

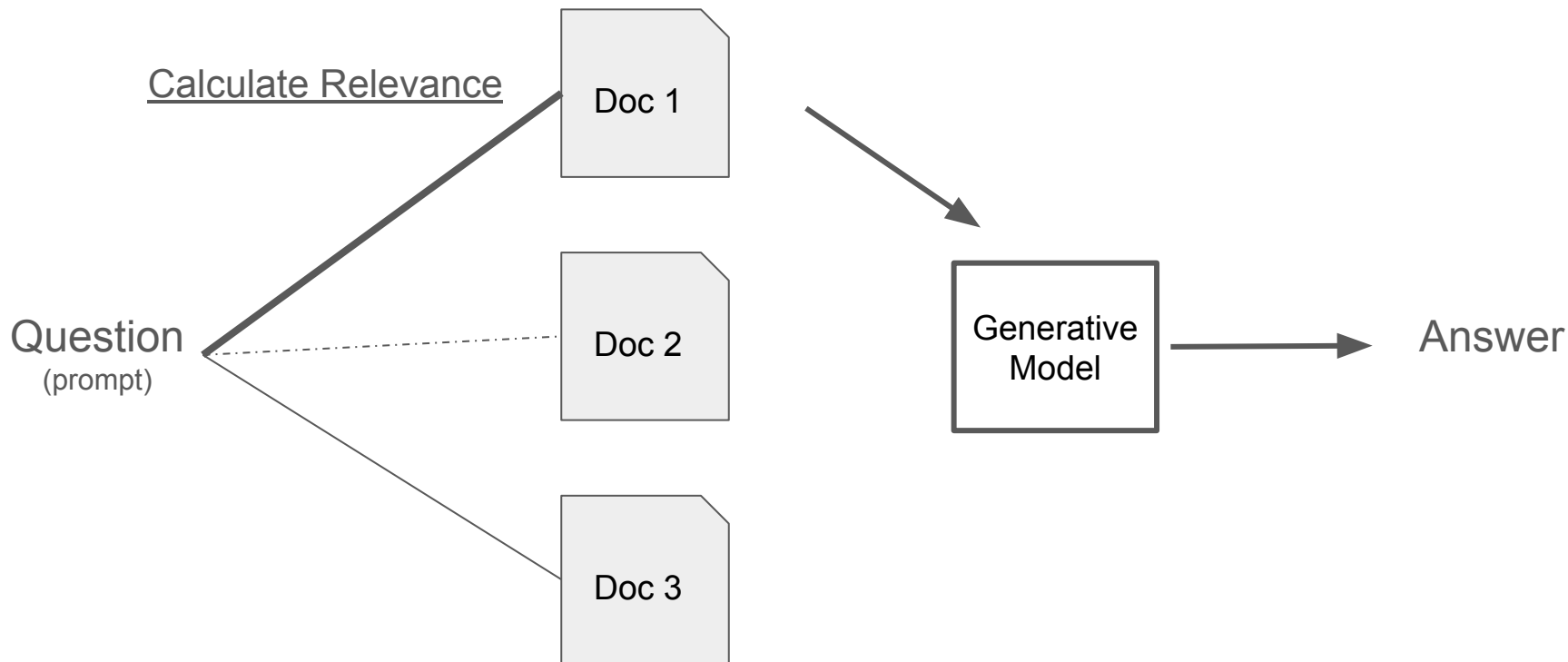
RAG-based Medical QA

Progress
Sep. 17, 2024

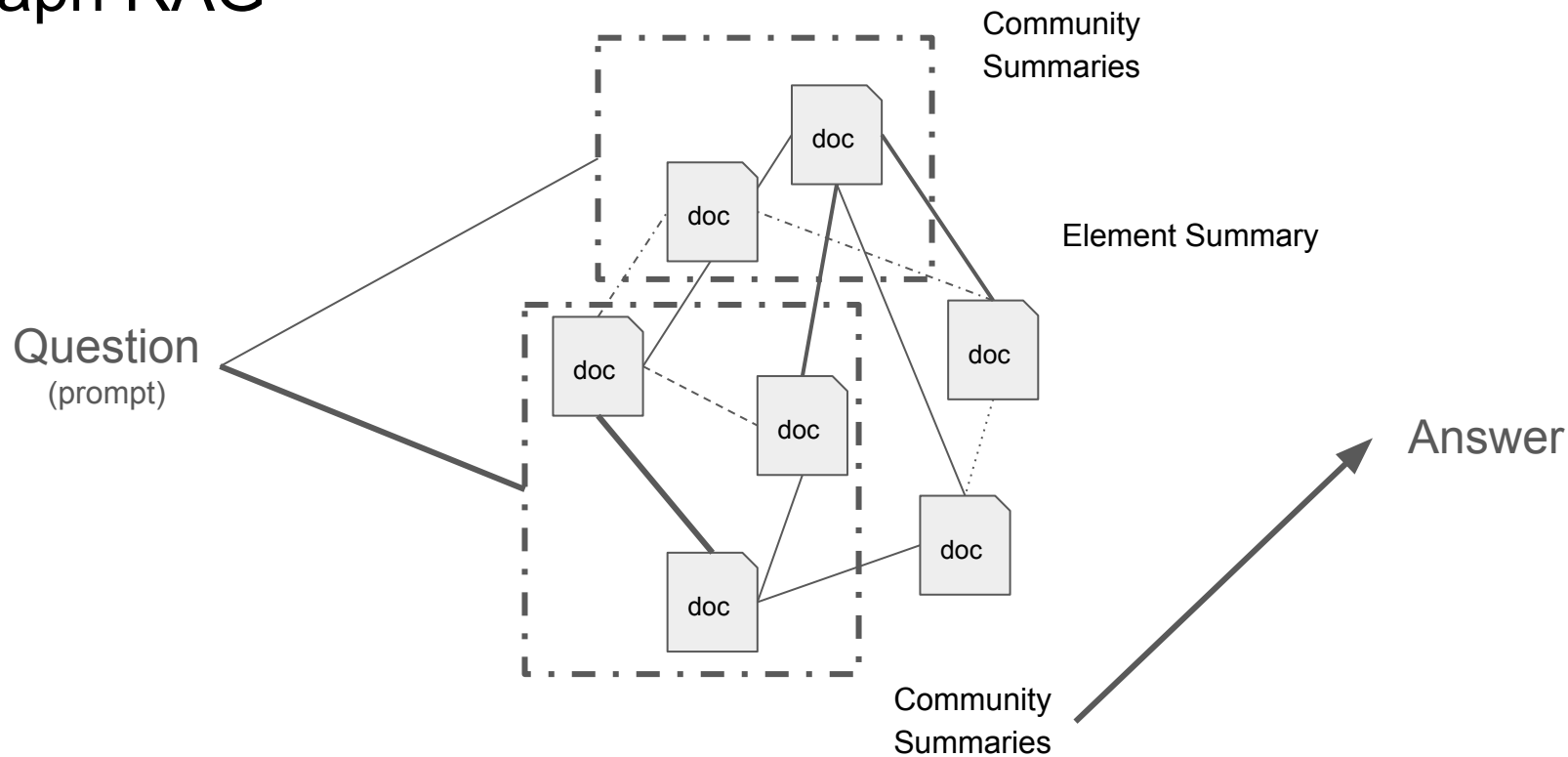
Outline

- Literature Review
 - From RAG to GraphRAG
- Developing Proof-of-Concepts
- Building Cloud-Based Systems

Retrieval Augmented Generation (RAG)



