Google Developer Experts

# Retrieval Augmented Generation (RAG)

Integrate LLMs into your custom dataset.

Tarun Jain
Data Scientist at AI Planet
@TRJ_0751

**You**

at what time does Iskcon temple open?

**You**

at what time does Iskcon temple open?

**ChatGPT**

The opening hours of ISKCON (International Society for Krishna Consciousness) temples can vary depending on the location. Each ISKCON temple may have its own schedule for opening and closing. To find the specific opening hours for a particular ISKCON temple, it is recommended to check the official website of that temple or contact the temple directly. Additionally, you can inquire about the opening hours by contacting the local ISKCON community or checking their social media pages for the most up-to-date information.

**You**

CONTEXT: Morning – 4:30 am to 5:20 am and
7:15 am to 1:15 pm
Evening – 4:15 pm to 8:00 pm

at what time does Iskcon temple open?

**ChatGPT**

Based on the provided schedule, the ISKCON temple is open during the following hours:

Morning:

- 4:30 am to 5:20 am
- 7:15 am to 1:15 pm

Evening:

- 4:15 pm to 8:00 pm

So, it appears that the temple opens at 4:30 am in the morning and remains open until 1:15 pm. After a break, it reopens in the evening at 4:15 pm and closes at 8:00 pm.

# Issues with Large Language Models

Challenges faced by LLMs

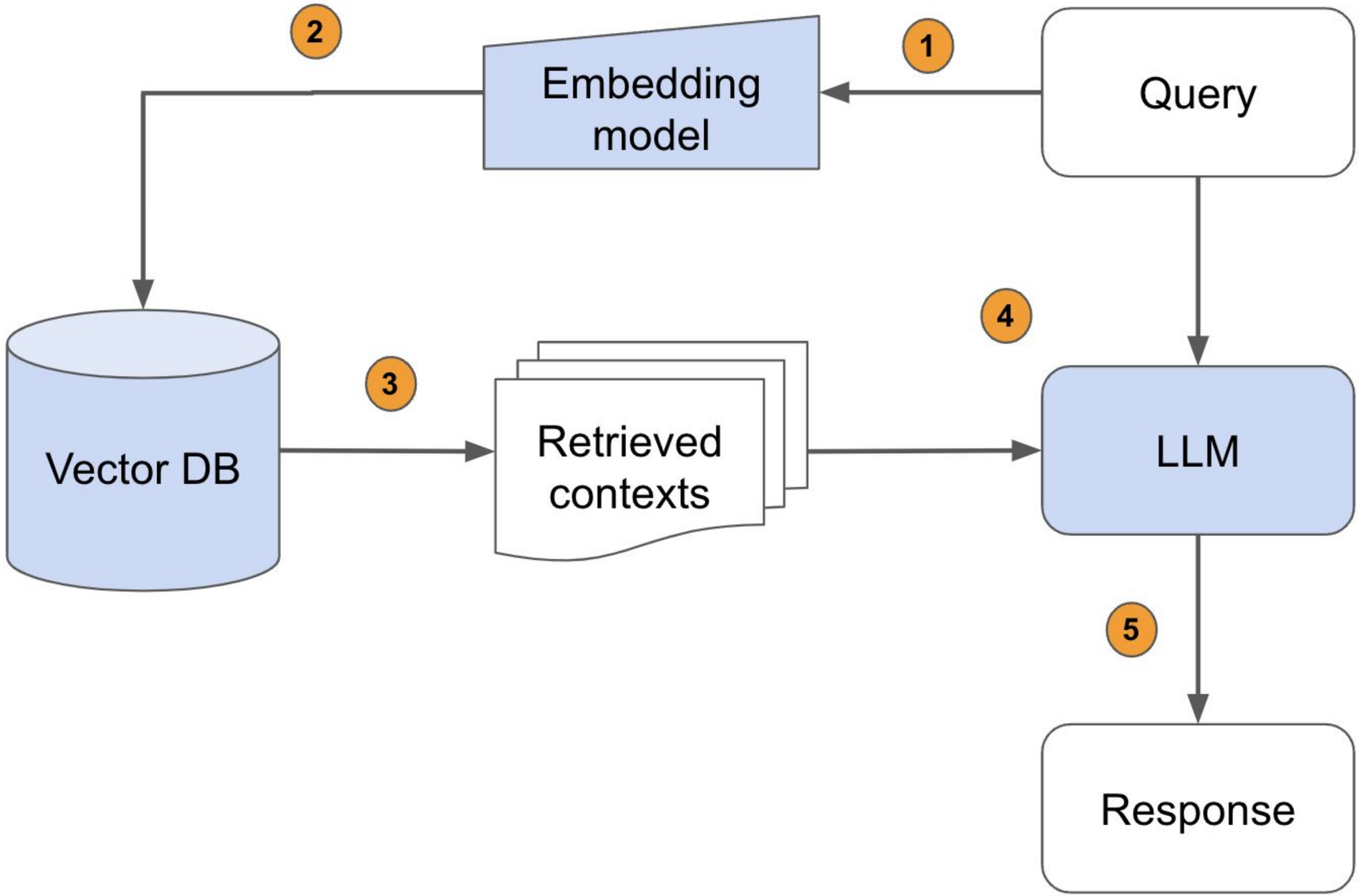
Experts

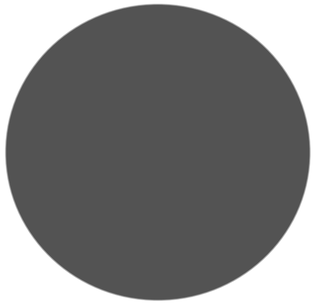
1. Hallucinations
2. Knowledge cut-off
3. Lack of Domain specific Factual responses

