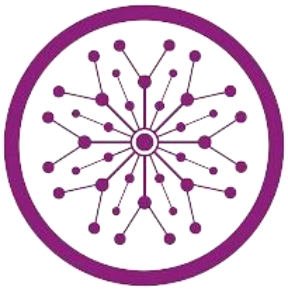
Programming for AI

Lab

Task 11



The Superior University

Faculty of Computer Science & Information Technology

2023-2027

Submitted To:

Sir Rasikh Ali

Submitted By:

Qaswa Chaudhary

Roll No:

SU92-BSDSM-F23-002

Section:

4A

Department:

Software Engineering

1. LangChain:

Definition: LangChain is an open-source framework designed to simplify the development of applications powered by Large Language Models (LLMs). It helps developers' chain together multiple components such as prompts, LLMs, memory, APIs, tools, and databases to build complex, context-aware AI applications.

Explanation: LangChain acts as a modular system where each part (like querying a database, calling an API, or storing conversation history) can be linked together in a chain. This allows developers to create AI systems that go beyond just responding to prompts—they can remember past interactions, retrieve data from external sources, and make more intelligent decisions.

Example Use Case: Building a customer support chatbot that fetches answers from a knowledge base using LLMs.

2. RAG (Retrieval Augmented Generation):

Definition: RAG is an AI architecture that enhances language models by incorporating external knowledge sources during text generation. It combines a retrieval system with a text generation model, allowing the LLM to generate more accurate and up-to-date responses.

Explanation: Instead of relying solely on the pre-trained knowledge of an LLM, RAG fetches relevant documents or data from a database (typically a VectorDB) based on the user's query. The language model then uses this retrieved information to craft a response, making it better suited for tasks like question-answering, summarization, and documentation.

Example Use Case: A chatbot that can answer company-specific questions by retrieving internal documents in real-time.

1. LLMs (Large Language Models):

Definition: LLMs are deep learning models trained on massive amounts of textual data to understand, interpret, and generate human language. These models use architectures like Transformers and have billions of parameters.

Explanation: LLMs work by predicting the next word in a sequence of words, making them capable of generating fluent text, answering questions, summarizing content, translating languages, and more. They are foundational to tools like ChatGPT, Google Bard, and others.

Examples: GPT-4, GPT-3.5, BERT, RoBERTa, LLaMA.

2. FAISS (Facebook AI Similarity Search):

Definition: FAISS is an open-source library developed by Facebook AI Research for efficient similarity search and clustering of dense vectors.

Explanation: FAISS is designed to handle large collections of vectors—numerical representations of text, images, etc.—and quickly find which vectors are most similar to a given one. It is especially useful in applications like image search, recommendation systems, and document retrieval in RAG.

Example Use Case: Searching for similar customer queries in a support database.

3. Vector:

Definition: In the context of AI and machine learning, a vector is a numerical representation of data (like words, sentences, or images) in a high-dimensional space.

Explanation: Vectors allow machines to process and compare data using mathematical operations. For instance, similar meanings of words will have vectors that are close together in vector space. These vectors are typically generated using techniques like Word2Vec, BERT embeddings, or CLIP for images.

Example: The word “king” might be close to “queen” in vector space, showing semantic similarity.

4. VectorDB (Vector Database):

Definition: A Vector Database is a specialized database system designed to store and retrieve high-dimensional vectors efficiently.

Explanation: VectorDBs support similarity searches, enabling applications to find the most relevant vectors to a query vector using metrics like cosine similarity or Euclidean distance. These databases are optimized for fast indexing and searching in large datasets.

Examples: Pinecone, Weaviate, Milvus, Chroma.

Use in RAG: Used to store document embeddings that can be retrieved based on user queries.

5. Generative AI:

Definition: Generative AI refers to a class of artificial intelligence that can create new content, such as text, images, music, or code, rather than simply analyzing existing data.

Explanation: These models learn the patterns and structure of training data and generate novel outputs that resemble the input data. They are widely used in creative applications, automation, and even scientific research.

Examples:

- ChatGPT (text)
- DALL·E and Midjourney (images)
- GitHub Copilot (code)

6. GANs (Generative Adversarial Networks):

Definition: GANs are a type of generative model made up of two neural networks—a Generator and a Discriminator—that compete with each other in a game-theoretic setup.

Explanation: The generator creates fake data, while the discriminator evaluates whether the data is real or fake. Over time, both networks improve, resulting in the generator producing extremely realistic data. GANs are famous for their ability to create synthetic images, deepfakes, and art.

Example Use Case: Generating lifelike human faces that do not actually exist.