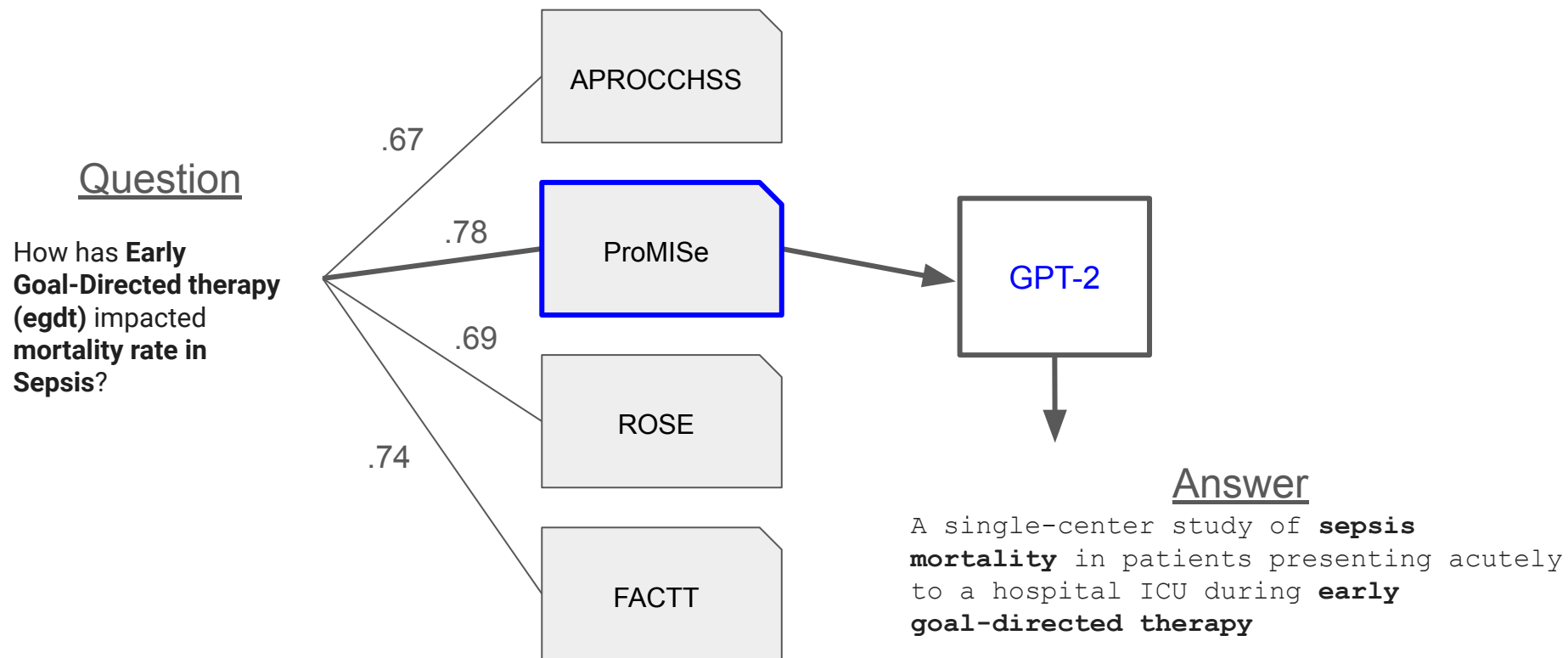
Graph RAG



Example Questions

Question	Answer	Source
What is the outcome of steroid (combination) therapy in treating patients other comorbidities with severe sepsis ?	Two papers agreed with the use of steriod therapy the cumulative trial data suggest that if steroids do in fact improve survival in septic shock, the effect is modest at best.	APROCCHSS
How has Early Goal-Directed therapy (egdt) impacted mortality rate in Sepsis ?	There is no impact in mortality rate using egdt.	ProMISe
How does early neuromuscular blockage improve mortality ?	Neuromuscular blockage in reducing mortality rate is still unknown as there was no significant difference using them and without using them.	ROSE
What is the benefit of conservative fluid management in treatment of ARDS ?	Conservative fluid management plays a beneficial role in reducing mortality	FACTT

Using BERT to Represent Document



* Using **BERT** *cls* token to **represent abstracts** and **cosine similarity** to **calculate relevance** of the prompt and the abstracts.

