TEAM

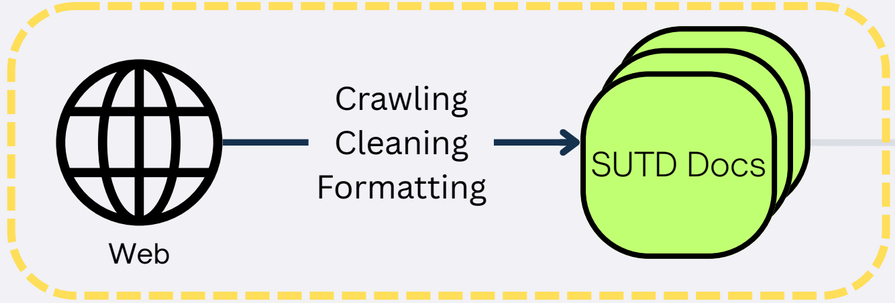
BYTEUS

Aditya Kumar	1006300
Mahima Sharma	1006106
Caitlin Chiang	1006537
Chin Wei Ming	1006264

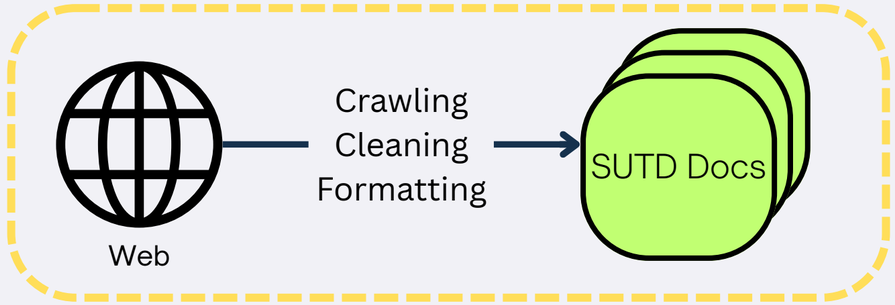
RAG PIPELINE

Requirements of the Chatbot

- Direct Answer
- Source of the Answer
- Further Links



RAG PIPELINE



Requirements of the Chatbot

- Direct Answer
- Source of the Answer
- Further Links

Initial Approach

Tried saving SUTD pages as PDFs

Change to LangChain RecursiveWebLoader

Crawled pages with JavaScript rendering

