Openai's GPT-4omni Transformer due its robust results and reasoning capabilities.
- Openai's Embedding Model to convert our data such as (Pdfs, Docs, Txt files, CSV or excel files) into vectors, because LLMs work with vectors/embeddings to generate related output.
- Pinecone Vector database, as it is one of the fastest, efficient and secure vector databases nowadays used in GenAI applications. All our data will be stored there. In comparison we can also use ChromaDB, Quadrant and faiss vector dbs but their efficiency is not good as compared to pinecone. https://www.pinecone.io/
- Prompt Engineering to direct the behaviour of LLM.
- All these Components will be integrated using Python programming language and langchain framework (nowadays most used framework in GenAI application development).
- Finally all endpoints will be created using fastapi to integrate with the UI.

Frontend Techs:

- Frontend (UI components) will be based on:
    - React js
    - React Router dom
    - React Icons

Database:

- Postgresql

Requirements:

- We will be needing two things from you to start.
    1. Openai API key to train and integrate their LLMon our data.
    2. Pinecone Vectordb subscribed account.