



SUPERIOR UNIVERSITY

NAME:	M.JUNAID
ROLL NO:	BSAI-109
SECTION:	4B
SUBJECT:	Programming for Artificial Intelligence
Lab task:	11
SUBMITTED TO:	Sir Rasikh

LangChain (Language Chain)

Definition:

LangChain is a framework designed to help developers build applications powered by **Large Language Models (LLMs)** by chaining together components like prompts, memory, tools, and databases.

Key Features:

- Helps connect LLMs with external data sources (e.g., APIs, databases).
- Supports **retrieval-augmented generation (RAG)** workflows.
- Modular: allows combining chains like prompt templates, LLM calls, and document retrievers.

Use Case:

Chatbots that pull info from private documents, customer support tools, or any AI app that combines reasoning + retrieval.

RAG (Retrieval-Augmented Generation)

Definition:

RAG is an architecture where an LLM generates responses **based on external information retrieved in real-time**, rather than relying purely on its training data.

How It Works:

1. **Query is embedded into a vector.**
2. The system retrieves **relevant documents** (from a VectorDB).
3. These documents are passed into the LLM as context.
4. The LLM generates a **more accurate and informed answer**.

Why It's Useful:

LLMs like GPT don't have access to real-time data or private knowledge. RAG helps overcome this limitation by retrieving up-to-date facts from external sources.

LLMs (Large Language Models)

Definition:

LLMs are deep learning models trained on vast amounts of text to understand and generate human-like language.

Examples:

- GPT-4 (OpenAI)
- PaLM (Google)
- LLaMA (Meta)

Capabilities:

- Text generation
- Translation
- Summarization
- Q&A
- Code generation

Limitation:

LLMs can **hallucinate** (make things up), have a **knowledge cutoff**, and can't access **real-time information** unless integrated with RAG or tools like LangChain.

FAISS (Facebook AI Similarity Search)

Definition:

FAISS is an open-source library developed by Facebook for **fast similarity search** and **clustering of dense vectors**.

Use Case in AI:

- Used in RAG systems to **retrieve similar documents** based on vector embeddings.
- It powers **Vector Databases** by providing fast search over high-dimensional vectors.

Vector (Embedding)

Definition:

A vector in AI is a **numeric representation of data**, like text or images, that captures **semantic meaning** in a fixed-size format.

Example:

- The sentence "I love cats" might be converted to [0.3, 0.8, -0.1, ...]

Why It's Important:

- Enables similarity search (similar texts have similar vectors).
- Used in LLM pipelines for comparing, clustering, or retrieving information.

VectorDB (Vector Database)

Definition:

A VectorDB is a specialized database designed to **store and search vectors** efficiently using similarity metrics (like cosine similarity).

Examples:

- Pinecone
- Weaviate
- Chroma
- Milvus

Role in RAG:

- Stores vector embeddings of documents.
- Given a query vector, it retrieves the most semantically similar documents for context-aware generation.

Generative AI

Definition:

Generative AI refers to models that can create **new content** such as text, images, music, code, etc., based on training data.

Types:

- Text: LLMs (e.g., GPT)
- Images: GANs, diffusion models (e.g., DALL-E, Stable Diffusion)
- Music: Jukebox
- Code: Codex

How It Works:

- Trained on massive datasets to learn patterns.
- Can **generate outputs** that were never explicitly seen in training data.

GANs (Generative Adversarial Networks)

Definition:

GANs are a type of generative AI model consisting of two neural networks:

- **Generator:** Creates fake data (e.g., fake images).
- **Discriminator:** Tries to distinguish between real and fake data.

How They Work:

- Generator improves by fooling the discriminator.
- Discriminator improves by catching fakes.

Used For:

- Creating photorealistic images
- Deepfakes
- Art generation
- Image super-resolution