

RAG - Retrieval Augmented Generation

LLMs are trained only with data and LLMs are limited by the recency of data. How good the data is, that good is only LLM.

Also, let's say we have a chatbot and it has to respond only from the data we have with us, not something very generic

User ---> sends request ---> LLM

LLM is trained on some generic internet data but we want LLM to respond based on our data only

1. One option is to fine-tune LLM to make it respond in the way we want. We can train LLM more to respond to our data
2. Second option is send the entire dataset for every request but the problem is, we can't keep sending TBs of data for every request, it could consume bandwidth, computation, token cost. Most important is cost based on the tokens we are consuming from the data
3. Third option is, RAG. Reference documents could be PDF or docs, you don't want to send the entire private data to LLM (that's insecure)

PDFs are converted into multiple chunks, then Embeddings and stored in VectorStore. When user sends a query it goes to VectorStore. VectorStore returns matches. So the User query + matches go into LLM and LLM responds to User finally.

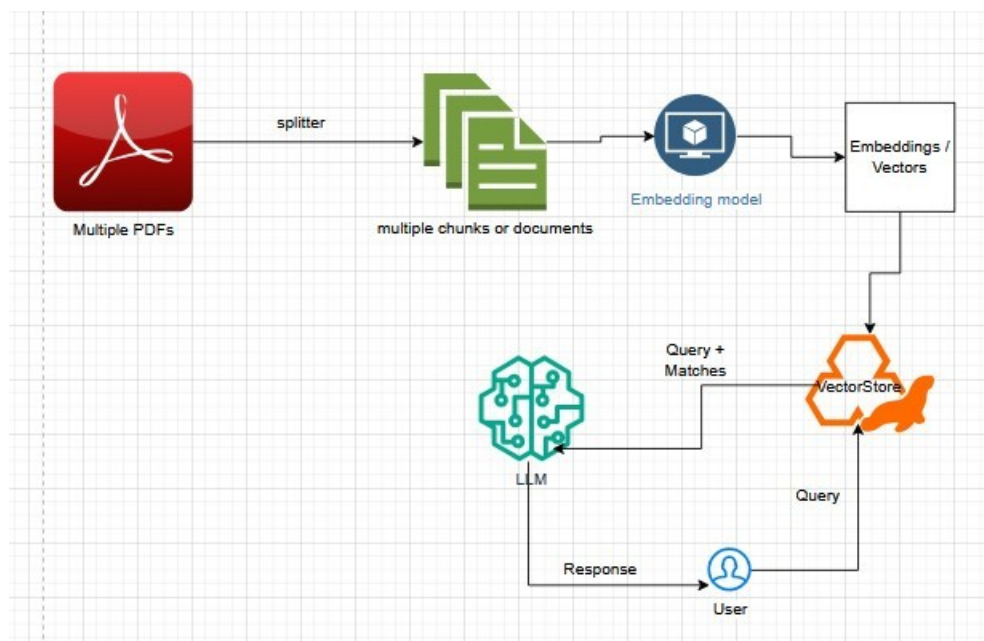
How do we include our documents into LLM?

PDF/docs ---> splitter ---> multiple chunks or documents ---> Embedding model ---> Embeddings/Vectors ---> VectorStore

User ---> Query ---> VectorStore (SimilaritySearch) ---> matches + query ---> LLM

LLM ---> responds ---> User

Check RAG_Concept pic



// you are talking to LLM but you are getting response based on your Product list
// for RAG

HTTP

http://localhost:8080/api/ask/suggest products for vacation

Save

Share

GET

http://localhost:8080/api/ask/suggest products for vacation

Send

Docs

Params

Authorization

Headers (7)

Body

Scripts

Tests

Settings

Cookies

Query Params

	Key	Value	Description	*** Bulk Edit
	Key	Value	Description	

Body

Cookies

Headers (5)

Test Results

200 OK

8.46 s

304 B

...

Raw

Preview

Visualize

1

For vacation, consider the "Waterproof Travel Backpack" for its multi-compartment design, padded laptop sleeve, and water-resistant fabric.

HTTP

http://localhost:8080/api/ask/suggest products for workout

Save

Share

GET

http://localhost:8080/api/ask/suggest products for workout

Send

Docs

Params

Authorization

Headers (7)

Body

Scripts

Tests

Settings

Cookies

Query Params

Key	Value	Description	Bulk Edit
Key	Value	Description	

Body

Cookies

Headers (5)

Test Results

200 OK

4.06 s

400 B

Raw

Preview

Visualize

1 I suggest the "Yoga Mat (6mm, Non-Slip)" for your workout needs. It is a high-density, eco-friendly mat suitable for yoga, Pilates, or floor exercises. It features an anti-tear design, is lightweight, easy to roll, and sweat-resistant.

// with chat memory advisors

// GET http://localhost:8080/api/ask/suggest products for cooking

GET

http://localhost:8080/api/ask/suggest products for cooking

Send

Docs

Params

Authorization

Headers (7)

Body

Scripts

Tests

Settings

Cookies

Query Params

Key	Value	Description	Bulk Edit
Key	Value	Description	

Body

Cookies

Headers (5)

Test Results

200 OK

6.87 s

670 B

Raw

Preview

Visualize

1 Based on the provided context, here are some products you might find useful for cooking:

2

3 1. **Silicone Baking Mats (Set of 2)**

4 - Description: Non-stick mats for baking with reusable, heat-resistant silicone.

5 - Price: \$13.00

6

7 2. **Kitchen Scale**

8 - Description: Precise measurement for cooking and baking, with tare function.


9 - Price: \$14.75

10 - Features: LCD screen, Slim design, Units in oz/g/ml/lb, Auto-off

11

12 If you need more suggestions or specific types of products, please let me know!

// After implementing PromptTemplate

 http://localhost:8080/api/ask/suggest products for cooking

Save

Share

GET

http://localhost:8080/api/ask/suggest products for cooking

Send

Docs

Params

Authorization

Headers (7)

Body

Scripts

Tests

Settings

Cookies

Query Params

	Key	Value	Description	...	Bulk Edit
	Key	Value	Description		

Body

Cookies

Headers (5)

Test Results

200 OK

2.27 s

326 B

...

Raw

Preview

Visualize

1

Title: "Silicone Baking Mats (Set of 2)"

2

Price: \$13

3

Category: Cooking/Baking

4

Description: Non-stick mats for baking with reusable, heat-resistant silicone.