DAY – 17

On Day 17, the focus was on to finalize the document retrieval system and hook it into our Large Language Model (LLM) using the LangChain framework.

We conducted thorough testing of the RAG system:

- Confirmed the Retrieval mechanism correctly fetches the top-k most semantically relevant chunks based on complex test queries.

- Validated the Augmentation phase, ensuring the retrieved chunks are appended to the user's prompt as context.

We developed the primary Retrieval Chain in LangChain:

- **Prompt Template:** Created a system prompt instructing the LLM to act as a research assistant and to always ground its answers in the provided document context.

- **Handling Hallucinations:** Implemented safety checks in the prompt design to instruct the LLM not to answer if no relevant context is retrieved.

- Chained the RAG Retriever (which fetches the external knowledge) to the Gemini LLM (which generates the response).

CONCLUSION

- Successfully built the core RAG intelligence for the Capstone, demonstrating proficiency in making LLMs more context-aware and reliable.

- Practical application of the Augmentation process, which is crucial for grounded responses using user-specific content.