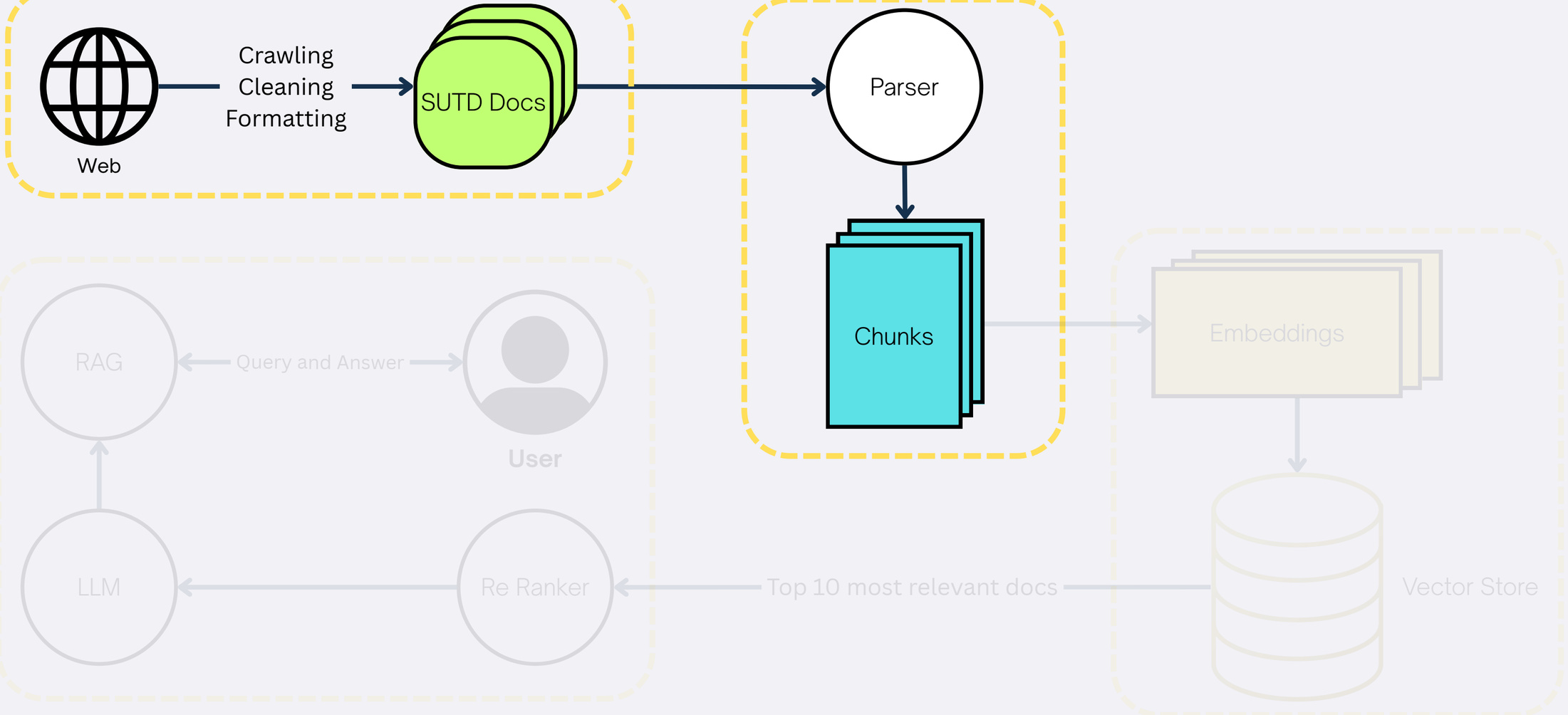
Target only SUTD Site

Focus Solely on SUTD Website

RAG PIPELINE

Requirements of the Chatbot

- Direct Answer
- Source of the Answer
- Further Links



RAG PIPELINE

One Chunk per page

Large Chunks that exceeds context window

Document-based Chunking

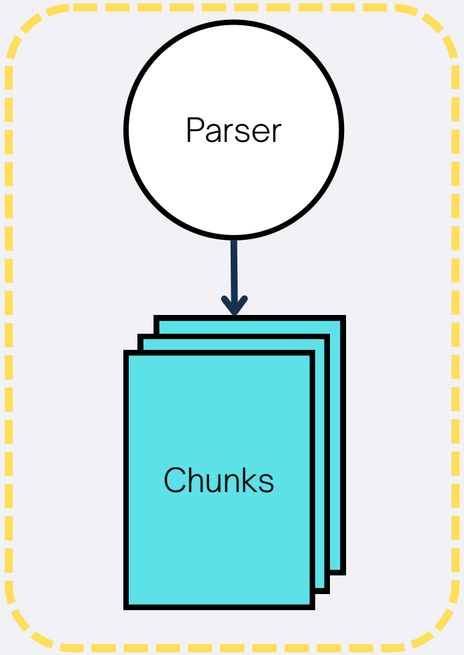
Hierarchical chunking using markdown headers

Graph-Based URL Tree

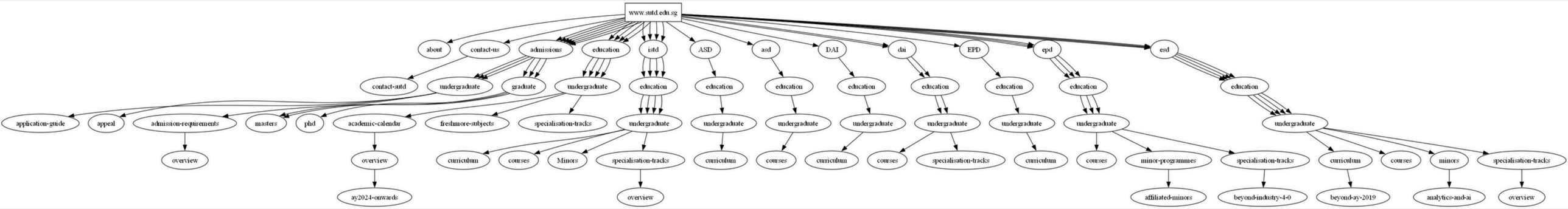
combining parent-child sections with similar topics

