

**The Superior University**

|  |  |  |
| --- | --- | --- |
| Name: Abdul Moeez | Roll No: 017 | Course: OS Lab |
| Semester: 4th | Section: BSAI 4A | Department: Artificial Intelligence |
| Submitted To: Sir Rasikh | Total Marks: | Date: |

**Key Differences Between Common AI Technologies**

### LangChain

### LangChain is a framework designed to help developers build applications that use large language models (LLMs). It provides tools for managing prompts, chaining together multiple LLM calls, integrating with external data sources (like APIs or databases), and building custom workflows. LangChain is particularly useful for creating conversational agents, question-answering systems, and LLM-powered tools that interact with external environments.

### **RAG (Retrieval-Augmented Generation)**

### RAG is a technique that combines information retrieval with text generation. It enhances LLM outputs by retrieving relevant documents from a knowledge base (often a vector database) and using those documents to inform the model's response. This approach helps mitigate hallucinations (false or made-up content) in LLM outputs by grounding responses in factual, retrieved data.

### LLMs (Large Language Models)

### LLMs are deep learning models trained on massive text corpora to understand and generate human-like language. Examples include GPT, BERT, and PaLM. These models can perform a range of language tasks, such as translation, summarization, question answering, and more. They form the core of many generative AI systems.

### 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's widely used to search for similar embeddings in high-dimensional spaces, which is essential for implementing vector databases and supporting RAG pipelines.

### Vector

### In the context of AI and machine learning, a vector is a numerical representation of data. For language, vectors (or embeddings) are used to represent words, sentences, or documents in a multi-dimensional space, allowing similarity computations. Vectors capture semantic meanings and relationships between data points.

### VectorDB (Vector Database)

### A VectorDB is a specialized database designed to store and query vector embeddings. It supports fast similarity searches using algorithms like approximate nearest neighbors (ANN). Examples include Pinecone, Weaviate, and FAISS-based systems. VectorDBs are critical for enabling efficient retrieval in RAG setups.

### **Generative AI**

### Generative AI refers to systems capable of generating new content (text, images, audio, etc.) based on learned patterns. It includes models like LLMs, image generators, and music generators. These models are trained to create data rather than just classify or interpret it.

### GANs (Generative Adversarial Networks)

### GANs are a specific type of generative model consisting of two neural networks: a generator and a discriminator. The generator creates new data instances, while the discriminator evaluates them. Through this adversarial training process, GANs learn to produce highly realistic data, especially images and videos.