**The Superior University, Lahore**

|  |  |  |  |
| --- | --- | --- | --- |
| Course Title: | PAI-LAB | | |
| Programme Name: | BSDS | | |
| Semester: | 4rd | Section: | BSDSM-4A |
| Student’s Name: | Muhammad Faizan Majid | | |
| Instructor Name: | Prof. Rasikh | | |

**Lab 11: Differences Between Key AI Concepts**

**1. LangChain**

LangChain is a framework designed to simplify the development of applications powered by large language models (LLMs). It provides tools and abstractions for integrating LLMs with external data sources, memory for context retention, and chains for orchestrating complex workflows. LangChain enables developers to build applications like chatbots or question-answering systems by combining LLMs with tools like retrieval systems or APIs.

**2. RAG (Retrieval-Augmented Generation)**

RAG is a hybrid approach that combines retrieval-based methods with generative models. It retrieves relevant documents or data from a knowledge base (using techniques like vector search) and feeds them into a generative model (e.g., an LLM) to produce contextually informed responses. RAG enhances LLMs by grounding their outputs in external knowledge, improving accuracy and relevance, especially for domain-specific queries.

**3. LLMs (Large Language Models)**

LLMs are deep learning models trained on vast text datasets to understand and generate human-like language. Examples include GPT, BERT, and LLaMA. They excel at tasks like text generation, translation, and summarization but rely on their training data and may lack real-time or specialized knowledge unless augmented with techniques like RAG or fine-tuning.

**4. FAISS (Facebook AI Similarity Search)**

FAISS is an open-source library developed by Facebook AI for efficient similarity search and clustering of dense vectors. It is used to perform fast nearest-neighbor searches in high-dimensional spaces, making it ideal for applications like semantic search or recommendation systems. FAISS is often used in RAG pipelines to retrieve relevant documents based on vector embeddings.

**5. Vector**

In the context of AI, a vector is a numerical representation of data (e.g., text, images) in a high-dimensional space. Vectors capture semantic or contextual meaning, enabling mathematical operations like similarity comparisons. For example, word embeddings (e.g., Word2Vec) or sentence embeddings (e.g., from BERT) represent text as vectors for tasks like search or clustering.

**6. VectorDB (Vector Database)**

A Vector Database is a specialized database designed to store, index, and query high-dimensional vectors efficiently. It supports similarity searches (e.g., cosine similarity) to find vectors closest to a query vector. Examples include Pinecone, Weaviate, and Milvus. VectorDBs are critical for applications like semantic search, RAG, and recommendation systems.

**7. Generative AI**

Generative AI refers to algorithms that create new content, such as text, images, audio, or video, based on patterns learned from training data. It includes models like LLMs, GANs, and diffusion models. Generative AI powers applications like content creation, chatbots, and art generation, leveraging probabilistic methods to produce novel outputs.

**8. GANs (Generative Adversarial Networks)**

GANs are a type of generative AI consisting of two neural networks: a generator and a discriminator. The generator creates synthetic data (e.g., images), while the discriminator evaluates whether the data is real or fake. Through adversarial training, GANs produce highly realistic outputs, commonly used in image generation, style transfer, and data augmentation.