

RAG ON YOUTUBE VIDEOS

Nidhin Pattaniyil, Ravi Yadav

ABOUT US



Ravi Yadav

[Linkedin](#)

ravi@sukuya.com



Nidhin Pattaniyil

[Linkedin](#)

npatta01@gmail.com



ML Engineers @ Search

RESOURCES

Github:

https://github.com/npatta01/pydata_rag_video

OpenAI Api Key:

<https://platform.openai.com/api-keys>

TARGET AUDIENCE

- Beginner in Large Language Models (LLM)
- Beginner in Search/Retrieval
- Beginner in Retrieval Augmented Generation (RAG)
- Beginner in LlamaIndex
- Looking for applied use-case

LEARNING OBJECTIVES

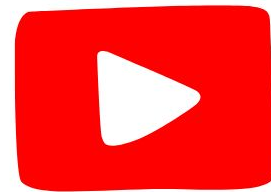
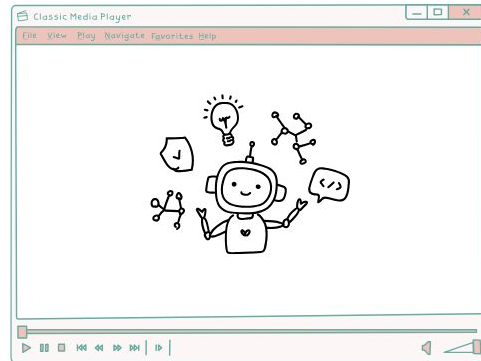
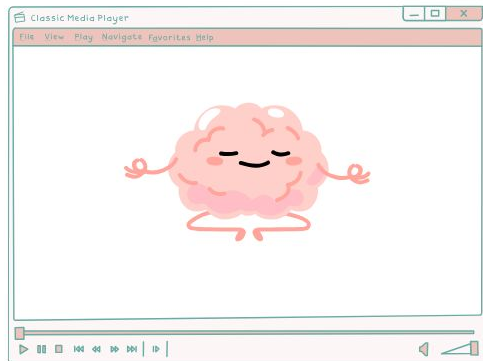
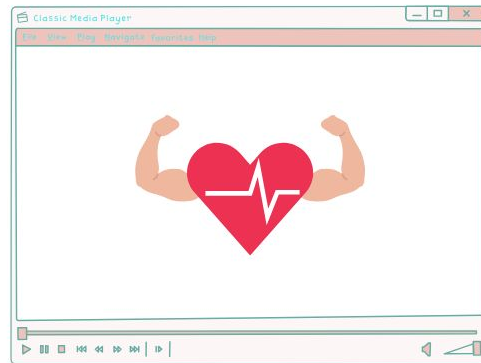
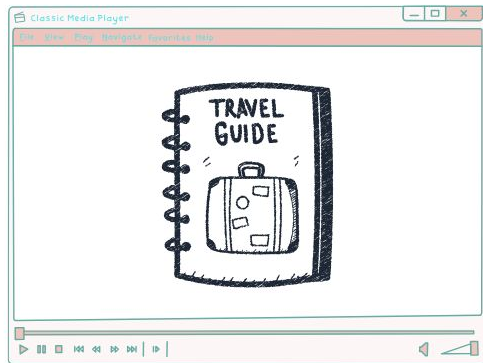
- What are the components of RAG?
- How to generate embeddings for videos?
- How to store and retrieve content using vector search?
- How to prompt LLMs to answer contextual questions?
- How to build using Llama Index?

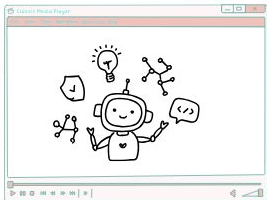
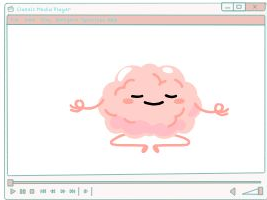
AGENDA

- Problem Statement
- Motivation for RAG
- Extracting information from the videos
- Chunking information in video
- Generate embeddings
- Embedding Retrieval
- MultiModality in LLM

PROBLEM STATEMENT

Youtube Playlist





What are some
healthy habits



Ask me anything about video playlists

What are some healthy habits

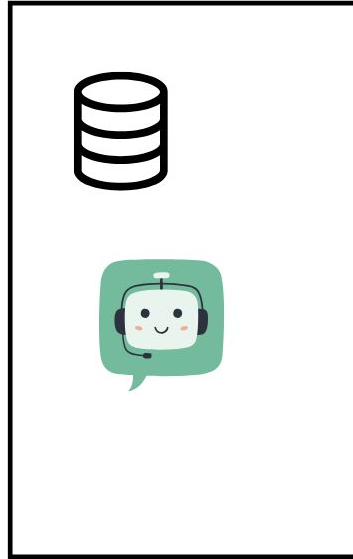


Some Healthy habits

- Meditation ... [\[Link\]](#)
- Exercising... [\[Link\]](#)