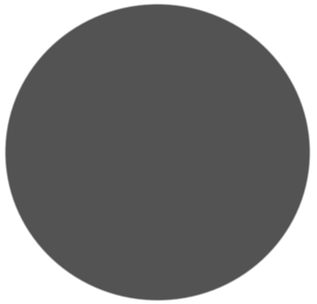
# Retrieval Augmented Generation

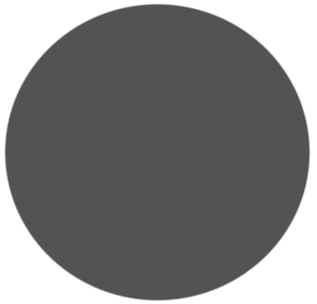
Chat with your own data

Experts


2
Embedding model
1
Query
Vector DB
3
Retrieved contexts
4
LLM
5
Response



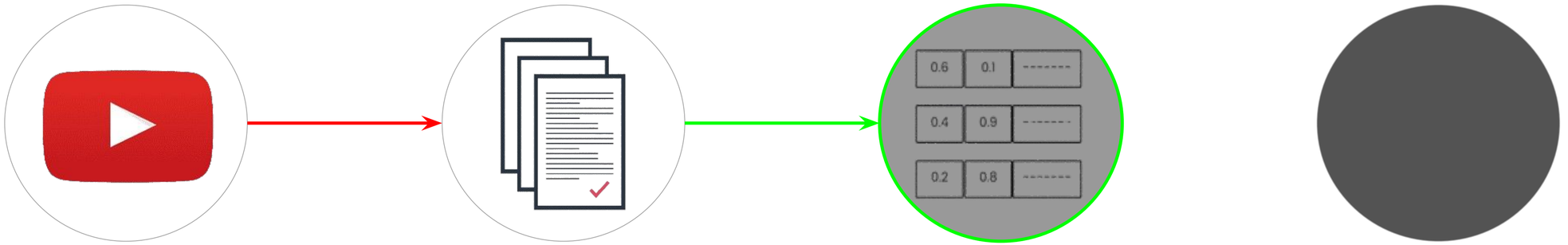




2 hours
Podcast

2 hours
Podcast

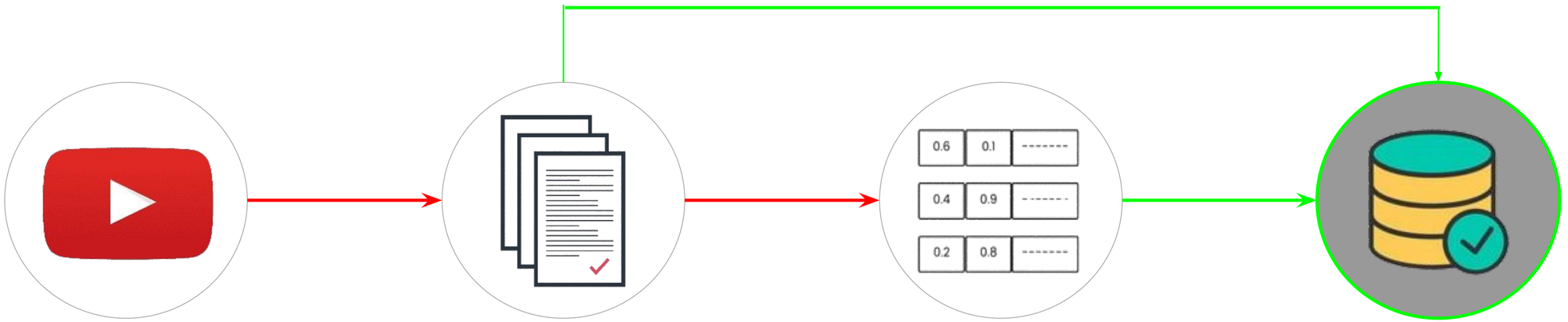
~ 25 chunks
of transcript




2 hours
Podcast

~ 25 chunks
of transcript

Convert the
Text to Vector
Qdrant FastEmbed
or OpenAI




2 hours
Podcast

~ 25 chunks
of transcript

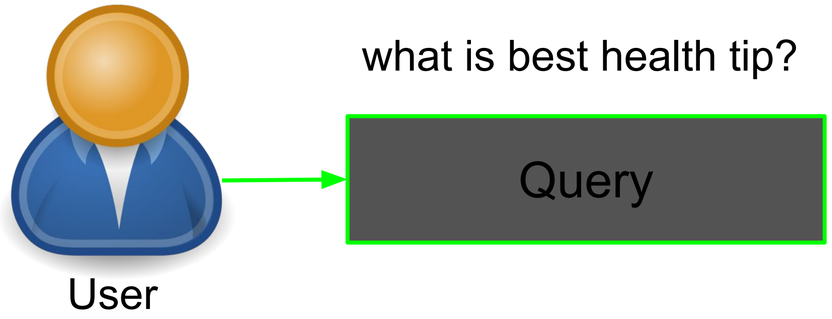
Convert the
Text to Vector
Qdrant FastEmbed

