Artificial Intelligence (AI) is a rapidly growing field of computer science that focuses on creating intelligent machines capable of performing tasks that normally require human intelligence. These tasks include learning from experience, understanding natural language, recognizing patterns, making decisions, and solving complex problems. AI has become an essential part of modern technology and is widely used in industries such as healthcare, finance, education, transportation, and entertainment.

One of the most important branches of Artificial Intelligence is Machine Learning. Machine Learning allows systems to automatically learn and improve from data without being explicitly programmed. Instead of following predefined rules, machine learning models identify patterns in data and use those patterns to make predictions or decisions. Common applications of machine learning include spam detection, recommendation systems, fraud detection, and image recognition.

Deep Learning is a subset of Machine Learning that uses neural networks with multiple layers to process complex data. Deep learning models are inspired by the structure of the human brain and are particularly effective in tasks such as image classification, speech recognition, and natural language understanding. These models require large datasets and significant computational power but provide highly accurate results.

Natural Language Processing (NLP) is another key area of Artificial Intelligence that focuses on enabling machines to understand and generate human language. NLP techniques are used in chatbots, virtual assistants, language translation systems, and sentiment analysis tools. Modern NLP systems rely on transformer-based architectures and large language models to process text efficiently and accurately.

Large Language Models (LLMs) are advanced AI systems trained on massive amounts of textual data. These models can generate human-like responses, summarize documents, answer questions, and assist users in various tasks. LLMs are widely used in customer support, content generation, education, and research applications.

Vector databases play a crucial role in AI-powered applications that rely on semantic understanding. Instead of storing data in traditional tables, vector databases store numerical representations called embeddings. These embeddings capture the semantic meaning of text, images, or other data types. Similar data points have vectors that are close to each other in vector space.

FAISS (Facebook AI Similarity Search) is a popular library used for efficient similarity search and clustering of dense vectors. It is designed to handle large-scale datasets and perform fast nearest-neighbor searches. FAISS is commonly used in semantic search engines, recommendation systems, and retrieval-augmented generation pipelines.

Retrieval-Augmented Generation (RAG) is a powerful technique that combines information retrieval with language generation. In a RAG system, relevant documents are first retrieved from a vector database based on the user query. These documents are then passed as context to a language model, which generates a response grounded in the retrieved information. This approach improves accuracy and reduces hallucinations.

LangChain is a framework that simplifies the development of applications powered by large language models. It provides tools for document loading, text splitting, embedding generation, vector storage, prompt templates, and chaining multiple components together. LangChain is widely used to build intelligent chatbots that can answer questions based on external documents such as PDFs.

Text splitting is an important step in document processing. Large documents are divided into smaller chunks to ensure efficient embedding and retrieval. Chunk size and overlap play a key role in maintaining context while improving search accuracy. Proper text splitting ensures better performance in vector-based search systems.

In conclusion, Artificial Intelligence, Machine Learning, Natural Language Processing, and vector databases form the foundation of modern AI systems. Tools like FAISS and LangChain enable developers to build powerful and scalable applications. By combining document retrieval with language models, developers can create intelligent systems that provide accurate, contextual, and meaningful responses to users.