The Superior University

Name: Fatima Nadeem	Roll No:076	Course: PAI Lab
Semester:4th	Section:4B	Department:SE
Submitted To: Sir Rasikh	Total Marks: 10	Date:

Lab Task 11

LangChain:

LangChain is a Python framework designed to help developers build applications using **Large Language Models (LLMs)** more effectively. It allows chaining together LLM prompts, memory, retrieval systems, and external tools like APIs or databases. LangChain is particularly useful for building **chatbots**, **agents**, and **RAG (Retrieval-Augmented Generation)** systems.

RAG (Retrieval-Augmented Generation):

RAG is a technique that combines **LLMs** with external knowledge sources (like documents) to improve the accuracy and relevance of responses. Instead of relying only on pre-trained model knowledge, RAG retrieves relevant information from a database and includes it in

the prompt to the LLM, enhancing performance on domain-specific or factual queries.

LLMs (Large Language Models):

LLMs are advanced AI models trained on massive text data to understand and generate human-like language. Examples include **GPT**, **BERT**, and **Claude**. They can perform a wide range of tasks, including summarization, translation, question answering, and content generation, but may not always have up-to-date or factual knowledge without augmentation.

FAISS (Facebook AI Similarity Search):

FAISS is a library developed by Facebook AI that enables **efficient similarity search** and clustering of dense vectors. It's widely used in **vector databases** to retrieve similar data (e.g., documents or images) quickly based on vector similarity, which is crucial in RAG pipelines for retrieving relevant documents.

Vector:

A **vector** in AI represents data (like text, images, or audio) as a list of numbers. This numerical form allows machines to compute similarity between items. For example, a sentence can be converted into a vector using embeddings so that similar sentences have similar vector representations.

VectorDB (Vector Database):

A VectorDB stores vectors and allows fast similarity search using tools like FAISS, Pinecone, or Weaviate. It is essential in

retrieval-based systems like RAG, where finding the most relevant documents to a user's query is based on vector similarity rather than keyword matching.

Generative AI:

Generative AI refers to any AI system that can create new content—text, images, audio, code, etc. It includes models like **GPT for text** or **Stable Diffusion for images**. These models learn patterns from training data and use them to generate new, original outputs based on user prompts.

GANs (Generative Adversarial Networks):

GANs are a specific type of generative AI composed of two neural networks—a **generator** and a **discriminator**—competing against each other. The generator creates fake data, and the discriminator tries to detect if it's real or fake. Through this process, GANs become excellent at generating realistic images, videos, and other types of data.