

Case study: Developing a Complex Generative AI application - QA RAG Chatbot

Instructor

Dipanjan Sarkar

AI Advisor at Analytics Vidhya

Google Developer Expert - ML & Cloud Champion Innovator

Published Author

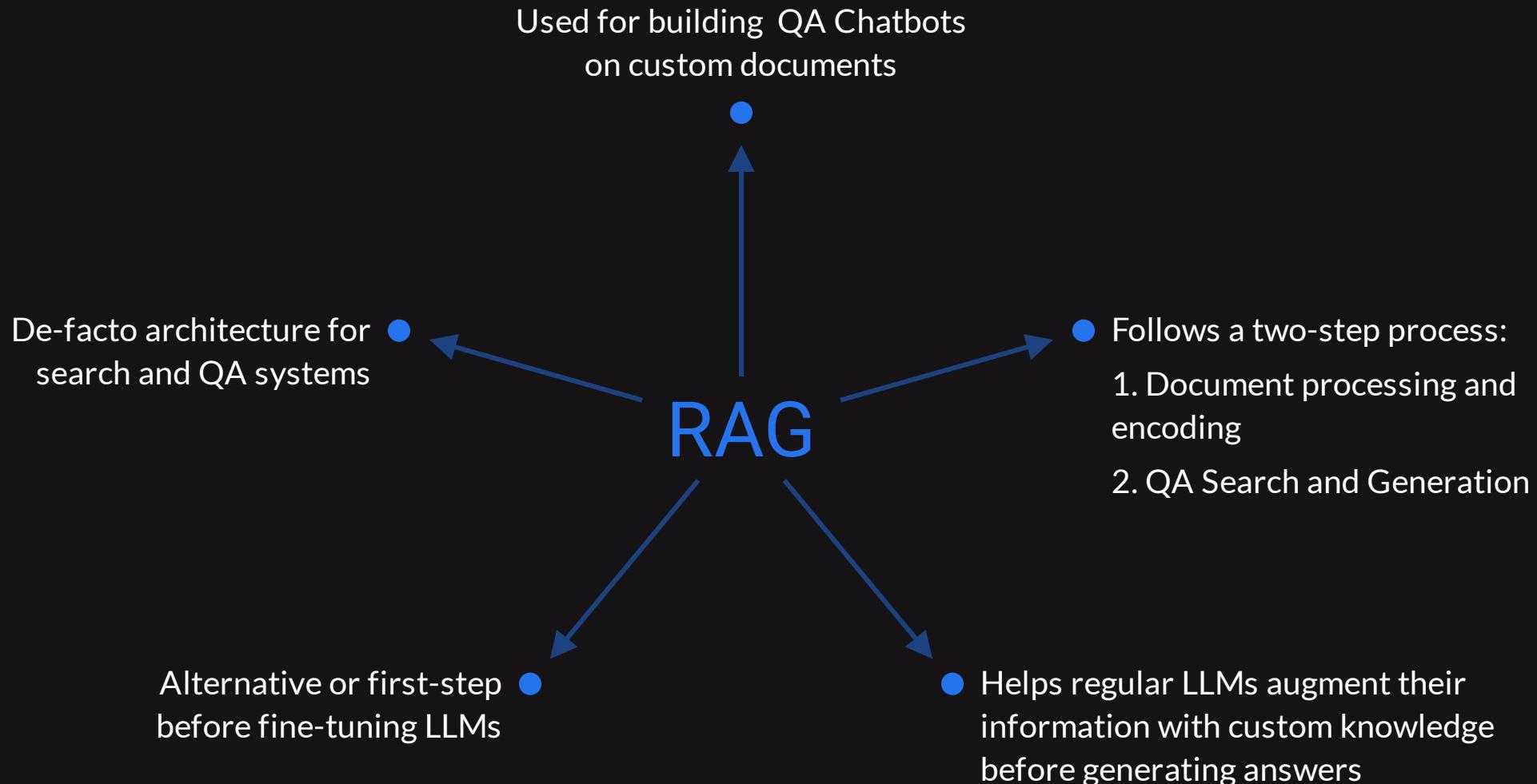


Outline

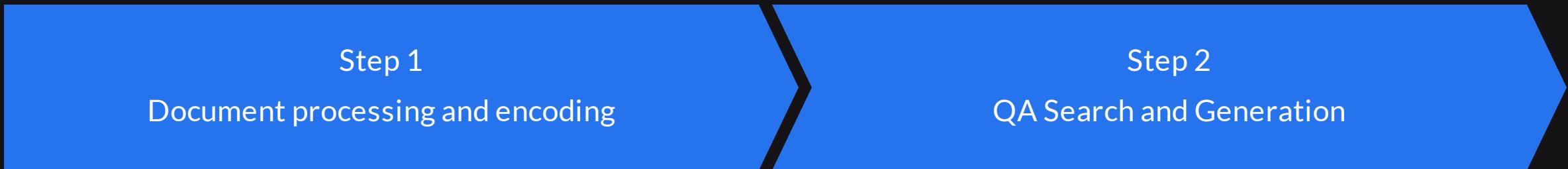
- What is a RAG System
- RAG System Architecture
- Key tools to build a UI-based QA RAG Chatbot App
- UI-based QA RAG Chatbot Architecture
- Authentication Keys for App
- Store ChatGPT and NGrok API Keys in YAML format
- RAG Case Study - QA Chatbot for Rare Diseases

What is a RAG System

RAG stands for Retrieval Augmented Generation



What is a RAG System



- Custom documents are processed and chunked
- These chunks are converted to embeddings with a LLM (transformer)