Requirements of the Chatbot

- Direct Answer
- Source of the Answer
- Further Links



0 RAG PIPELINE



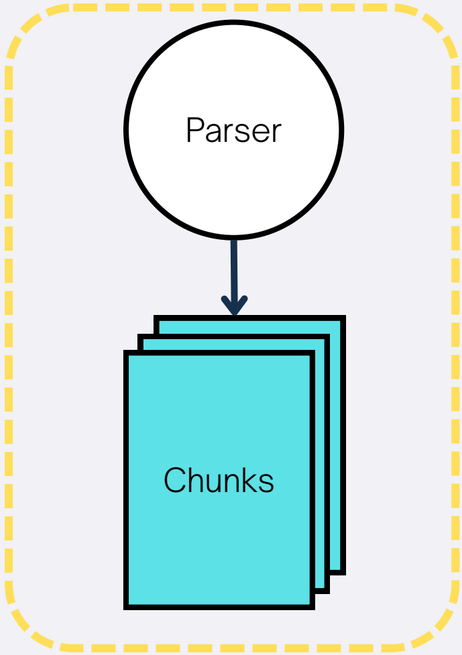
Graph-Based URL Tree

combining parent-child sections with similar topics

RAG PIPELINE

Requirements of the Chatbot

- Direct Answer
- Source of the Answer
- Further Links



One Chunk per page

Large Chunks that exceeds context window

Document-based Chunking

Hierarchical chunking using markdown headers

Graph-Based URL Tree

combining parent-child sections with similar topics

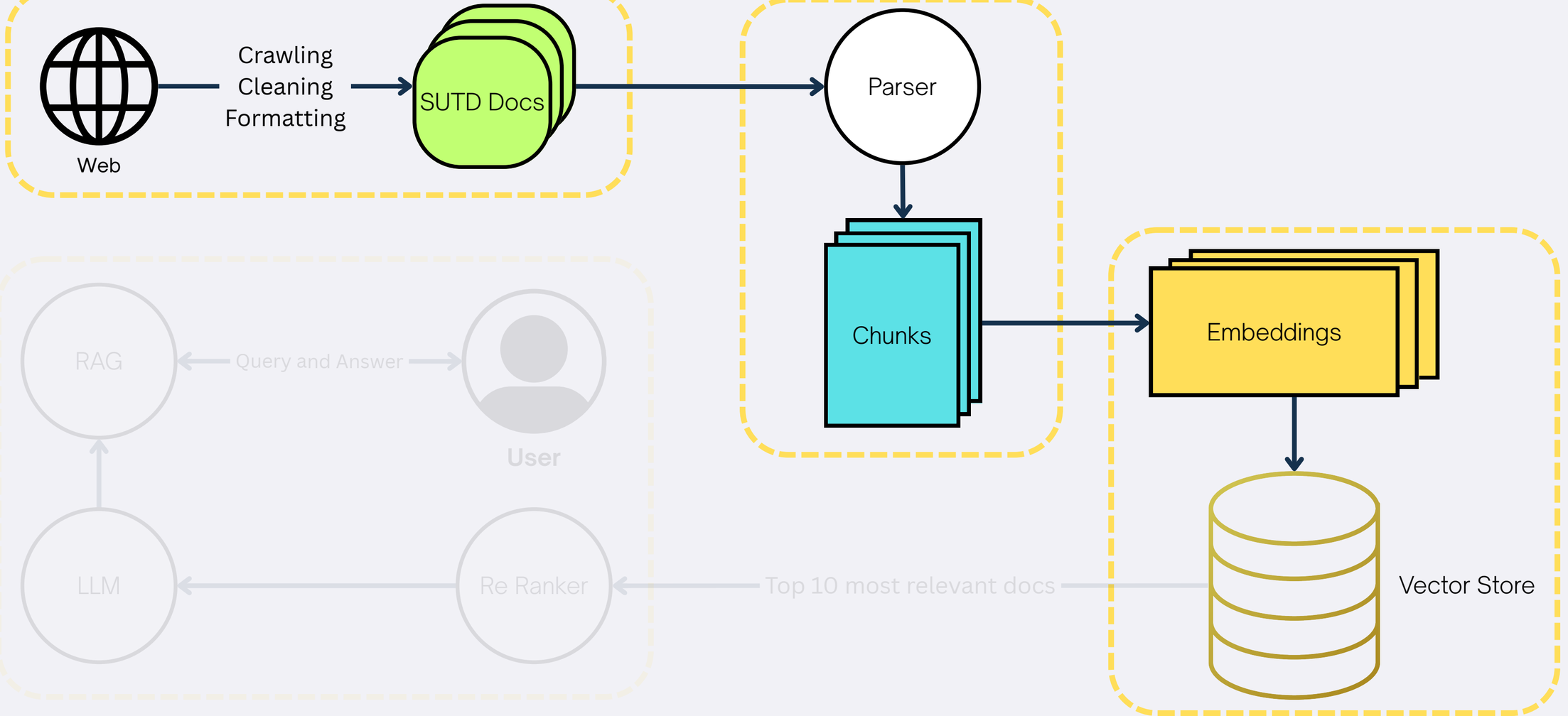
Semantic Chunking

Chunking based on semantic similarity.

chunk_size = 1000
chunk_overlap = 100
min_chunk_size = 100

RAG PIPELINE

Requirements of the Chatbot
- Direct Answer
- Source of the Answer
- Further Links