Index/Store
your document

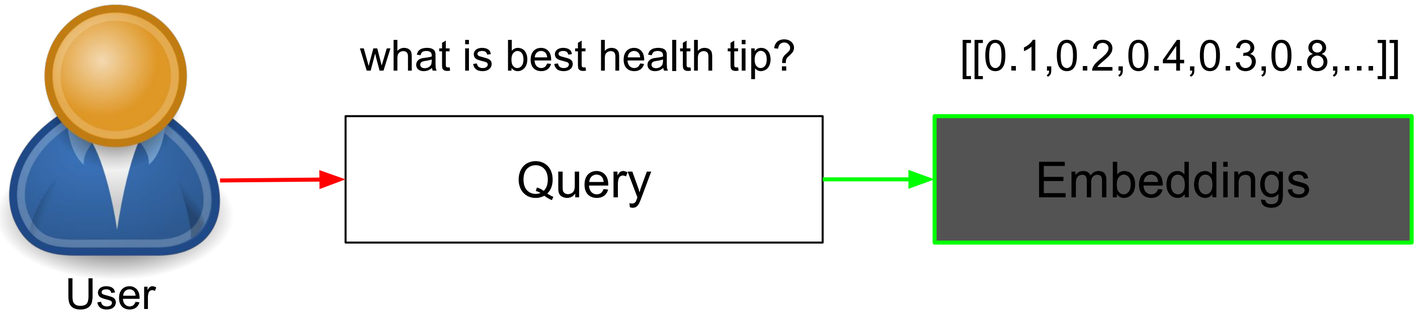
# R- Retrieval

User

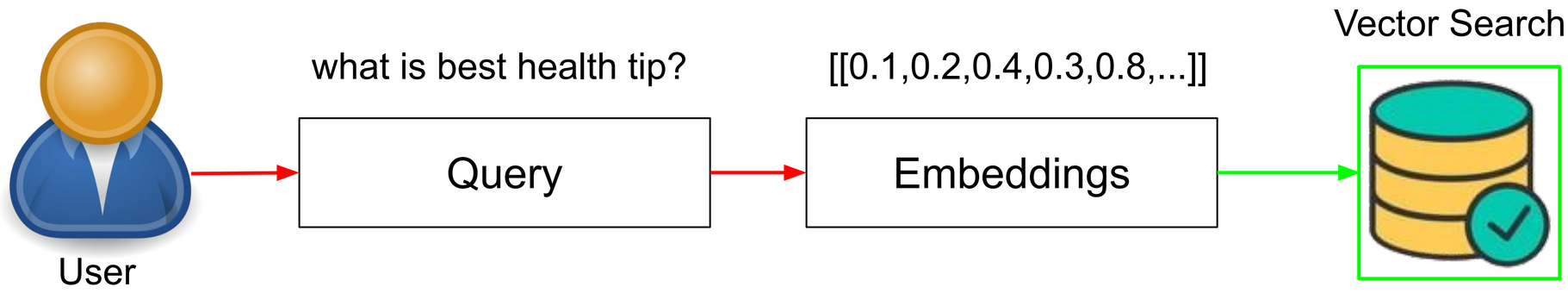
# R- Retrieval



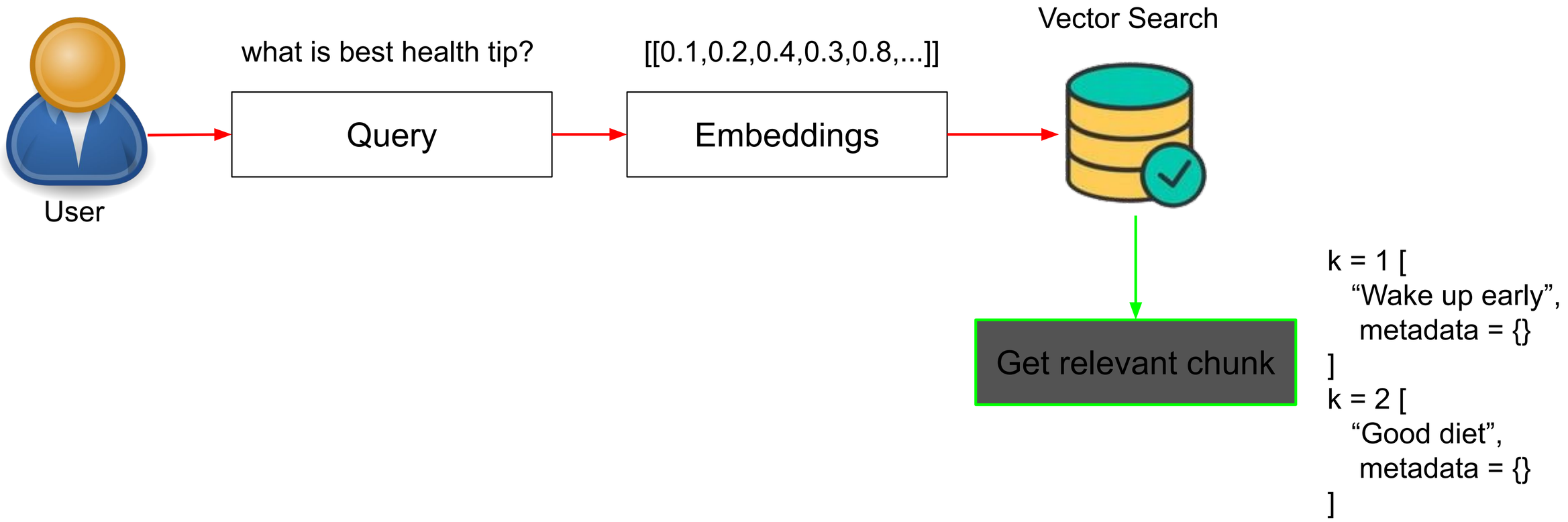
# R- Retrieval



# R- Retrieval


Vector Search
what is best health tip?
[[0.1,0.2,0.4,0.3,0.8,...]]
Query
Embeddings
User

# R- Retrieval

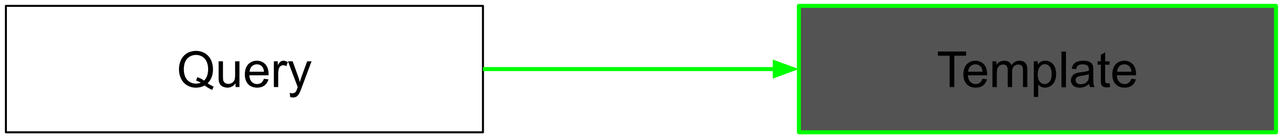

User
what is best health tip?
Query
[[0.1,0.2,0.4,0.3,0.8,...]]
Embeddings
Vector Search
Get relevant chunk
k = 1 [
“Wake up early”,
metadata = {}
]
k = 2 [
“Good diet”,
metadata = {}
]

# A- Augmented

Query

what is best health tip?

# A- Augmented




what is best health tip?

**<|system|>**{System Prompt}**<|end|>**
**<|user|>** {User Question} **<|end|>**
**<|assistant|>**

# A- Augmented

what is best health tip?

