**Lab Task 11**

**Questions:**

**1. LangChain (Language Chain)**

LangChain is an open-source platform designed to help developers build applications on top of Large Language Models (LLMs) like ChatGPT. It allows the models to connect with external sources, read memory, and execute complex workflows.

**Key Features:**

* Links LLMs with real-world tools like web search, APIs, databases, or PDFs.
* Supports multi-step conversations (like recalling previous inputs).
* Helps in developing applications like AI assistants, chatbots, and Q&A bots.

**Example:**

You're building a hospital chatbot. With LangChain, your chatbot can:

* Read from a medicine database,
* Answer questions based on the latest information,
* And remember last questions asked from the same user.

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

RAG is an intelligent way to make language models give improved and more accurate answers.

* Retrieval: It first searches for the most applicable documents related to the user's query.
* Generation: It then uses those documents to generate a better-informed and accurate answer.

**Why It's Useful:**

ChatGPT and other LLMs have a flaw—they can respond only based on what they've been trained. RAG enables them to "look up" new or special information from a bespoke document or database before responding.

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

Large Language Models are sophisticated artificial intelligence models trained on enormous quantities of text data from the internet, books, articles, and so forth. They are engineered to comprehend, process, and produce language in human-like fashion.

**Examples:**

* GPT-3 / GPT-4 (from OpenAI)
* BERT (from Google)
* LLaMA (from Meta)

**What They Can Do:**

* Respond to questions
* Translate between languages
* Generate stories or email
* Generate code
* Summarize text
* And so much more!

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

FAISS is a library developed by Facebook AI Research to assist computers in scanning large volumes of vector data very efficiently. When text is encoded into vectors (numbers), FAISS assists you in:

* Storing them with efficiency,
* And retrieving the most similar ones rapidly.

**Example:**

Suppose you've stored a database of 1 million paragraphs as vectors, and you ask a question, FAISS can quickly tell you the top 5 paragraphs nearest to your question.

**Used In:**

* RAG systems
* Recommender systems
* Memory-based chatbots

**5. Vector**

A vector is a list of numbers that encode something such as a word, sentence, image, or document in a format that makes sense to AI. In AI, we don't contrast the words "cat" and "dog" directly. Rather, we convert them into vectors that define their meaning.

**Example:**

* "Cat" → [0.23, 0.67, 0.15, -0.03]
* "Dog" → [0.25, 0.64, 0.14, -0.01]

Since their vectors are close, the AI knows these two are similar animals.

**6. What is VectorDB?**

A Vector Database is a specialized database for storing, searching, and retrieving vectors.

It does not store numbers or plain text but stores vector embeddings (numerical representations).

**Popular VectorDBs:**

* Pinecone
* FAISS
* Weaviate
* Chroma
* Milvus

**Why Use VectorDB?**

Legacy databases search for exact matches. But in AI, we need to search by meaning. VectorDBs help by searching similar content based on vectors.

**Example:**

You ask a question. Your system:

1. Converts your question into a vector,

2. Search the VectorDB,

3. Retrieves semantically similar documents,

4. Passes them to the LLM for a response (that's RAG!).

**7. Generative AI**

Generative AI is a type of artificial intelligence that can generate new content based on what it learned from data. This content may be:

* Text (ChatGPT)
* Multimedia content such as images (DALL·E, MidJourney)
* Music (Google MusicLM)
* Code (GitHub Copilot)

**How It Works:**

It uses models trained on large datasets to learn patterns and then generate new, original content that is similar to the training data.

**Example:**

You give it a prompt: "Draw a dog playing guitar." A Generative AI like DALL·E can create an original picture of a dog playing a guitar, even though it has never seen one such before.

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

GANs are a specific type of Generative AI model.

They are two neural networks:

* Generator: Tries to create forged content.
* Discriminator: Tries to determine if the content is forged or not.

They work in an analogous fashion to a game:

* The Generator improves at generating counterfeit data,
* The Discriminator improves at identifying fakes.

Finally, the Generator becomes so proficient that it produces real-looking content such as faces,paintings, or even counterfeit videos (deepfakes).

**Example:**

You may have noticed fake celebrity faces or AI-generated artworks — a lot of these are created with GANs.

**Used In:**

* Face generation
* AI art
* Fashion design
* Deepfake videos