Lang Chain:

A framework for building applications using large language models (LLMs).

Use: Helps developers create LLM-based apps with memory, chains, tools (like web search), and agents.

Use Case: A chatbot that queries documents, remembers past interactions, and integrates APIs.

RAG (Retrieval-Augmented Generation)

A technique that combines retrieval of relevant data with generation from LLMs.

How it works: It first retrieves relevant documents (e.g., from a database) and feeds them into an LLM to generate more informed responses.

Example Use Case: Chatbots that give accurate answers based on company documents or knowledge bases.

LLMs (Large Language Models)

AI models trained on massive text datasets to understand and generate human language.

Examples: GPT (like ChatGPT), BERT, Claude

Use: Text completion, summarization, Q&A, translation, code generation

FAISS (Facebook AI Similarity Search)

A library developed by Meta for efficient similarity search on vectors.

Use: Quickly find similar documents, images, or data points by comparing vector embeddings.

Example Use Case: Finding the top 5 most relevant documents to a user query.

Vector

What it is: A numeric representation of data (like text, images, etc.) in a high-dimensional space.

Use: Enables comparison of meaning or similarity between items.

Example: A sentence like "How are you?" might be represented as [0.1, -0.2, 0.3, …].

VectorDB (Vector Database)

A database designed to store and search vectors efficiently.

Examples: Pinecone, Weaviate, Milvus, Qdrant.

Use: Supports RAG systems by enabling fast vector search to retrieve relevant context.

Generative AI

What it is: A type of AI that can create new content (text, images, music, etc.) from learned patterns.

Examples: ChatGPT (text), DALL·E (images), Jukebox (music).

Use: Content creation, code generation, data augmentation

GANs (Generative Adversarial Networks)

A machine learning framework with two neural networks: a generator and a discriminator.

How it works: The generator creates fake data, and the discriminator tries to detect fakes. They improve by competing.

Use: Image synthesis, deepfakes, art generation.