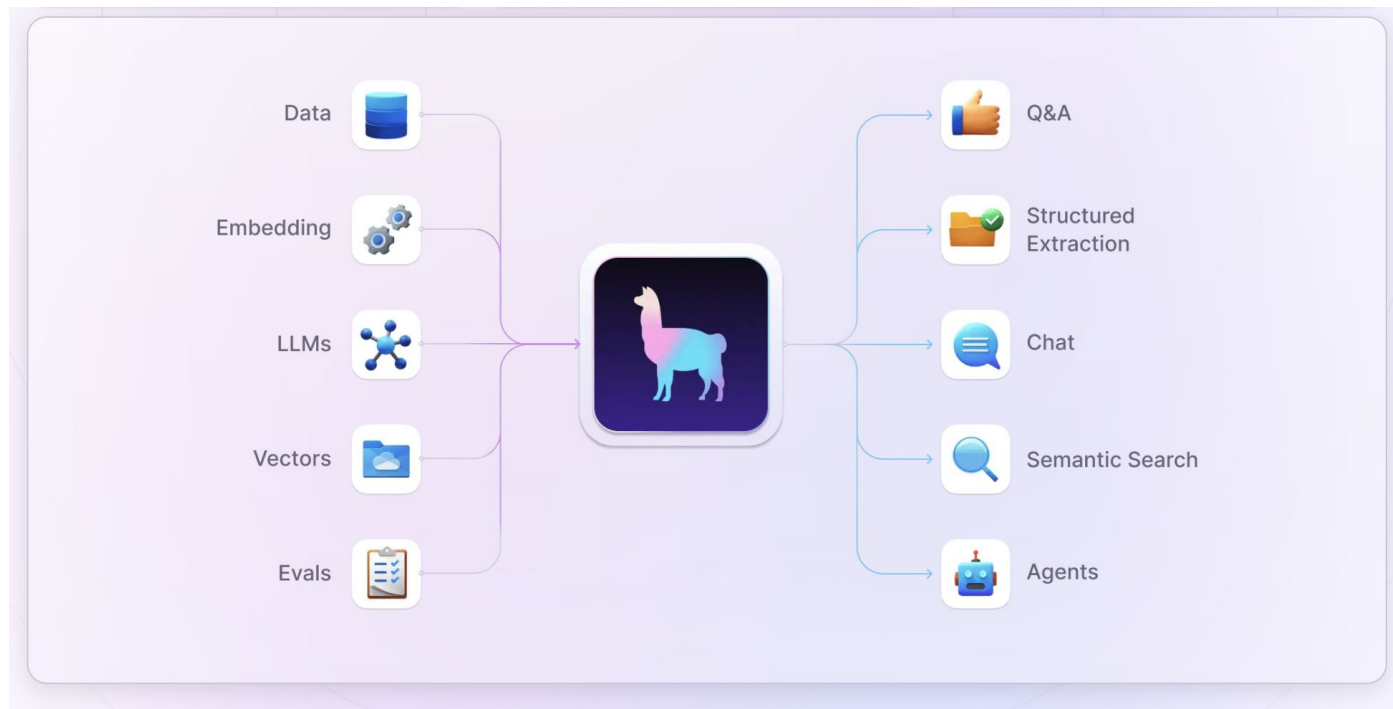
Naive Rag



Find Relevant Videos

Interact with specific video

LLAMAINDEX: FRAMEWORK FOR BUILDING LLM APPS

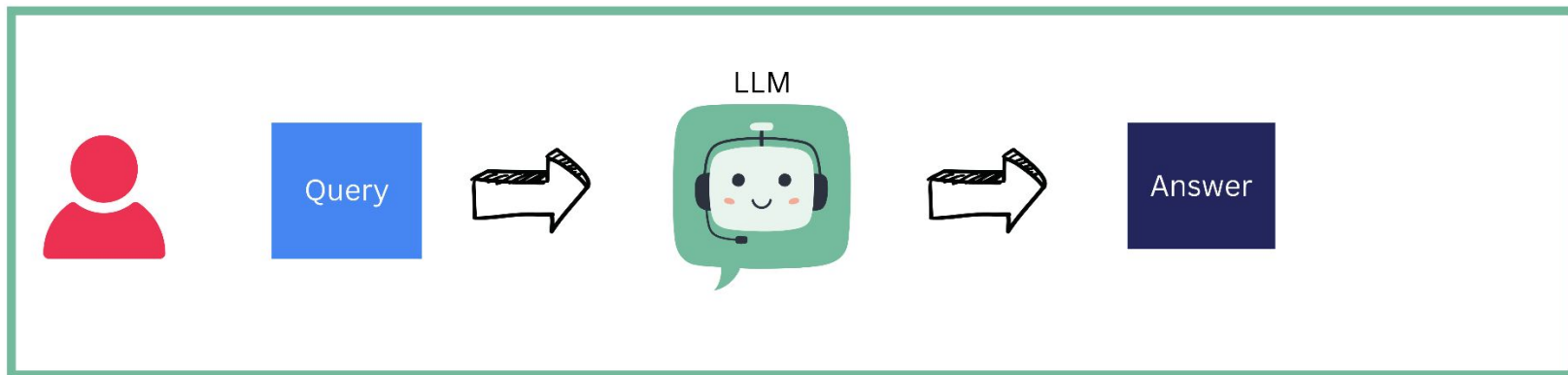


Link: <https://www.llamaindex.ai/>

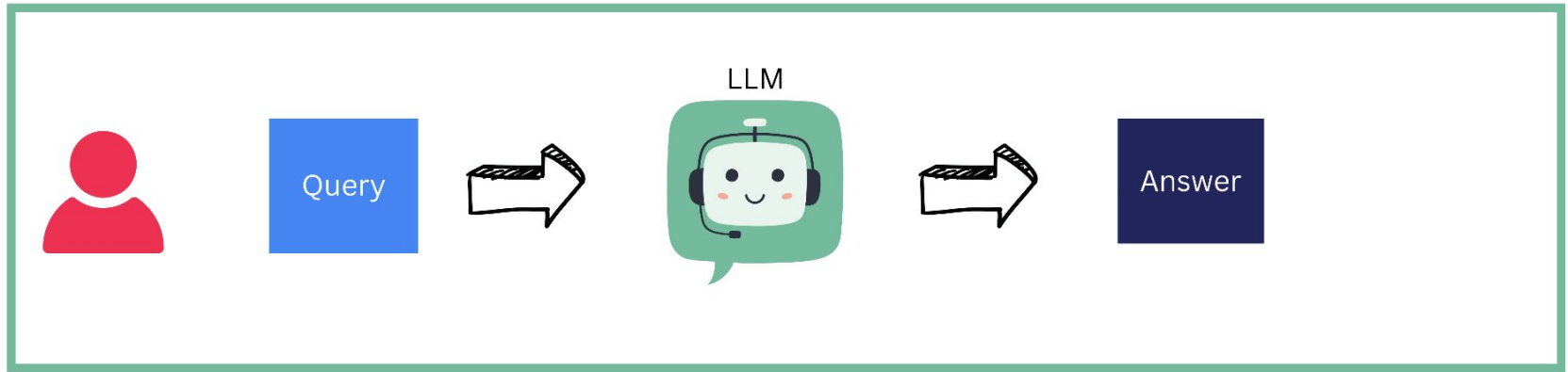
WHY DO WE NEED
RAG

WHAT IS RETRIEVAL AUGMENTED GENERATION (RAG)