<|system|>{System Prompt}<|end|>
<|user|> {User Question} <|end|>
<|assistant|>

<|system|> You are an AI assistant <|end|>
<|user|> what is best health tip? <|end|>
<|assistant|>

# G- Generation

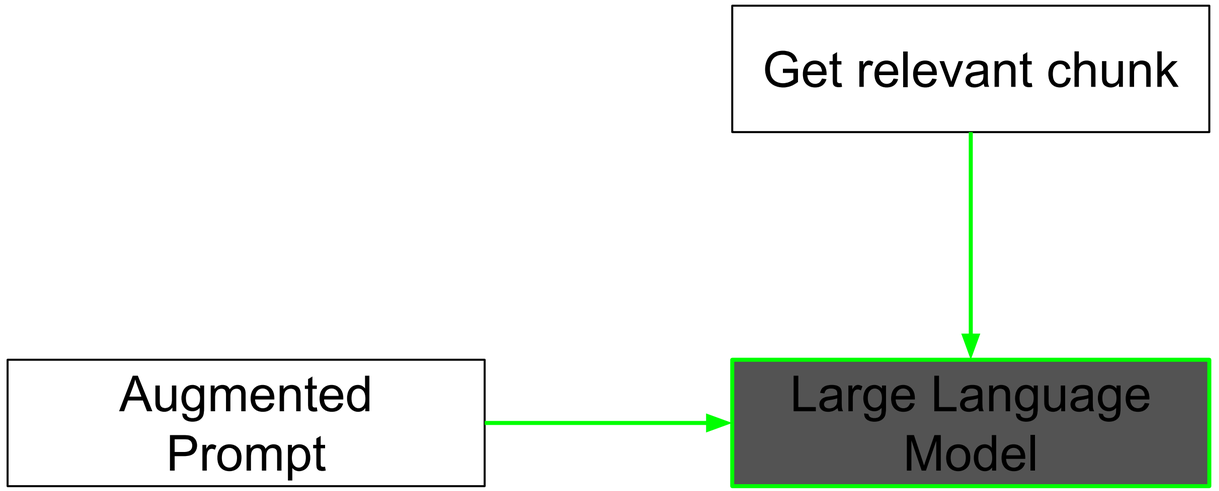
Get relevant chunk

k = 1 [ “Wake up early”,
metadata = {} ]
k = 2 [“Good diet”,
metadata = {} ]

Augmented Prompt

<|system|> You are an AI assistant <|end|>
<|user|> what is best health tip? <|end|>
<|assistant|>