- Chunks and embeddings are stored in a Vector Database (along with metadata)

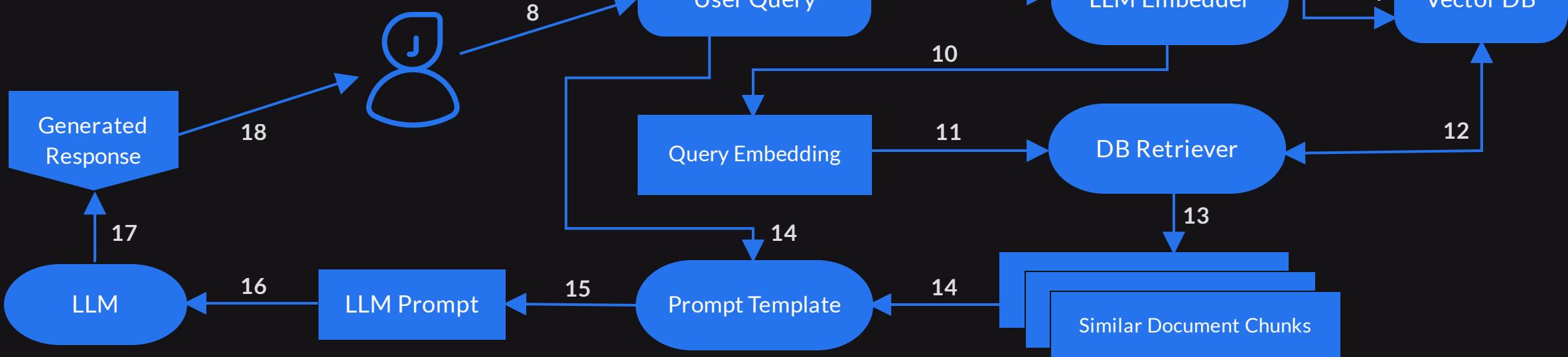
- Based on user query the system **retrieves** relevant document chunks
- Passes these chunks to the LLM along with the query to **augment** its knowledge
- The LLM **generates** a human-like response to the user query based on this information

RAG System Architecture

Step 1 - Document processing and encoding



Step 2 - QA Search and Generation



Key tools to build a UI-based QA RAG Chatbot App



[LangChain](#) - A Python framework to build
LLM systems

Key tools to build a UI-based QA RAG Chatbot App



LLMs like [Open AI ChatGPT](#)

Key tools to build a UI-based QA RAG Chatbot App



Google Gemini Pro

LLM Embedding Model from [Open AI](#) or
[Google Gemini](#) or [HuggingFace](#)