Next Steps: Proof-of-Concepts

- Build GraphRAG based model
- Generate Sample Questions
- Study ways to Evaluate Answers

Structured Simple LLM Model using Free API

```
1 from flask import Flask, render_template, request
2 from openai import OpenAI
3 from dotenv import load_dotenv
4 import os
5
6 # load environment
7 load_dotenv()
8
9 # fetch keys from environment dir
10 api_key = os.getenv("OPENAI_API_KEY")
11
12 client = OpenAI(api_key=api_key)
13
14 app = Flask(__name__)
15
16
17 @app.route("/", methods=["GET", "POST"])
18 def index():
19     if request.method == "POST":
20         user_input = request.form.get("user_input")
21
22         if not user_input:
23             return render_template("index.html", error="No input provided.")
24
25         # New API call using 'chat.completions'
26         response = client.chat.completions.create(
27             model="gpt-3.5-turbo", # Use the latest GPT model
28             messages=[
29                 {"role": "system", "content": "You are a helpful assistant."},
30                 {"role": "user", "content": user_input},
31             ],
32         )
33         result = response.choices[0].message["content"].strip()
34         return render_template("index.html", result=result)
35
36 return render_template("index.html")
37
```

AI Text Generator

Enter a prompt below and get a response from OpenAI!

Prompt

Enter your prompt here...