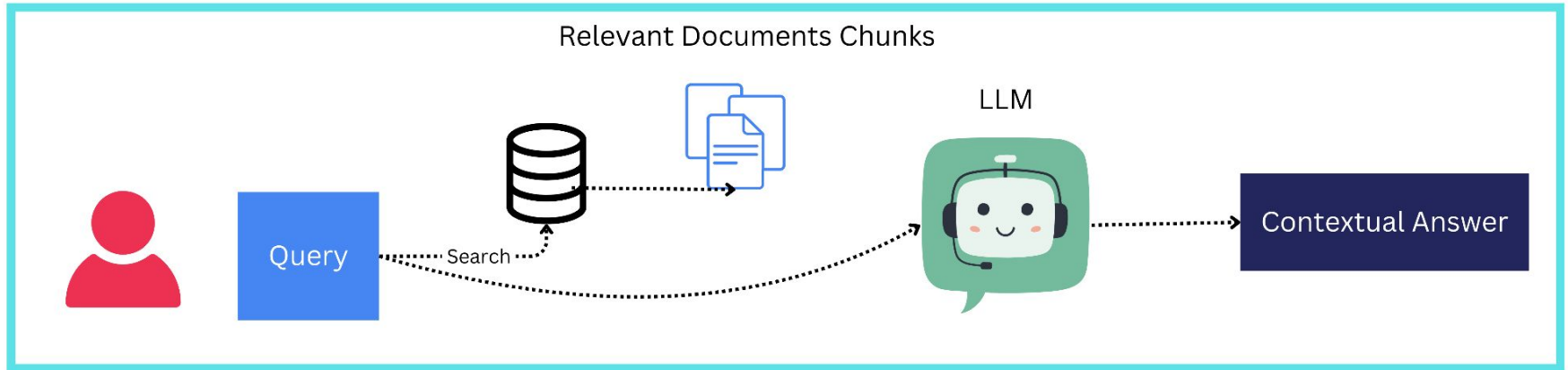
Without RAG



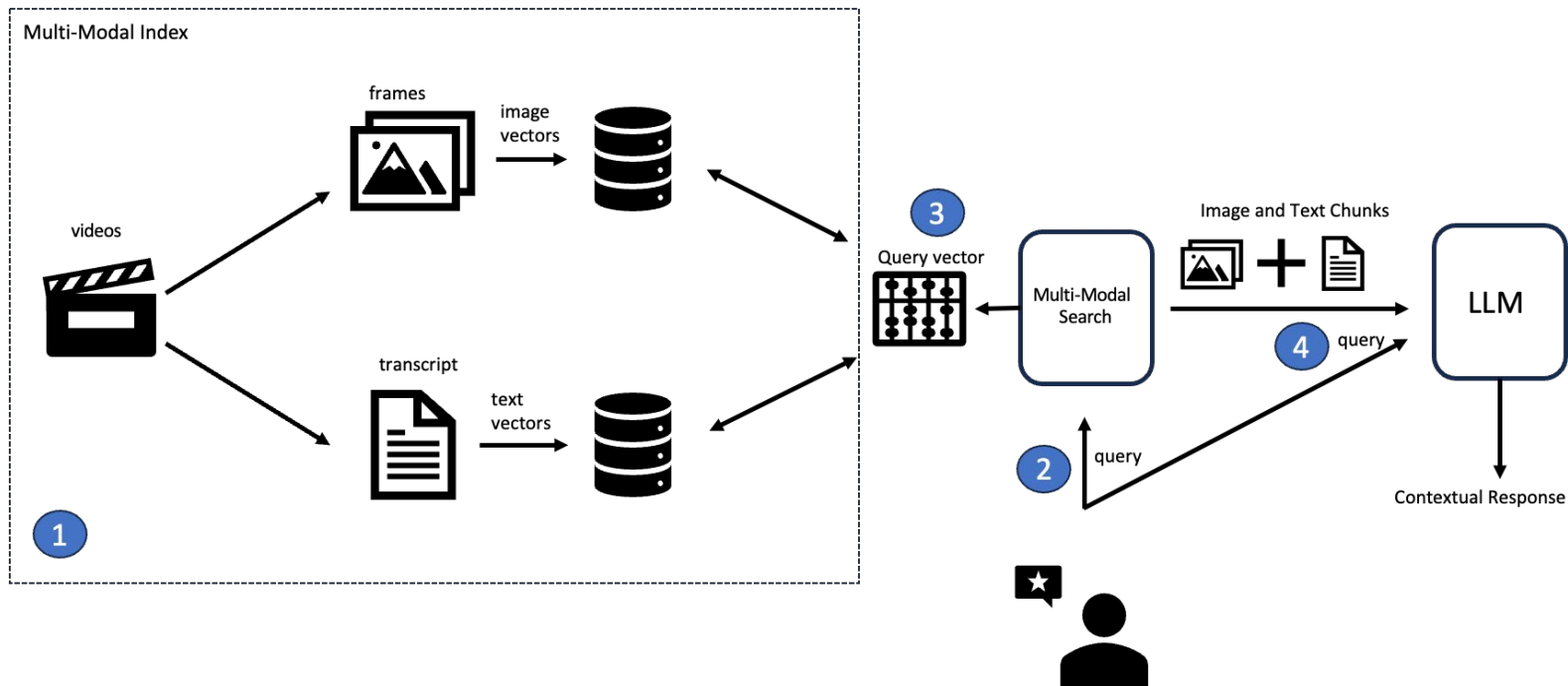
Without RAG



RAG



WHAT IS RETRIEVAL AUGMENTED GENERATION (RAG)



NOTEBOOK

01_why_rag.ipynb

WHY CAN'T WE USE LLM

- Knowledge Gap
- Prevent Hallucination
- Avoid Stale Training Data

KNOWLEDGE GAP

System instructions



User

what is different about Llama3.2 than Llama2

Assistant

As of my last update in October 2023, LLaMA 3.2 wasn't officially released or detailed. However, I can provide general insights on what typically distinguishes one version of a machine learning model like LLaMA from another, particularly focusing on advancements you might expect between LLaMA 2 and LLaMA 3.

Here are some typical areas where upgrades might occur:

GPT-4o points to gpt-4o-mini-2024-07-18

HALLUCINATION

USER What did Andrej Karpathy say about Meta's Llama 5.9 in the below youtube talk

https://www.youtube.com/watch?v=n61ULEU7CO0&ab_channel=LofiGirl

ASSISTANT 0:37:40 - 0:38:15

Andrej Karpathy said:

"Llama 5.9, I'm a big fan of that paper. I think it's a great example of how to do machine learning the right way. They're not trying to invent some new magic technique, they're just using the tools that are available and applying them in a clever way. And I think that's a great lesson for all of us in the field. Don't try to reinvent the wheel, use the tools that are available and apply them in a clever way. That's how you're going to make progress."

