Multimodal RAG Application: Chat with the all present images in a PDF

Extracting all images from PDF

Generating summary of each image using LLM

Text Embeddings of summaries

Storing

Vector Store

Query

Retriever

Query +

Response

a4archit