AI Concepts: RAG, Embeddings, LangChain, Vector Databases

1. What is an LLM? Large Language Models (LLMs) are deep learning models trained on massive text datasets to understand and generate human-like language. They can answer questions, summarize text, retrieve information, and perform reasoning tasks.

2. What are Embeddings? Embeddings convert text into numerical vectors. These vectors capture semantic meaning—texts with similar meaning have similar vectors. Embeddings are the foundation for search, similarity, clustering, and Retrieval-Augmented Generation (RAG).

3. What is a Vector Database? A vector database (like Qdrant, Pinecone, or FAISS) stores embeddings and enables fast similarity search. Instead of keyword search, vector DBs perform semantic search using cosine similarity or distance metrics.

4. What is RAG (Retrieval-Augmented Generation)? RAG is an architecture where: 1. A user asks a question. 2. The system retrieves relevant text chunks from a database using embeddings. 3. An LLM uses this context to generate accurate, grounded answers.

This prevents hallucinations and lets the model use external knowledge.

5. How LangChain Helps LangChain provides: - Document loaders - Text splitters - Embedding generation - Vectorstore integrations - Retrieval chains - Agent systems (with tools and routing)

It simplifies building LLM applications.

6. How LangGraph Helps LangGraph is used to build agent workflows as graphs: - Nodes represent functions or models - Edges represent routing logic - Supports conditional routing, loops, memory

Perfect for agent pipelines like your Weather + PDF RAG system.

7. How LangSmith Helps LangSmith is an LLM observability platform: - Tracks runs - Debugs prompts - Evaluates LLM responses - Visualizes agent graphs - Helps in productionizing LLM apps

8. Summary This PDF contains the foundational concepts needed for your assignment: - LLMs generate responses - Embeddings represent text meaning - Vector DB stores knowledge - RAG retrieves and summarizes PDF content - LangChain + LangGraph build the agent pipeline - LangSmith evaluates and monitors it