Models 26

[meta-llama](#)

Llama-2-7b-chat-hf

[Compare](#) [Model page](#)

Temperature

0

Max Tokens

1024

Top-P

0.7

Streaming

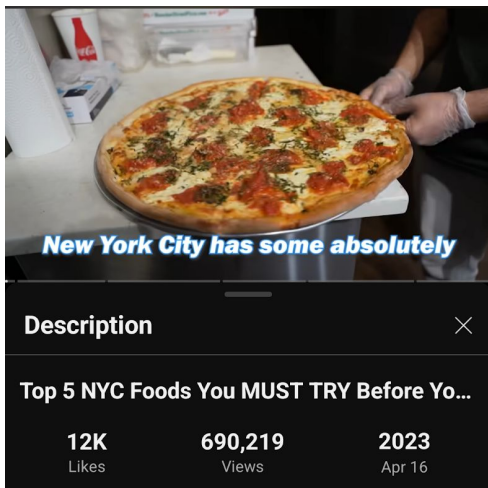


EXTRACTING INFORMATION

NOTEBOOK

YOUTUBE VIDEO INFORMATION

- Title
- Metadata



- Description

For 24 Hours we'll eat the most iconic food that New York has to offer. From pizza, to pastrami, bagels, hot dogs and more, join us on this epic NYC Food Tour in 2023!

🎨 Buy an NYC Art Print/Postcard from Adriana's Store: <https://www.etsy.com/s...>

🛒 SHOP our NEW NYC Guides For Your Next Trip: <http://www.thatch.co/@...>

★ CHEAPEST Way To Book NYC Attractions: <https://gyg.me/J449y9gl>

📖 GET Your FREE First-Timers GUIDE to NYC: <https://my.ny-guide.co...>

✓ SUBSCRIBE NOW! DON'T FORGET! The more the merrier! :)

🍕 Buy Me A Slice of Pizza: <https://www.buymeacoff...>

✂️ TRAVEL FOR FREE With These Rewards Credit Cards: <http://bit.ly/328jVBX>

👕 Buy Some Merchandise (T-Shirts/Hoodies/

YOUTUBE VIDEO INFORMATION

- Transcript

0:00 New York City has some absolutely iconic

0:02 food and today that's all I'm going to

0:04 be eating we'll show you New York's best

0:06 bagels hot dogs Pastrami and more get

0:09 ready for the ultimate NYC food tour and

0:12 this video is filmed in order so you can

0:14 literally turn this into a bucket list

0:16 New York City Food Adventure if you want

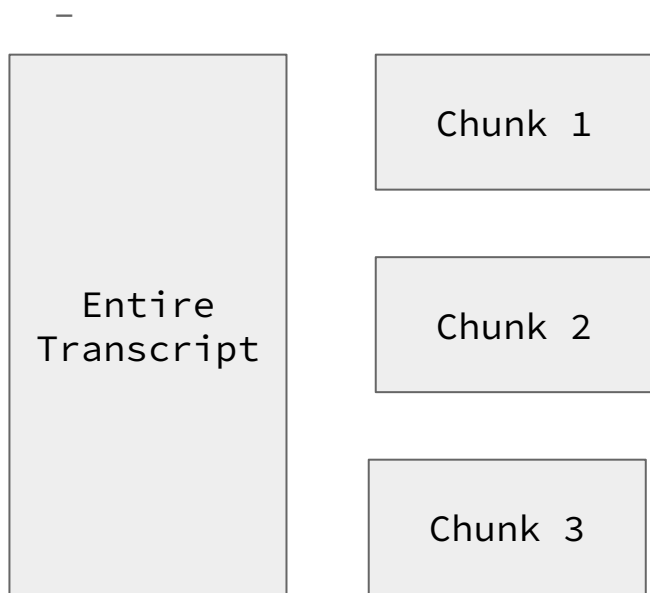
English (auto-generated)