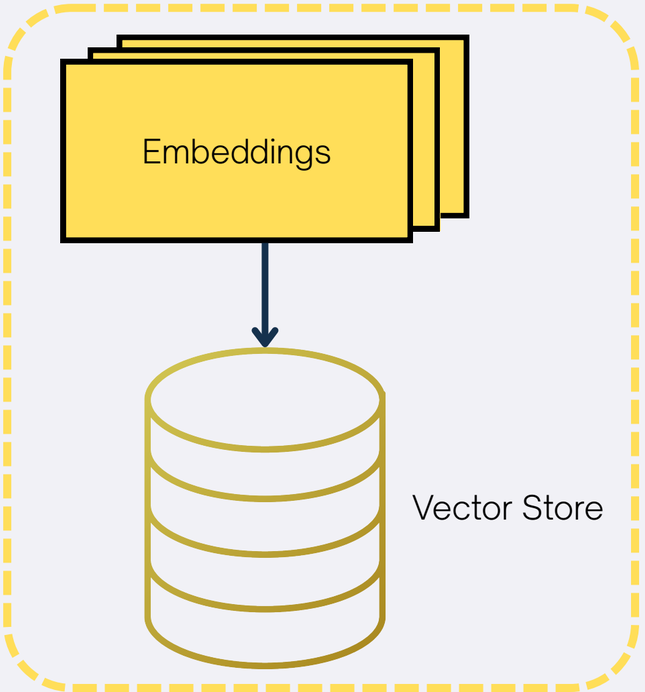
RAG PIPELINE

Requirements of the Chatbot

- Direct Answer
- Source of the Answer
- Further Links

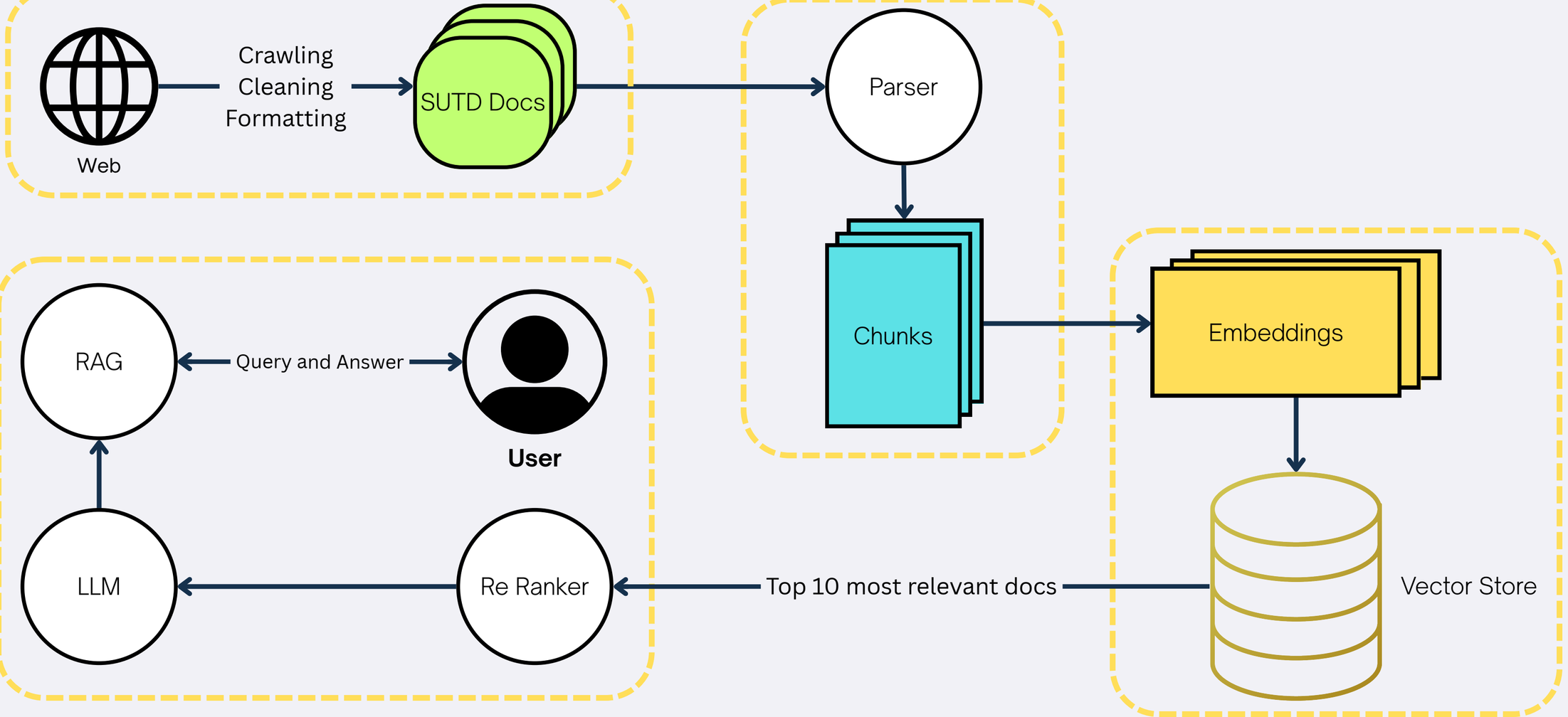
FAISS (Facebook AI Similarity Search)

A highly efficient vector store with fast query performance and accurate similarity matching.