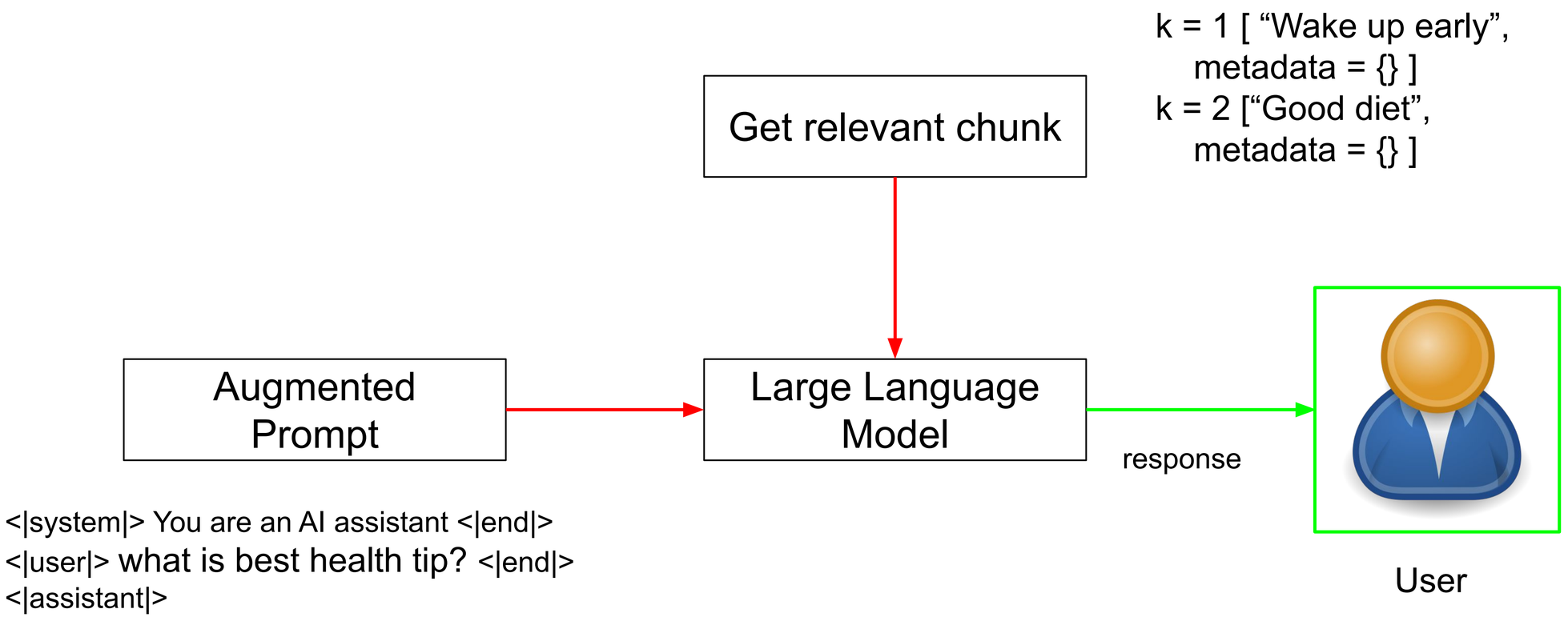
# G- Generation



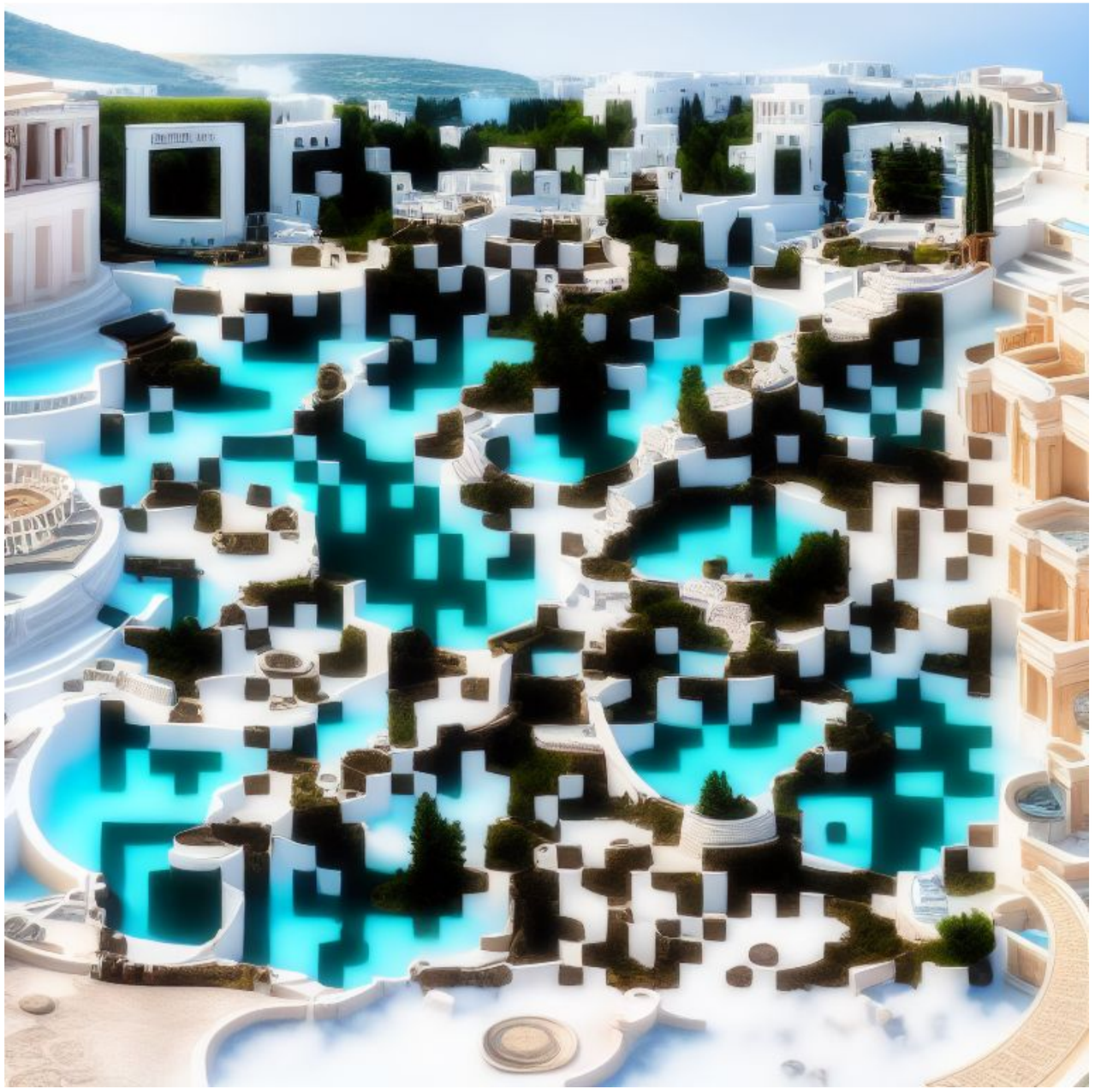

k = 1 [ “Wake up early”,
metadata = {} ]
k = 2 [“Good diet”,
metadata = {} ]