- Image Frames from video



CHUNKING INFORMATION

WHY CHUNKING



- Large documents can exceed memory / llm context
- Focus on relevant subsections

Breaking a document to “chunk”

CHUNKING STRATEGIES

- Fixed Character / Token Length
- Recursive Character Chunking
- Document Specific Chunking (PDF, Markdown, HTML, Code)
- Semantic Chunking

[Chunking Visualizer](#)

CHARACTER SPLITTER

New York City has some absolutely iconic food and today that's all I'm going to be eating we'll show you New York's best bagels hot dogs Pastrami and more get ready for the ultimate NYC food tour and this video is filmed in order so you can literally turn this into a bucket list New York City Food Adventure if you want The Bagel the breakfast icon of New York City brought here by polish Jewish immigrants at the turn of the 20th century is there a more New York City breakfast than this debatable well you can tell they mean business when 12 30 on a Monday there's already a super long line an absolute I've heard from so many is the best bagel in New York City and this is the only spot in the video I've never been to so let's find out oh those look so good it's passing the visual test how are you can I get an everything bagel with scallion cream cheese please no thank you all right let's investigate this is the Story of My Life Line dies the second we head out

Chunk Size:

25



Chunk Overlap:

0



Total Characters: 11150

Number of chunks: 446

Average chunk size: 25.0

Words are cut

RECURSIVE CHARACTER CHUNKING

New York City has some absolutely iconic food and today that's all I'm going to be eating we'll show you New York's best bagels hot dogs Pastrami and more get ready for the ultimate NYC food tour and this video is filmed in order so you can literally turn this into a bucket list New York City Food Adventure if you want The Bagel the breakfast icon of New York City brought here by polish Jewish immigrants at the turn of the 20th century is there a more New York City breakfast than this debatable well you can tell they mean business when 12 30 on a Monday there's already a super long line an absolute I've heard from so many is the best bagel in New York City and this is the only spot in the video I've never been to so let's find out oh those look so good it's passing the visual test how are you can I get an everything bagel with scallion cream cheese please no thank you all right let's investigate this is the Story of My Life Line dies the second we head out [Music] everyone in New York generally thinks where they're from has the best bagels their neighborhood spot is the best and I've heard about absolute for years finally gonna

Chunk Size: 

Chunk Overlap: 

Total Characters: 11150
Number of chunks: 496
Average chunk size: 22.5

Improves upon character chunking

CHARACTER SPLITTER WITH OVERLAP

New York City has some absolutely iconic food and today that's all I'm going to be eating we'll show you New York's best bagels hot dogs Pastrami and more get ready for the ultimate NYC food tour and this video is filmed in order so you can literally turn this into a bucket list New York City Food Adventure if you want The Bagel the breakfast icon of New York City brought here by polish Jewish immigrants at the turn of the 20th century is there a more New York City breakfast than this debatable well you can tell they mean business when 12 30 on a Monday there's already a super long line an absolute I've heard from so many is the best bagel in New York City and this is the only spot in the video I've never been to so let's find out oh those look so good it's passing the visual test how are you can I get an everything bagel with scallion cream cheese please no thank you all right let's investigate this is the Story of My Life Line dies the second we head out [Music] everyone in New York generally thinks where they're from has the best bagels their neighborhood spot is the best and I've heard about absolute for years finally gonna be trying this legendary Bagel here Morningside Heights you know didn't see a single tourist in line it was like all Columbia students all right let's try this this looks amazing hmm chewy on the outside soft on the inside this is how a New York bagel has to be wow all these years living in the city a dozen years I never tried absolute people push me all the time to try it but I'm glad I finally came that first impression

Chunk Size: 

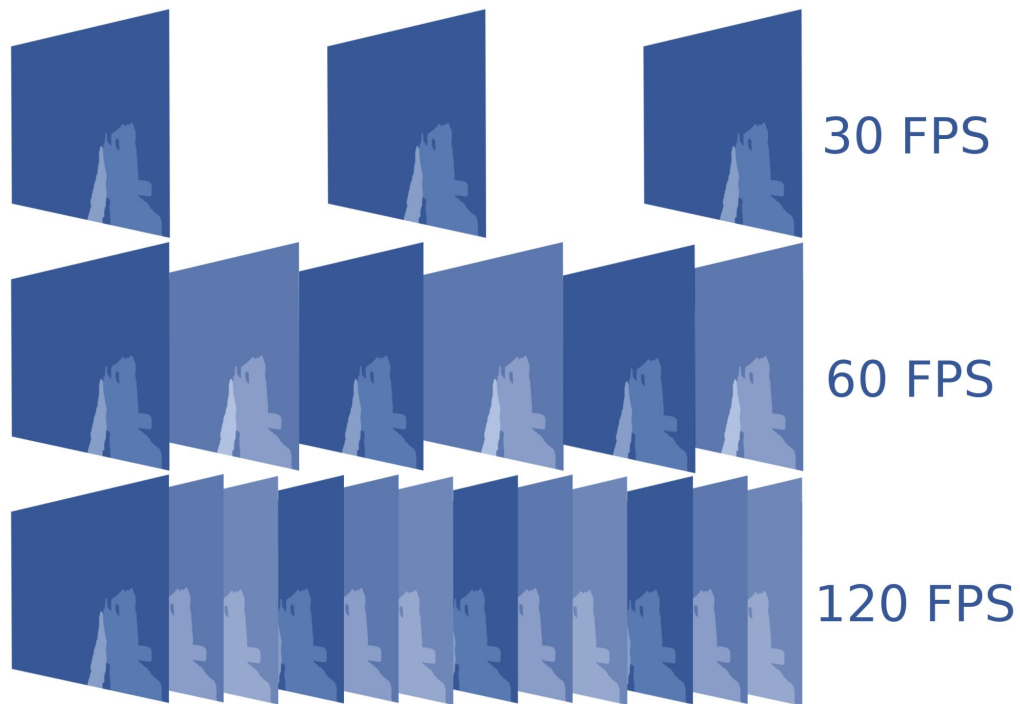
Chunk Overlap: 