HuggingFace

Key tools to build a UI-based QA RAG Chatbot App



[Streamlit or Chainlit](#) - A Python framework to build conversational UI



Key tools to build a UI-based QA RAG Chatbot App



Google Colab or a Python environment for running your apps



Key tools to build a UI-based QA RAG Chatbot App



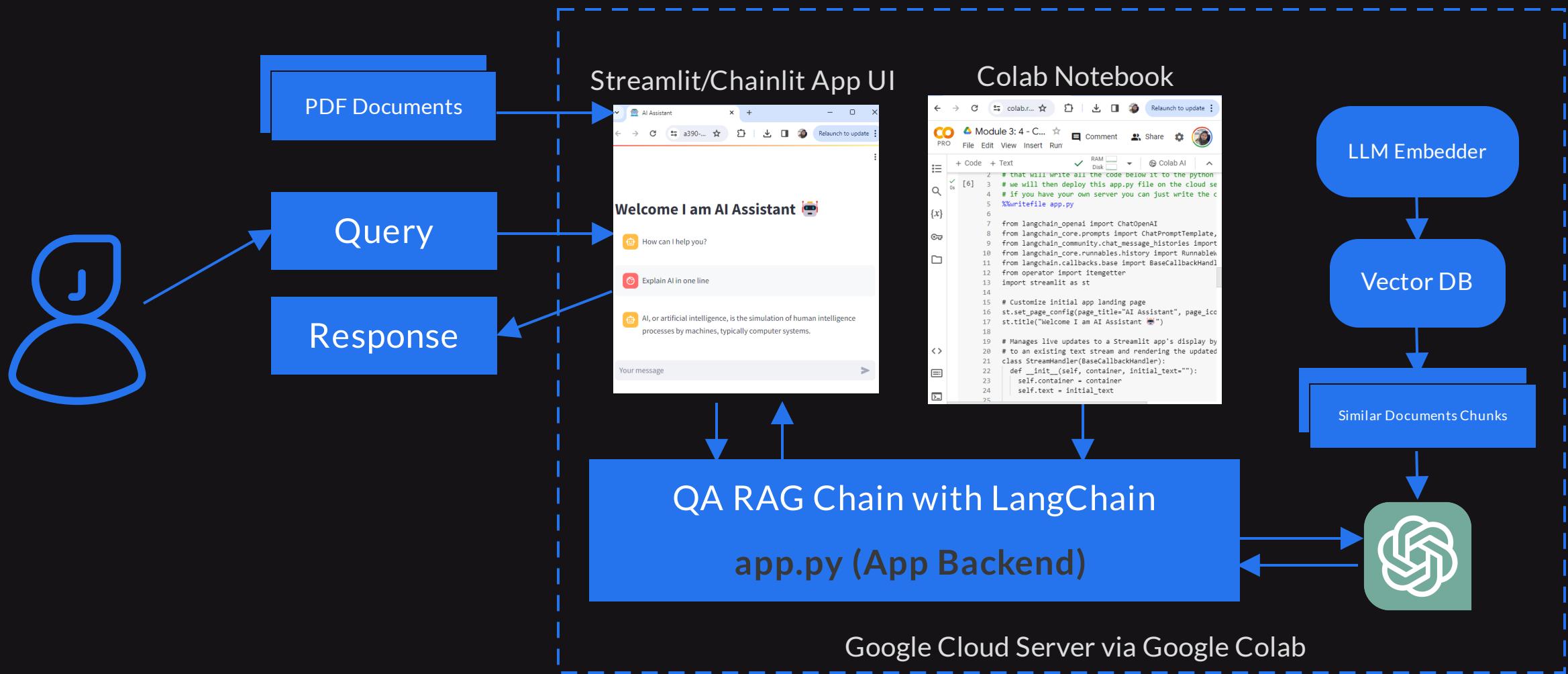
Chroma DB - A vector database to store embeddings and retrieve similar documents

Key tools to build a UI-based QA RAG Chatbot App

ngrok

ngrok to access deployed app publicly

UI-based QA RAG Chatbot Architecture



Authentication Keys for App

- ChatGPT Open AI Key (or use Google Gemini or HuggingFace LLMs)
- NGrok API Key

Store ChatGPT and NGrok API Keys in YAML format

chatgpt_api_credentials.yml ×

```
1 openai_key : sk-i0SmpANG1avZYP1DBIt
```

ngrok_credentials.yml ×

```
1 ngrok_key : 1rnMGVDt6L57aU12JSqNVB  
2
```

- Store each key as key-value pairs as shown on the left side in text files
- Make sure the files have a [.yml](#) extension
- We will load these files when building apps to not expose keys publicly in the app code