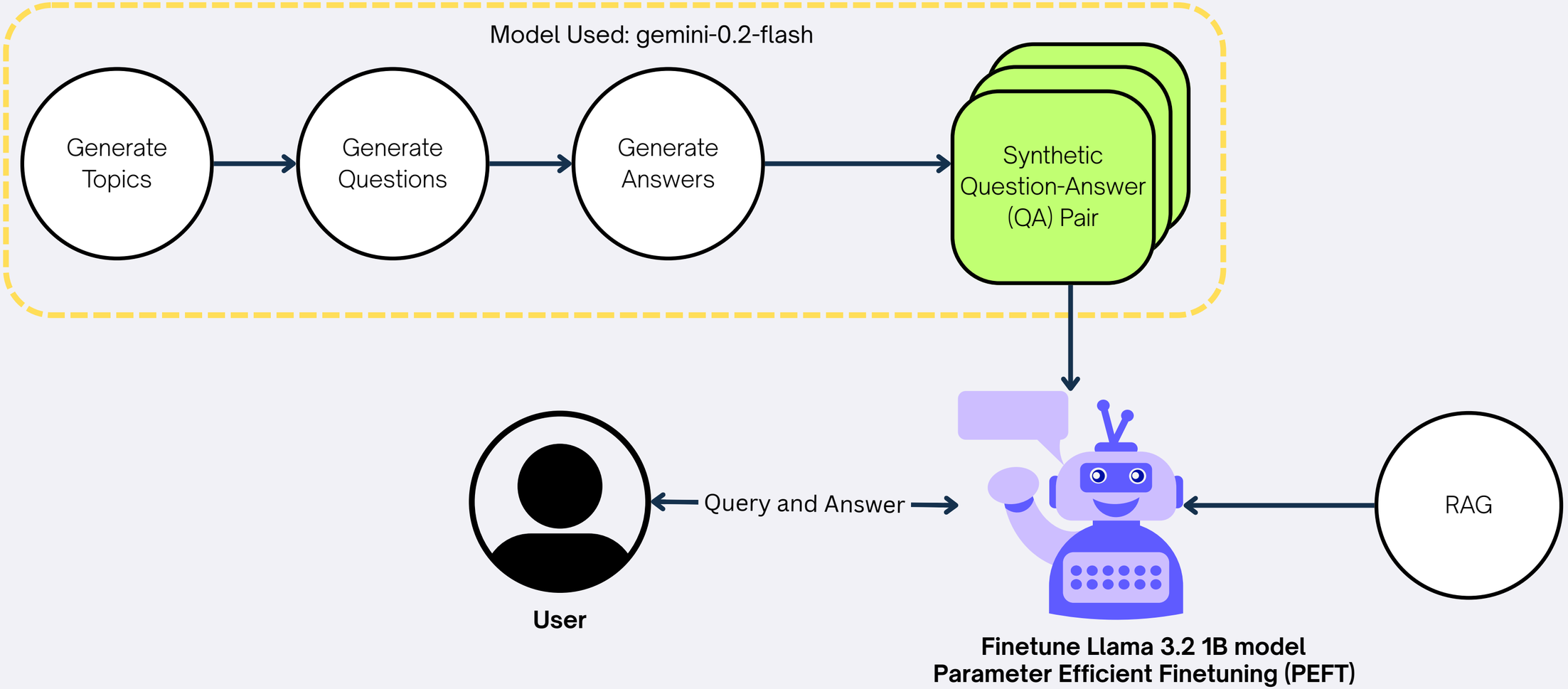


RAG PIPELINE

Requirements of the Chatbot
- Direct Answer
- Source of the Answer
- Further Links