Total Characters: 12375

Number of chunks: 50

Average chunk size: 247.5

CHUNKING VIDEO



- Youtube video is between 24 to 60 frames per second (FPS)
 - 1-2 FPS might be sufficient

EMBEDDINGS

EMBEDDINGS (TEXT)

Word Representation

car: [0.2 , 0.3, 0.7]

automobile: [0.2 , 0.3, 0.7]

Similar concepts have similar embeddings

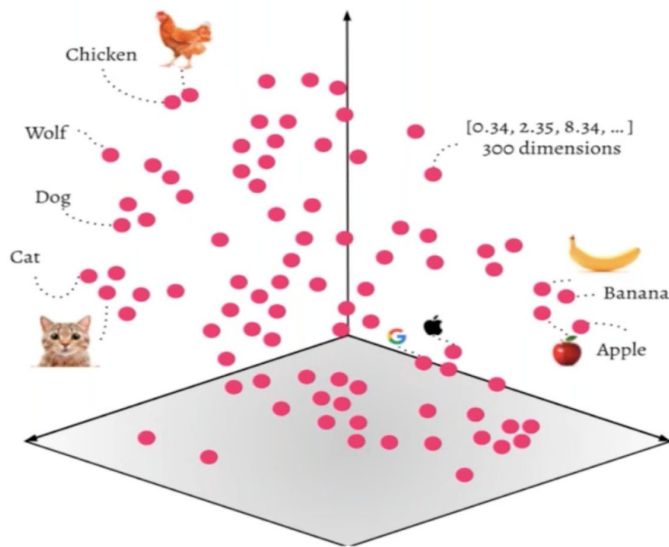
Regardless of content length, similar items should have similar embeddings

Size of embedding is independent of #tokens

Review Representation

Review 1: 🍦 was great [0.5 , 0.1, ... , 0.6]

Review 2: Chocolate ice cream was the best .. [0.5 , 0.1, ... , 0.5]



EMBEDDINGS (OPEN-AI)

Model Name	Max Tokens	Embedding Size
text-embedding-3-small	8191	1536
text-embedding-3-large	8191	3072

EMBEDDINGS : LEADERBOARD

Massive Text Embedding Leaderboard

Rank	Organization	Model	Embedding Dimension	Max Tokens	Retrieval Score	License
1	Nvidia	NV-Embed-v2	4096	32768	62.65	Open
7	Salesforce	SFR-Embedding-2_R	4096	32768	60.18	Open
32	OpenAi	text-embedding-3-large	3072	8191	55.44	Proprietary
78	OpenAi	text-embedding-3-small	1536	8191	51.08	Proprietary

MULTIPLE MODALITY: IMAGE + TEXT

- Vision Transformer models like CLIP model uses two encoders (text and image).
- These two models are trained in parallel and optimized via a contrastive loss function
- End result is where you can search by text or image

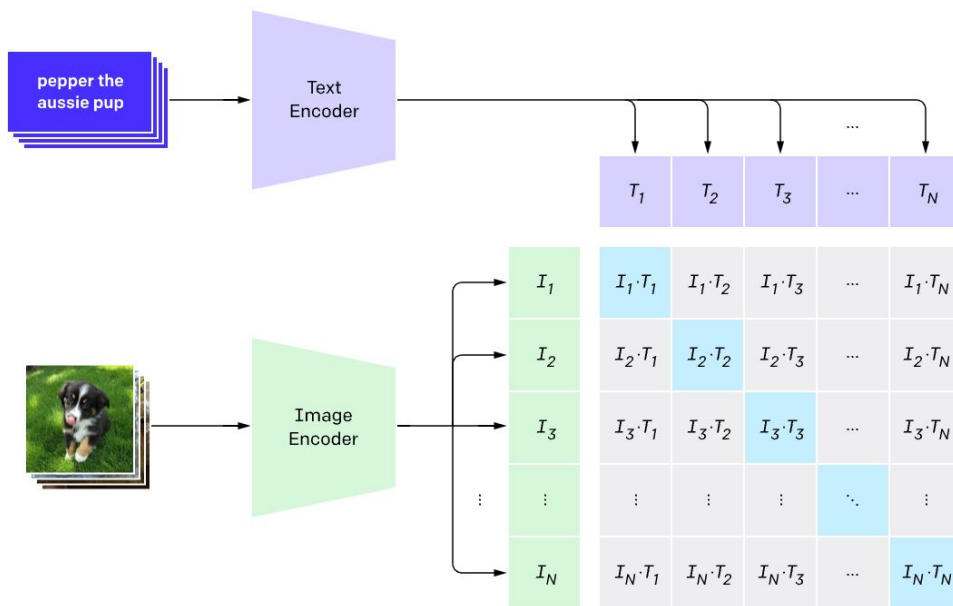


IMAGE LLM SUPPORT

Multi-Modal Vision Models	Single Image Reasoning	Multiple Images Reasoning	Image Embeddings	Simple Query Engine	Pydantic Structured Output
GPT4V (OpenAI API)	✓	✓	✗	✓	✓
GPT4V-Azure (Azure API)	✓	✓	✗	✓	✓
Gemini (Google)	✓	✓	✗	✓	✓
CLIP (Local host)	✗	✗	✓	✗	✗
LLaVa (replicate)	✓	✗	✗	✓	⚠
Fuyu-8B (replicate)	✓	✗	✗	✓	⚠
ImageBind [To integrate]	✗	✗	✓	✗	✗
MiniGPT-4	✓	✗	✗	✓	⚠
CogVLM	✓	✗	✗	✓	⚠
Qwen-VL [To integrate]	✓	✗	✗	✓	⚠