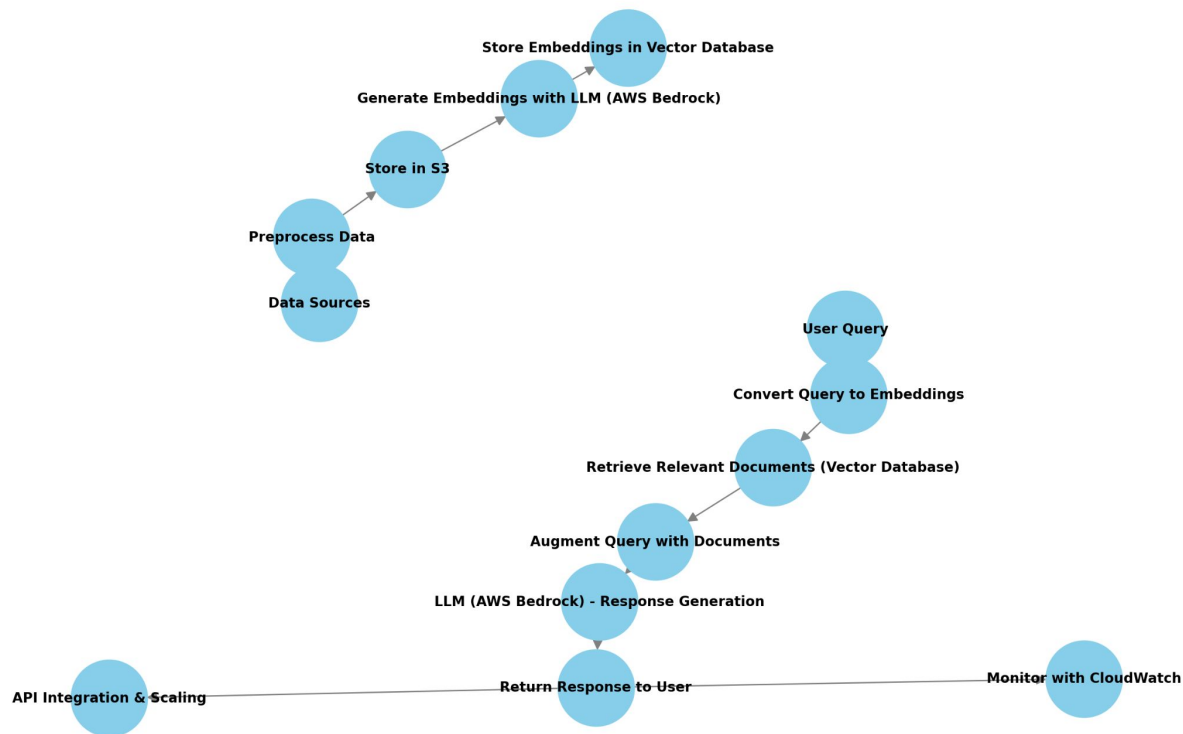
Submit

NO AVAILABLE CREDITS IN OPEN
AI ACCOUNT

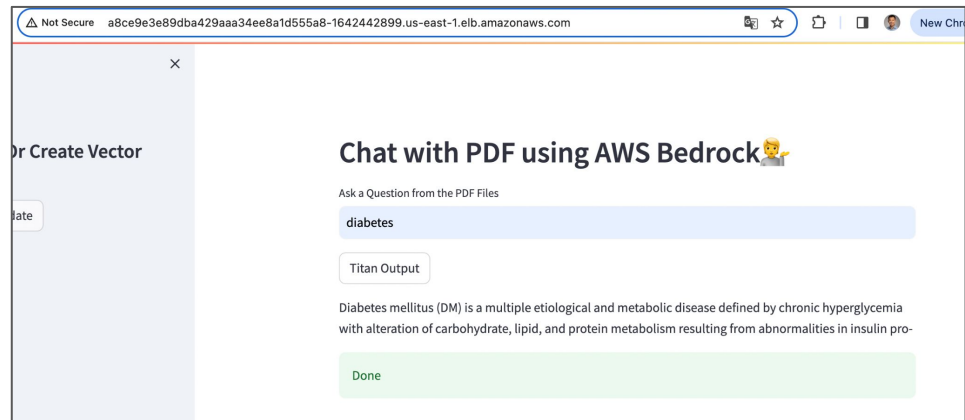
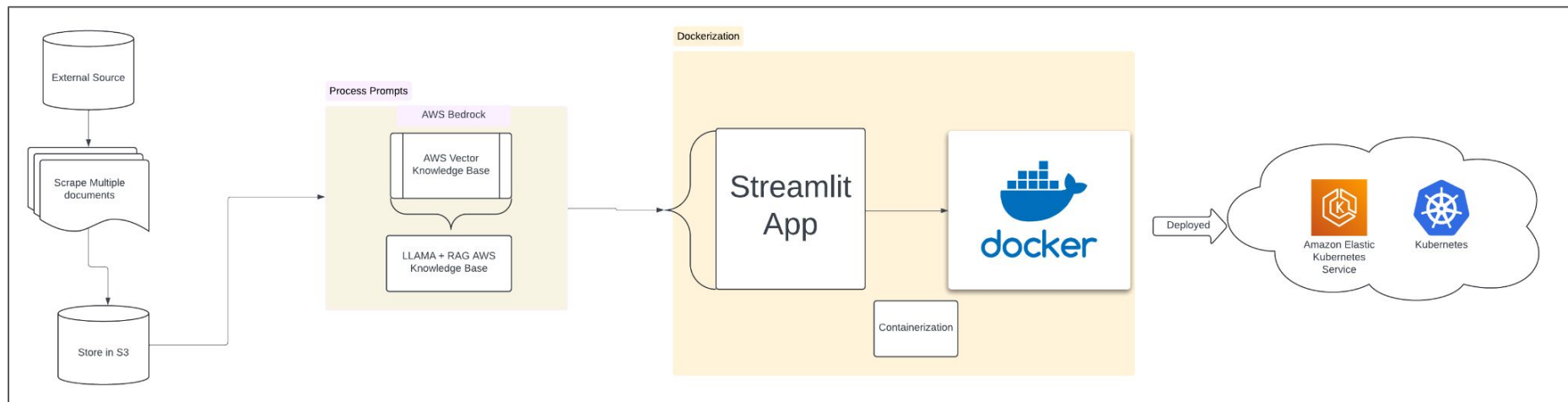
ALREADY ASKED MIDS ADMIN

Overview of the development

AWS Bedrock Chatbot Development Workflow



What we have



Next Steps: Dive into Data Source

- Continue researching a suitable LLM model for this project (either free or paid API)
- Build a simple vectorized database from PDF documents using an S3 bucket