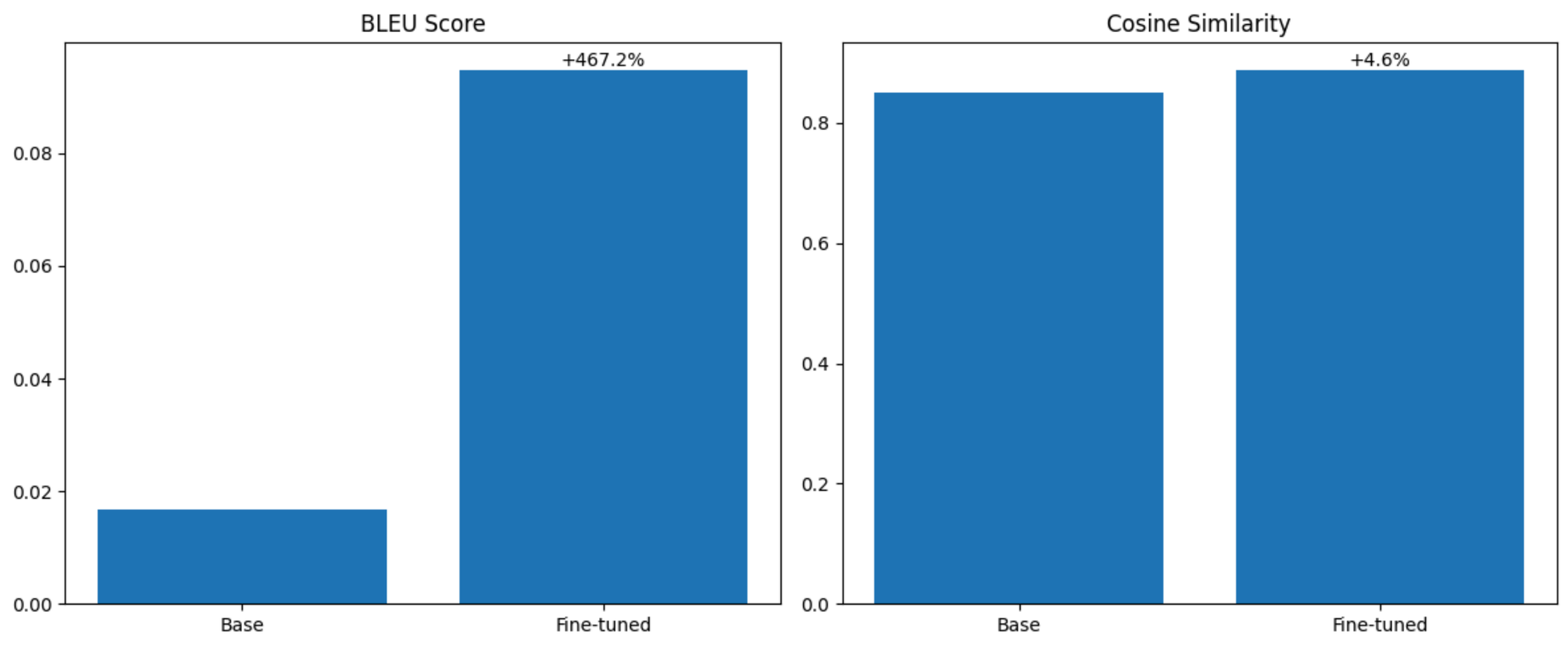


FINETUNING LLMS



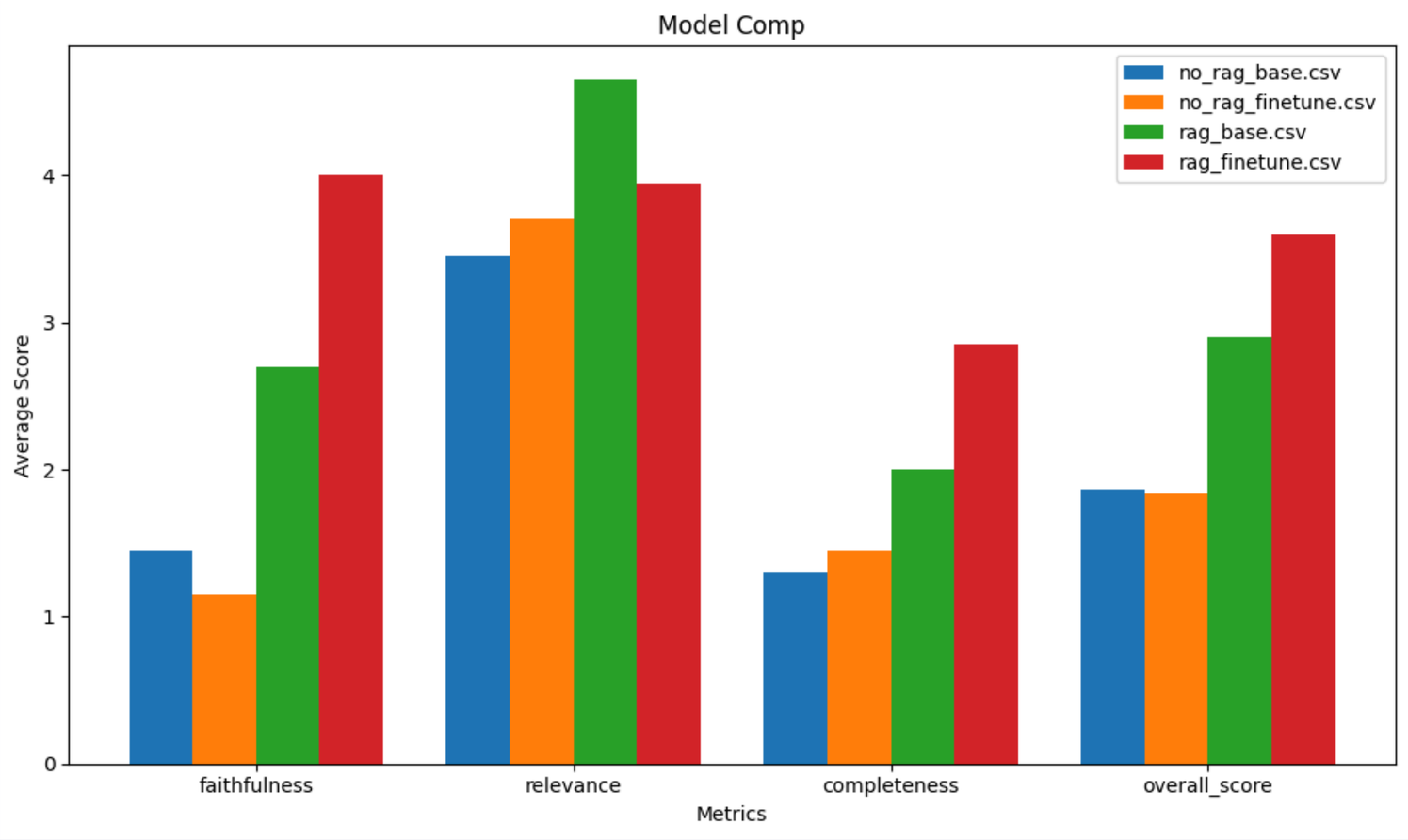