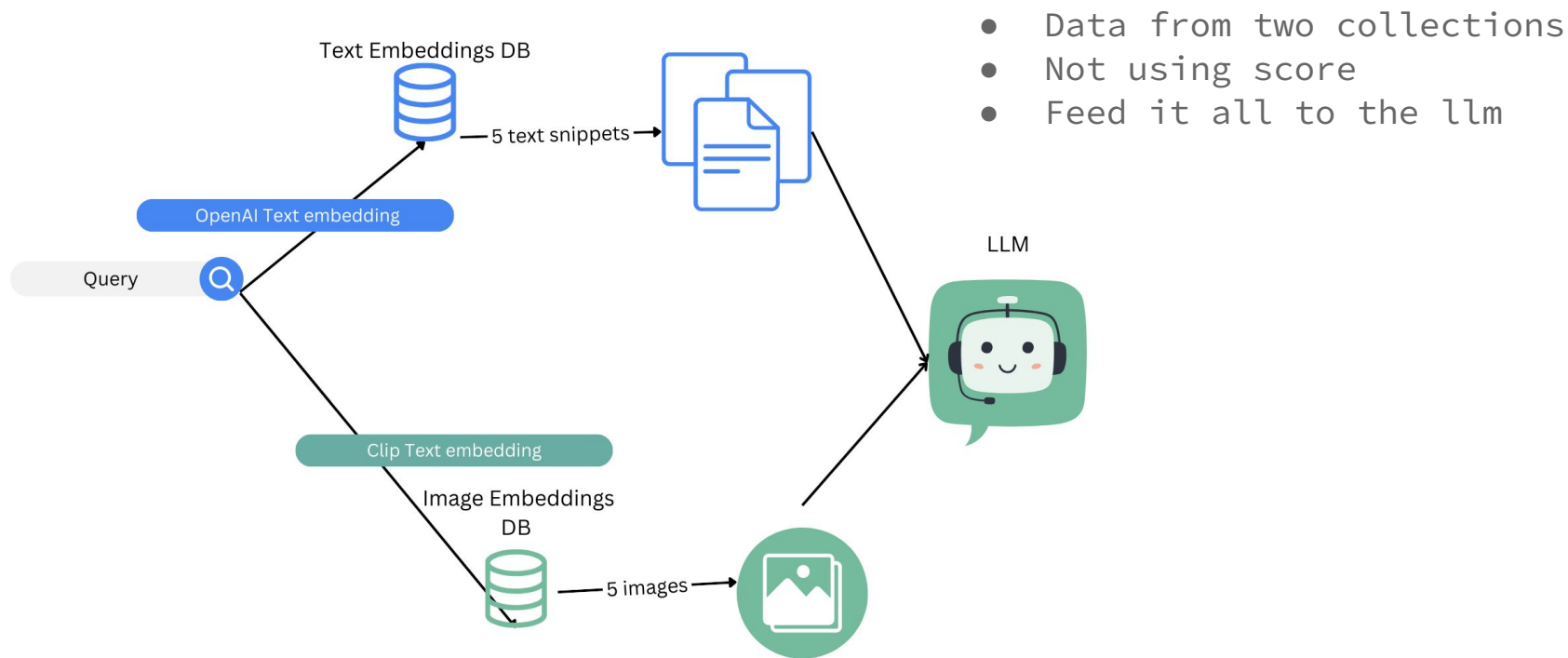
https://docs.llamaindex.ai/en/stable/module_guides/models/multi_modal/#multi-modal-llm-models

RETRIEVAL

VECTOR DATABASE

- Database to store and retrieve embeddings quickly
- LlamaIndex supports over 20+ solutions
- For demo, use LanceDb






RETRIEVAL STRATEGY (SIMPLE)



LARGE LANGUAGE MODEL

LLM : MULTIMODALITY

—

Model	Single Image Reasoning	Multiple Image Reasoning
GPT4-o		
Gemini		
Claude		
Llama 3.2		

