Retrieval-Augmented Generation (RAG)

RAG combines information retrieval with text generation.

Documents are embedded into vector databases like FAISS or Chroma.

At query time, relevant chunks are retrieved and passed to the LLM.