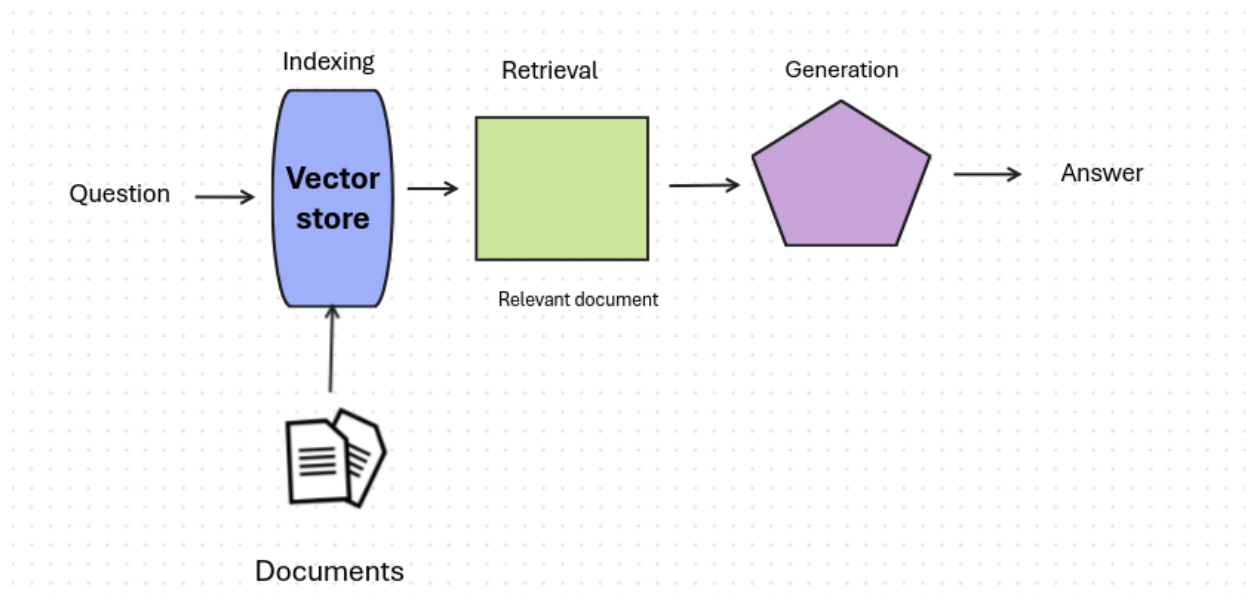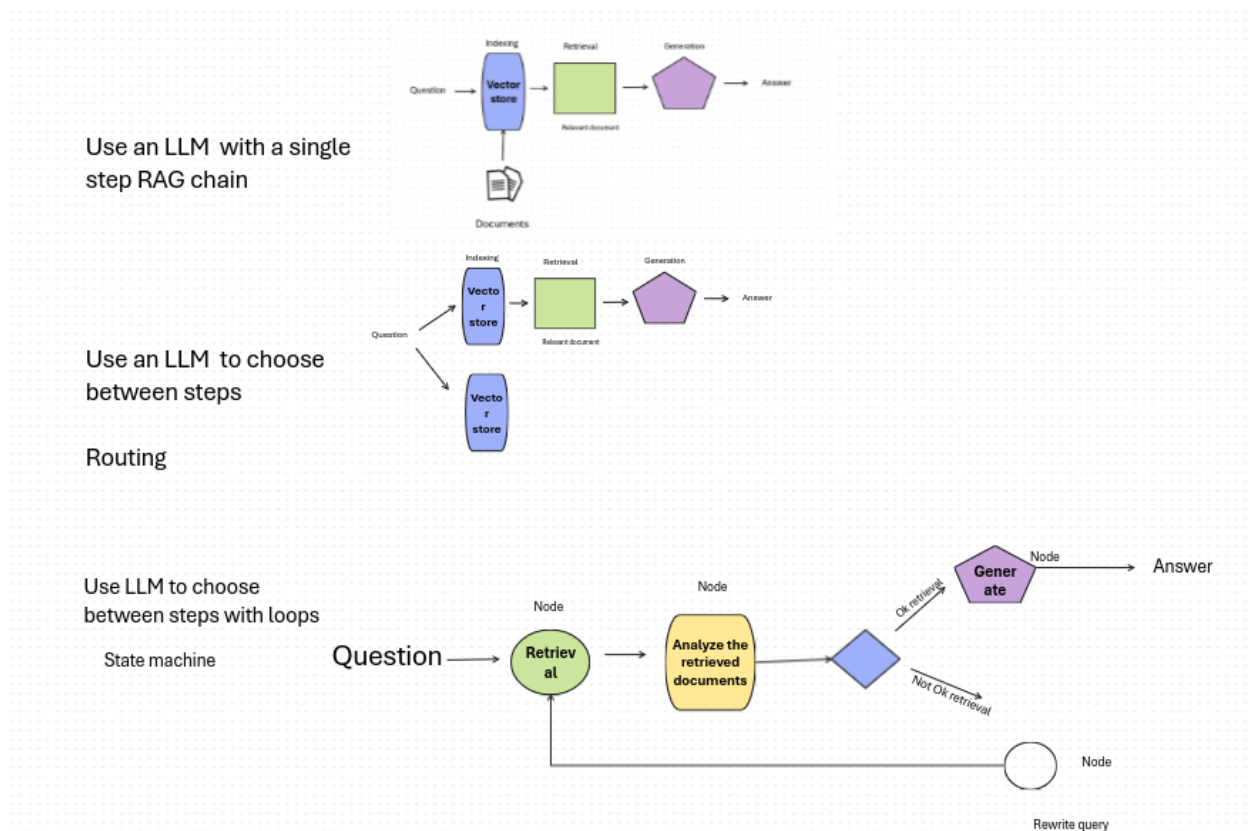# Active RAG

RAG Flow

Challenges

Implementing RAG sometimes require logical reasoning

- When to actually retrieve

- Are the retrieved documents actually good if not good then re try retrieval

- When to rewrite questions for better retrieval

Active RAG

LLM decides when and what to retrieve based on the retrieval or generation

Levels of control over LLMs in an RAG application

Use an LLM with a single step RAG chain

Use an LLM to choose between steps

Routing

Use LLM to choose between steps with loops

State machine

CRAG

Corrective RAG

Corrective-RAG (CRAG) is a strategy for RAG that does self-reflection on the retrieved documents.

Example:

Pseudo Code

- Take input

- Retrieve relevant documents

- If at least one document is relevant to the question, then it proceeds to generation

- If none of the documents are relevant then use web search to supplement retrieval