

BE AMs training

- What is RAG?

RAG (Retrieval Augmented Generation)

(We can give chatbot the access to our data)

RAG is a technique that enables Gen AI to retrieve and incorporate new information. In simple language, it means it can take the data from the personal documents that we provide, modify or augment it and then generate information specific to the user's query based on the new information and pretrained knowledge.

- Why is RAG needed in modern AI applications?

RAG is needed because it helps to minimize the contradictions and inconsistencies in the generated text by allowing us to prompt-engineer the LLM or GenAI to generate text based on the retrieved knowledge. It provides more accurate, reliable and efficient solution by combining the LLM's generative capabilities with the ability to retrieve and incorporate external knowledge.

- What is Langflow?

Langflow allows users to build AI applications like RAG with minimal or no coding (low-code), making it accessible to developers of varying skill levels and AI enthusiasts. It provides a drag and drop interface which enables users to connect various components to create complex AI workflows.

- How does Langflow help in building RAG based applications?

Langflow helps in building RAG applications by providing visual, low-code interface for designing and deploying complex AI workflows.

Langflow offers a drag and drop interface which enables users to connect different components like prompts, LLMs, vector databases and data sources. It provides pre-built modules and components for common RAG tasks like data loading, chunking, embedding and retrieval. Langflow also supports integration with various vector databases, including Astra DB. This allows developers to easily store and retrieve relevant information based on semantic similarity, a key aspect of RAG systems. The visual interface also allows users to easily debug and iterate on their RAG pipelines.