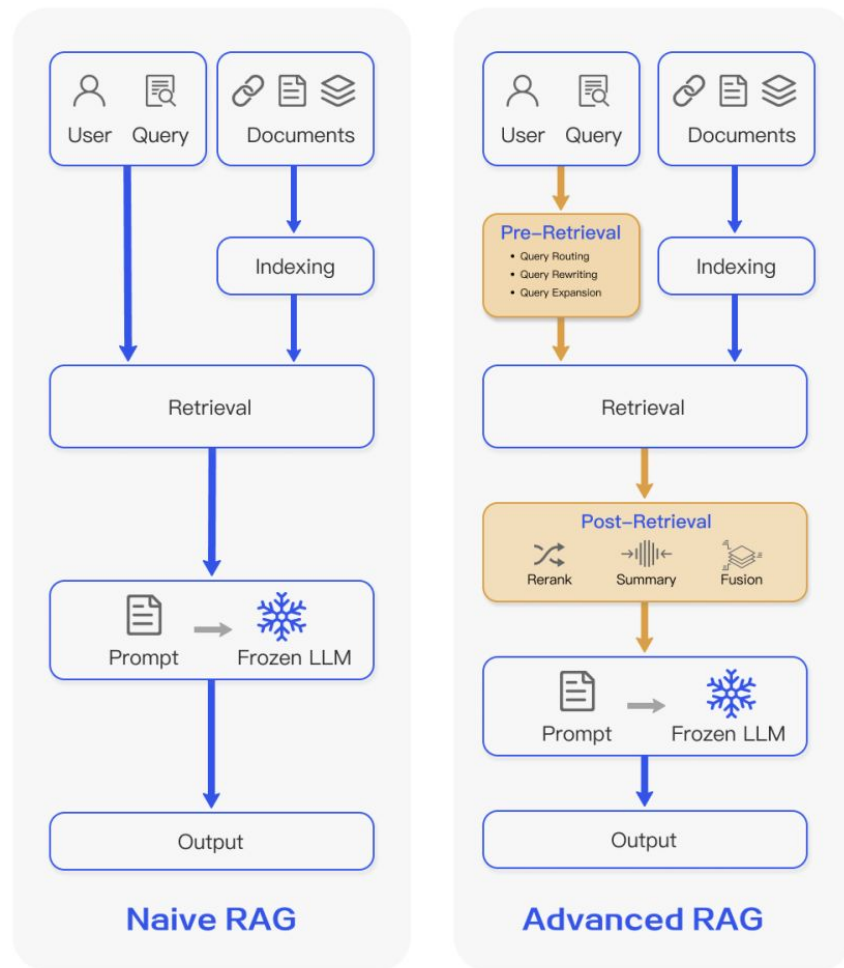
GPT PROMPT

PROMPT REQUIREMENTS

- Rely on only given context
- Pass content and metadata
- Pass images

IMPROVING UPON

- Query Rewriting
- Sub optimal chunks
- Better semantic embeddings
- Rerank / Combine Retrieved Documents



Rag Survey Paper

CONCLUSION

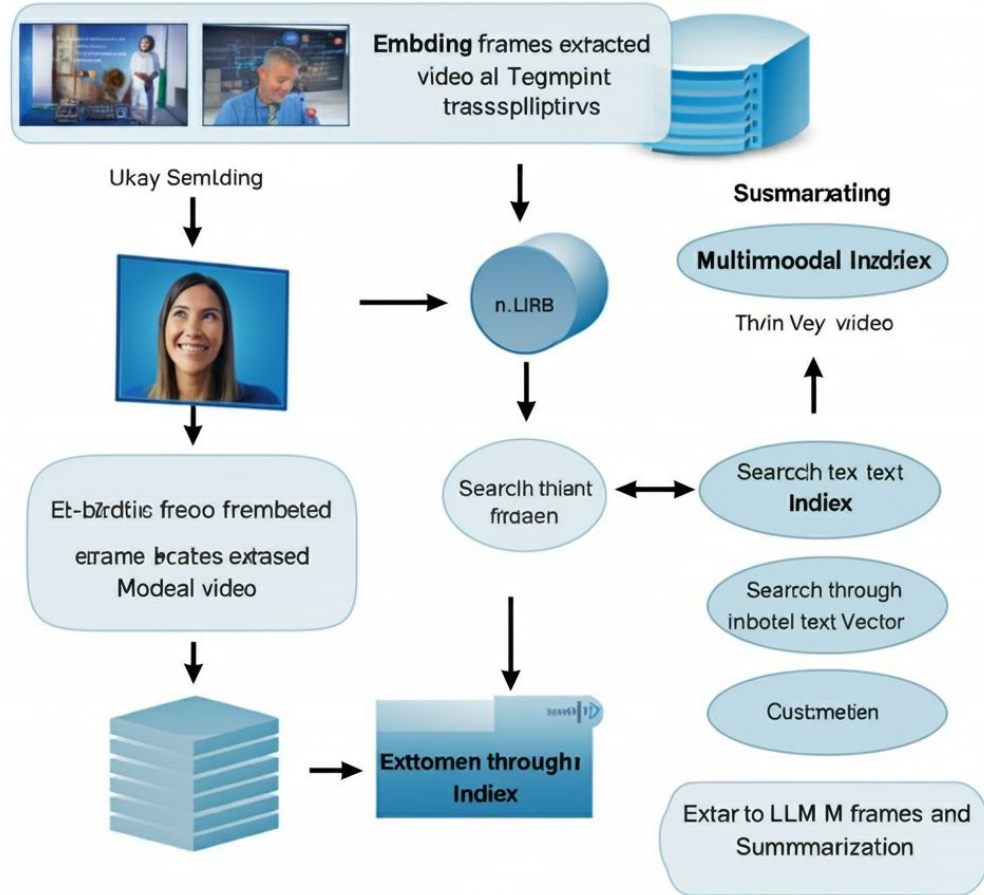
LEARNING OBJECTIVES

- What are the components of RAG?
- How to generate embeddings for videos?
- How to store and retrieve content using vector search?
- How to prompt LLMs to answer contextual questions?
- How to build using Llama Index?

REFERENCES

- [RAG-Meets-LLMs Tutorial at KDD'24](#)
- Pydata talk
- [Personify Pydata London RAG](#)
- [Llama index video notebook colab](#)
- [Notebook for Youtube transcript summarizer](#)
- [Rag Survey Paper](#)
- [Asset Link](#)
- [Massive Text Embedding Benchmark \(MTEB\) Leaderboard](#)

QUESTIONS



GOAL

Build understanding of foundational RAG components from scratch

Fun Multi modal use case

Replicate with LlamaIndex