<|system|> You are an AI assistant <|end|>
<|user|> what is best health tip? <|end|>
<|assistant|>

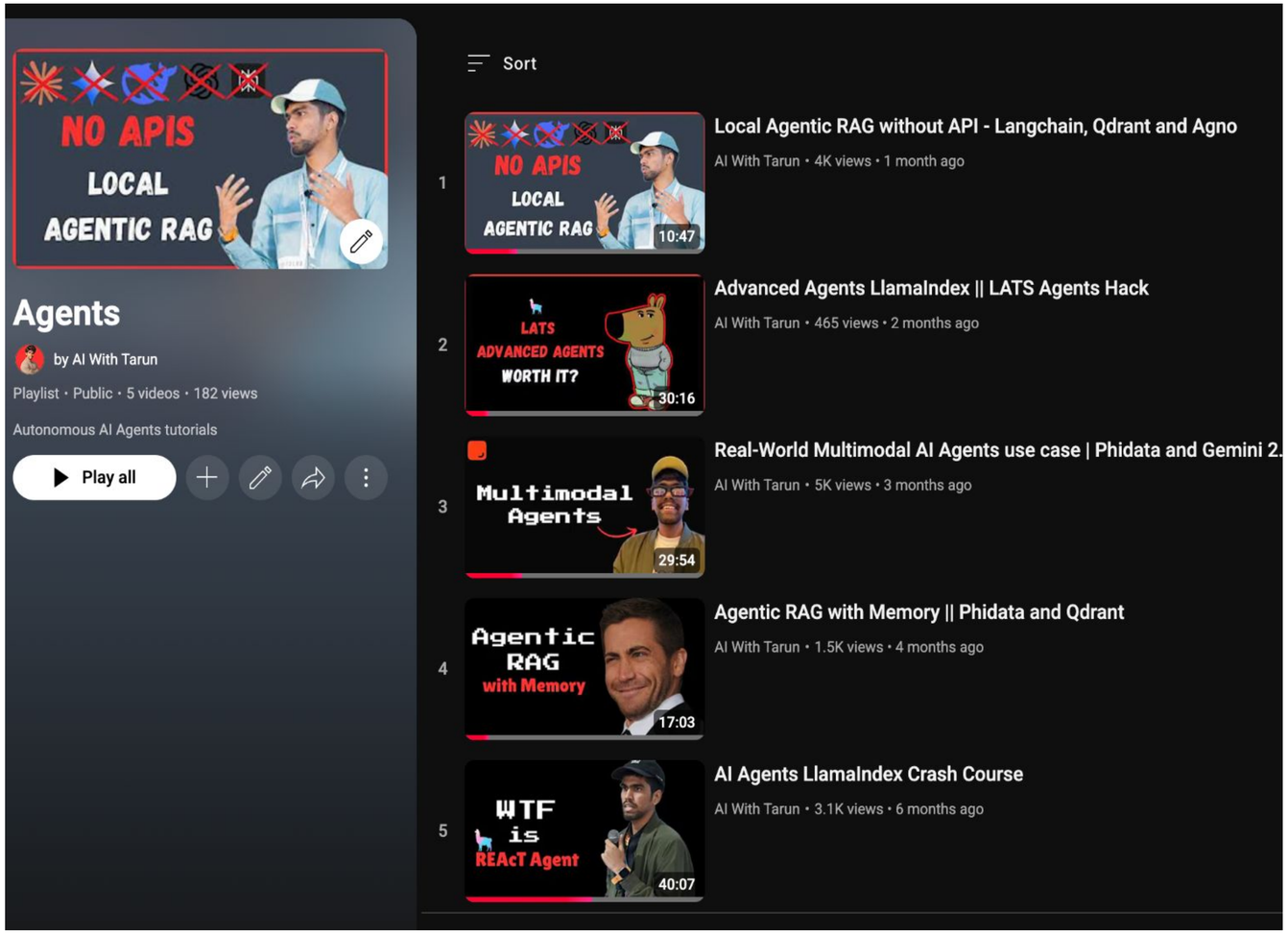
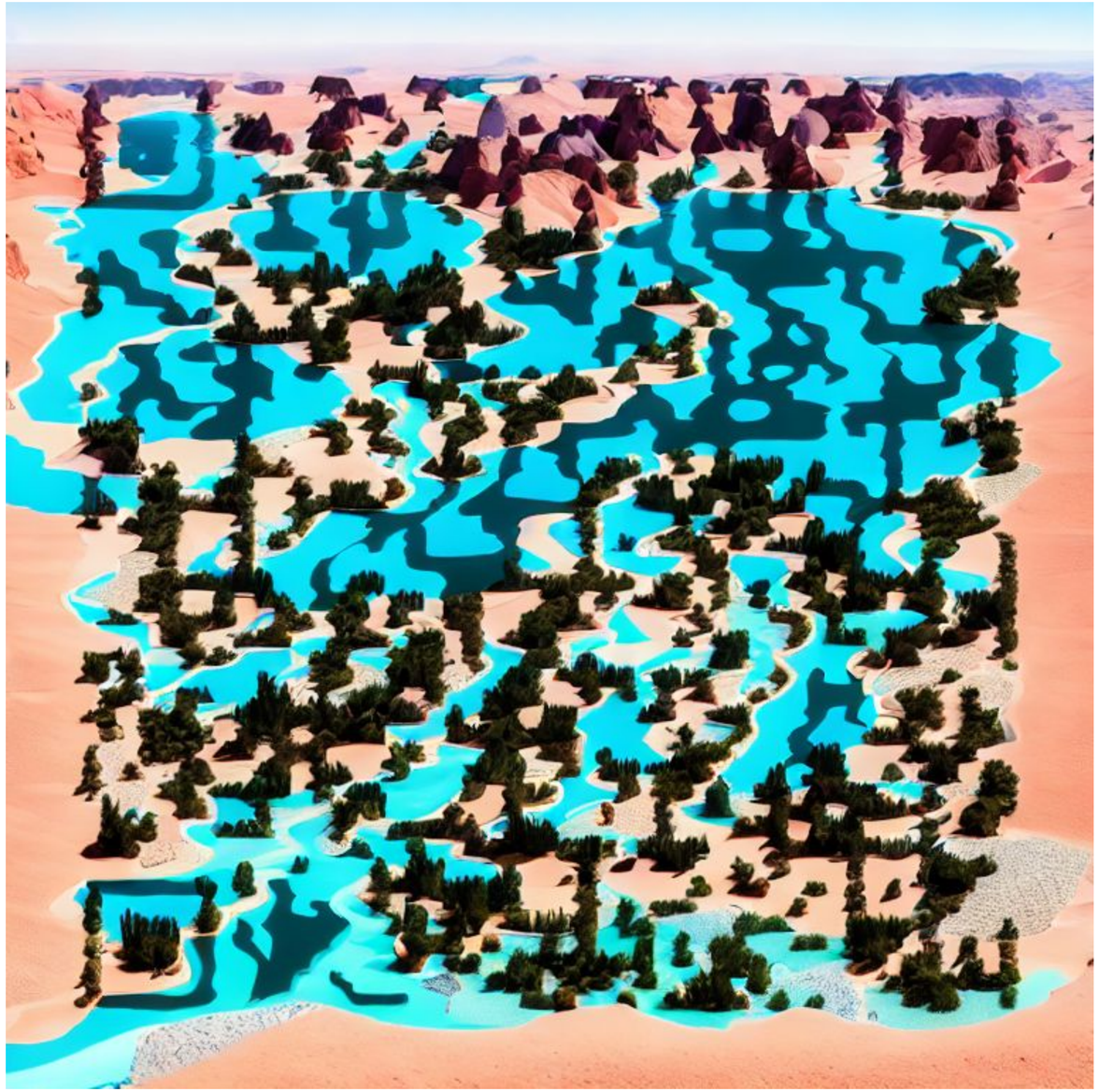
# G- Generation



Crash Course
Langchain
THE ONLY VIDEO YOU NEED



# Resource


NO APIS
LOCAL
AGENTIC RAG
Agents
by AI With Tarun
Playlist • Public • 5 videos • 182 views
Autonomous AI Agents tutorials
Play all
Sort
1
NO APIS
LOCAL
AGENTIC RAG
10:47
Local Agentic RAG without API - Langchain, Qdrant and Agno
AI With Tarun • 4K views • 1 month ago
2
LATS
ADVANCED AGENTS
WORTH IT?
30:16
Advanced Agents LlamaIndex || LATS Agents Hack
AI With Tarun • 465 views • 2 months ago
3
Multimodal
Agents
29:54
Real-World Multimodal AI Agents use case | Phidata and Gemini 2.
AI With Tarun • 5K views • 3 months ago
4
Agentic
RAG
with Memory
17:03
Agentic RAG with Memory || Phidata and Qdrant
AI With Tarun • 1.5K views • 4 months ago
5
WTF
is
REAcT Agent
40:07
AI Agents LlamaIndex Crash Course
AI With Tarun • 3.1K views • 6 months ago