RAG Case Study - QA Chatbot for Rare Diseases

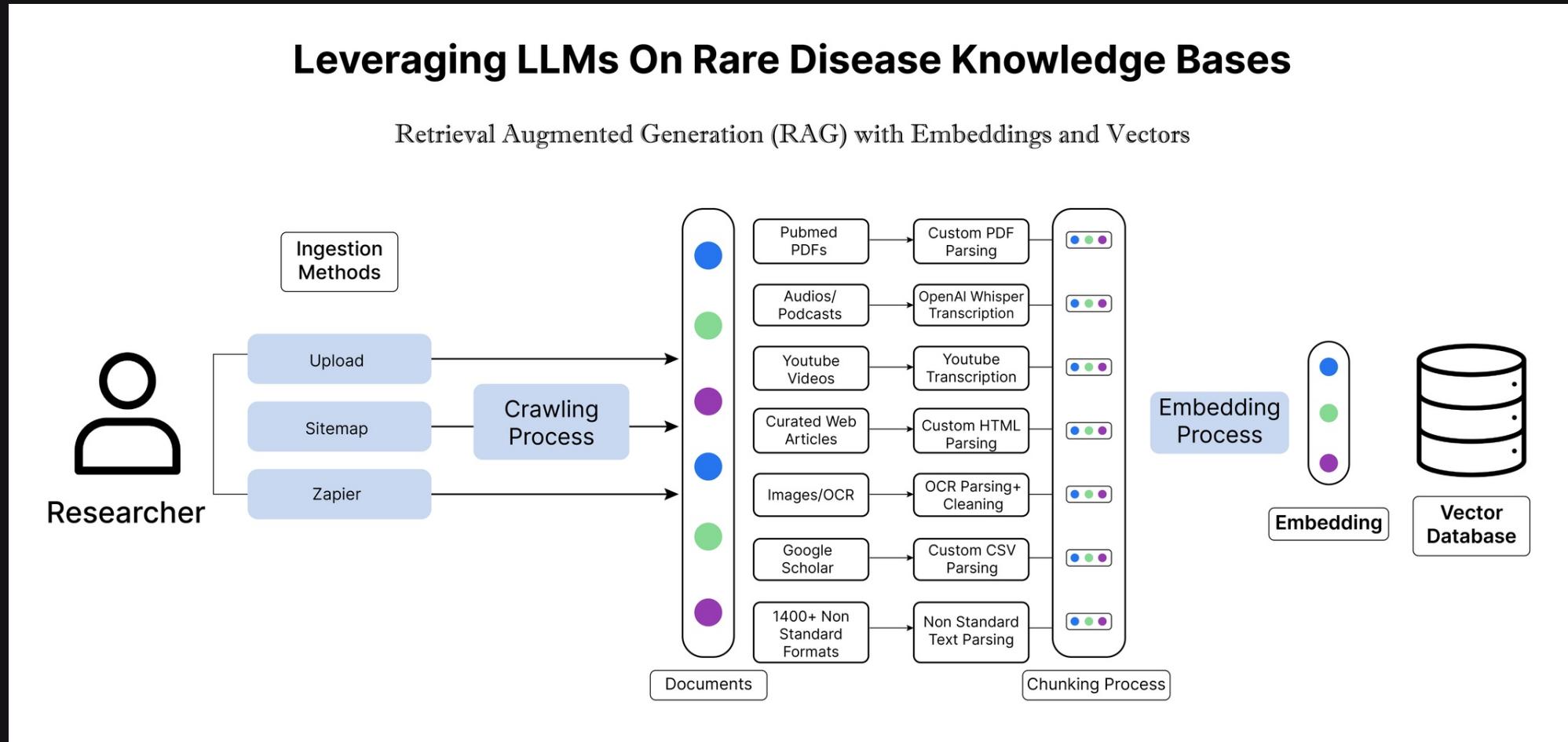
Alden Do Rosario built DentDiagnosisGPT using a RAG system

The screenshot shows a dark-themed chat application window. At the top, there's a purple header bar with a small user icon and a search bar containing the text "My child has been having kidney stones, is this a symptom of Dent's disease?". Below this is a blue sidebar with a circular icon containing a white "DENT" logo. The main conversation area has a light gray background. A user message is shown in a blue box: "Yes, the formation of kidney stones is a common symptom of Dent's disease. Patients with Dent's disease often experience recurrent episodes of kidney stones (nephrolithiasis), which can cause various symptoms including blood in the urine (hematuria), painful urination (dysuria), the urge to urinate often, abdominal pain (renal colic), blockage of the urinary tract, and repeated urinary tract infections. It's important to note that while kidney stones are a symptom of Dent's disease, they can also be caused by other conditions. Therefore, a thorough evaluation and possibly genetic testing are recommended to confirm a diagnosis of Dent's disease." Below this message is a follow-up question: "What primary symptom or concern about Dent's disease brings you in today? Would you like more information on the diagnosis process or other symptoms associated with Dent's disease?" Underneath, a "Transparency:" section explains the response was crafted based on context information about kidney stones being a symptom of Dent's disease. At the bottom of the message area, there are small icons for reply, like, and dislike, followed by a "Concerned about Dent's disease? Let's delve into your symptoms." button and a text input field placeholder "Shift + Enter to add new line". The footer of the window is white and displays the text "Powered by CustomGPT.ai".