05 EVALUATION

- BLEU SCORE & COSINE SIMILARITY



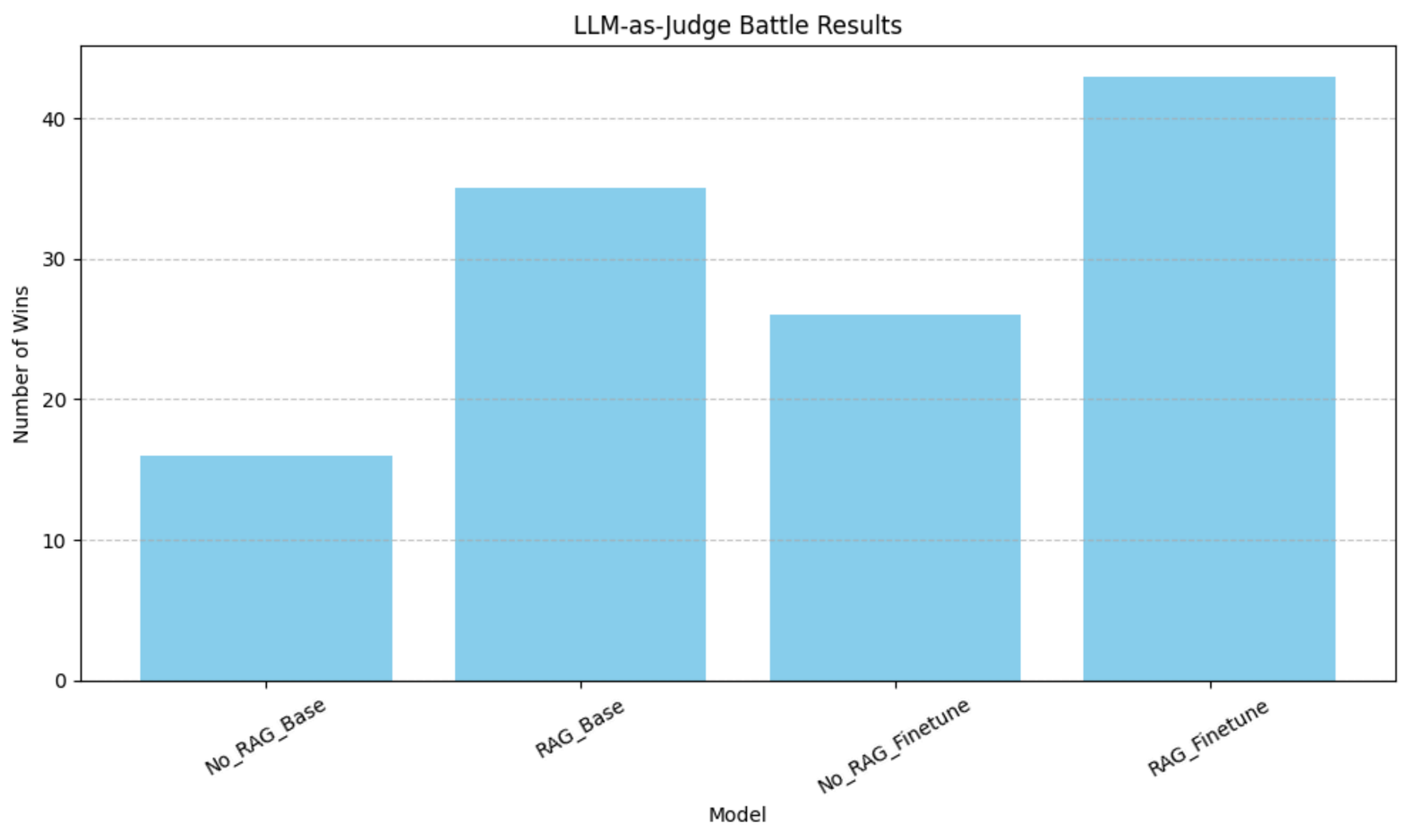
LLM-AS-A-JUDGE - QA EVALUATION

Model Used: gemini-0.2-flash



LLM-AS-A-JUDGE - BATTLES

Model Used: gemini-0.2-flash



DEMO