SOURCE: [HTTPS://ADOROSARIO.GITHUB.IO/DENTS/](https://adorosario.github.io/dents/)

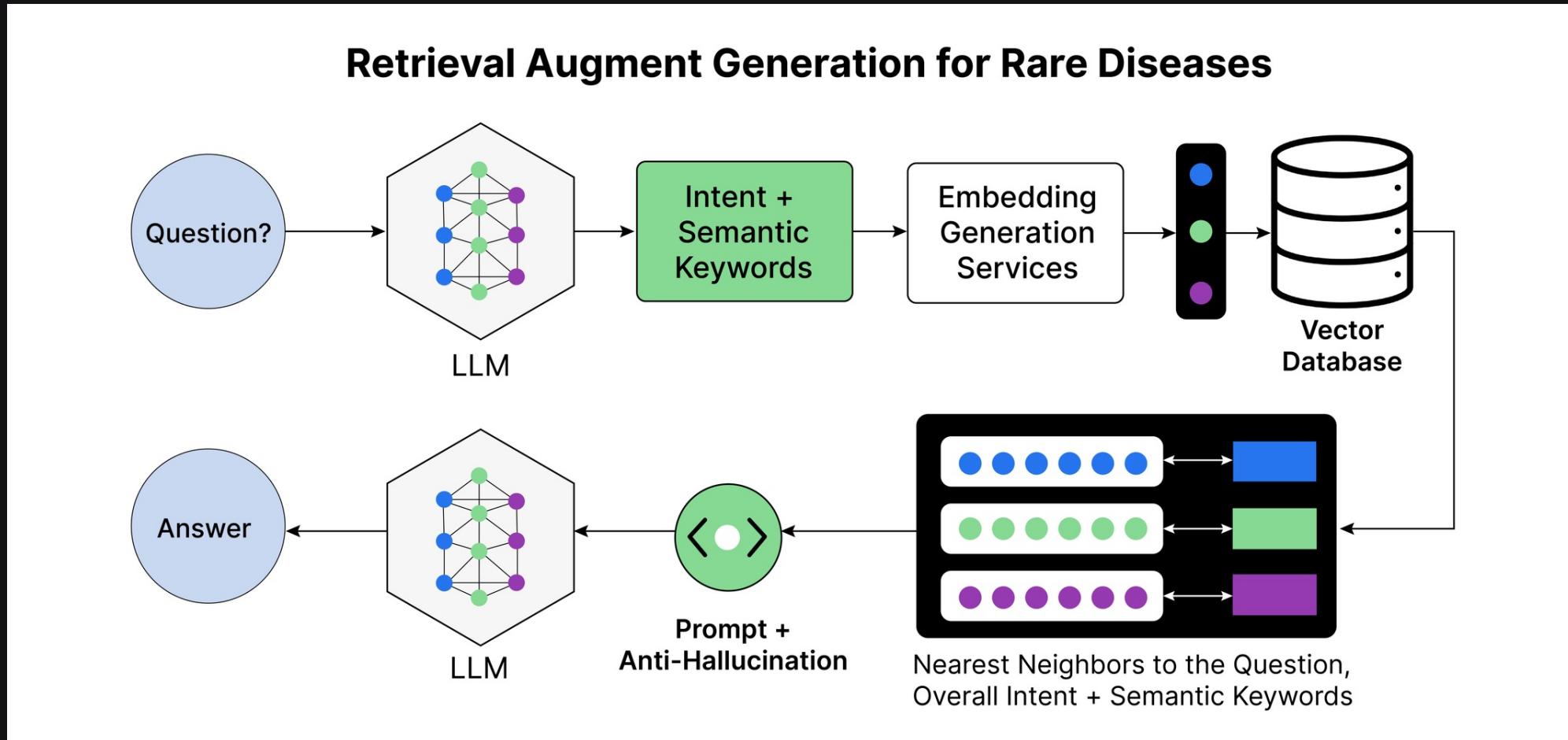
- It is a QA RAG chatbot that uses CustomGPT + ChatGPT to answer questions about the diagnosis of Dent's disease
- Leverages a two-step RAG architecture and workflow which we discussed earlier with some enhancements
- More details and live App available at:
 - <https://adorosario.github.io/dents/>

RAG Case Study - QA Chatbot for Rare Diseases - Step 1



Source: <https://generativeai.pub/case-study-how-ai-can-help-cure-rare-diseases-a55f46de32d1>

RAG Case Study - QA Chatbot for Rare Diseases - Step 2



Source: <https://generativeai.pub/case-study-how-ai-can-help-cure-rare-diseases-a55f46de32d1>

Thank You
