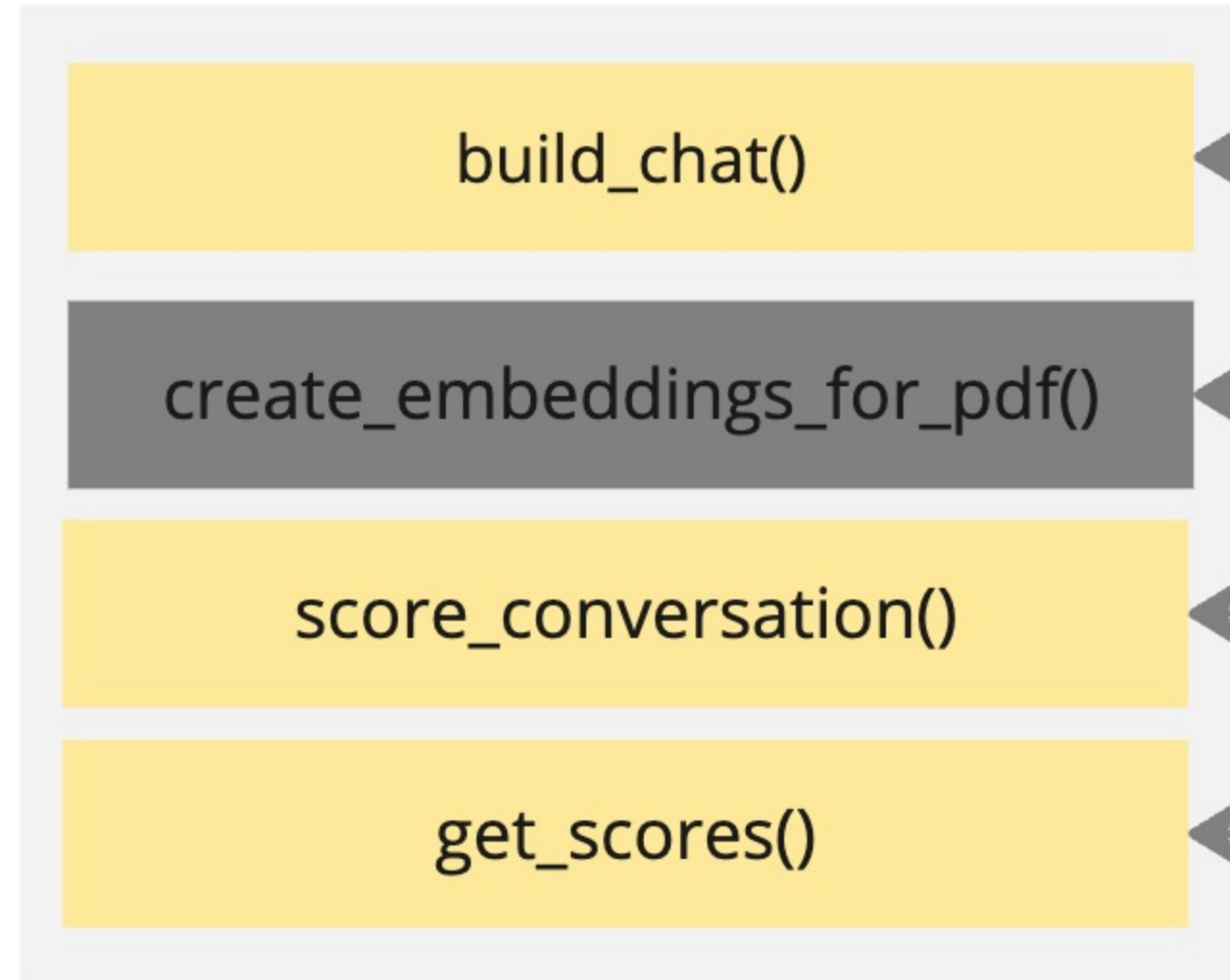


## **app.chat Module**



*We need to implement these! (plus tons of supporting code)*

**app.web**

*The app.web modules needs us to implement these 4 functions*

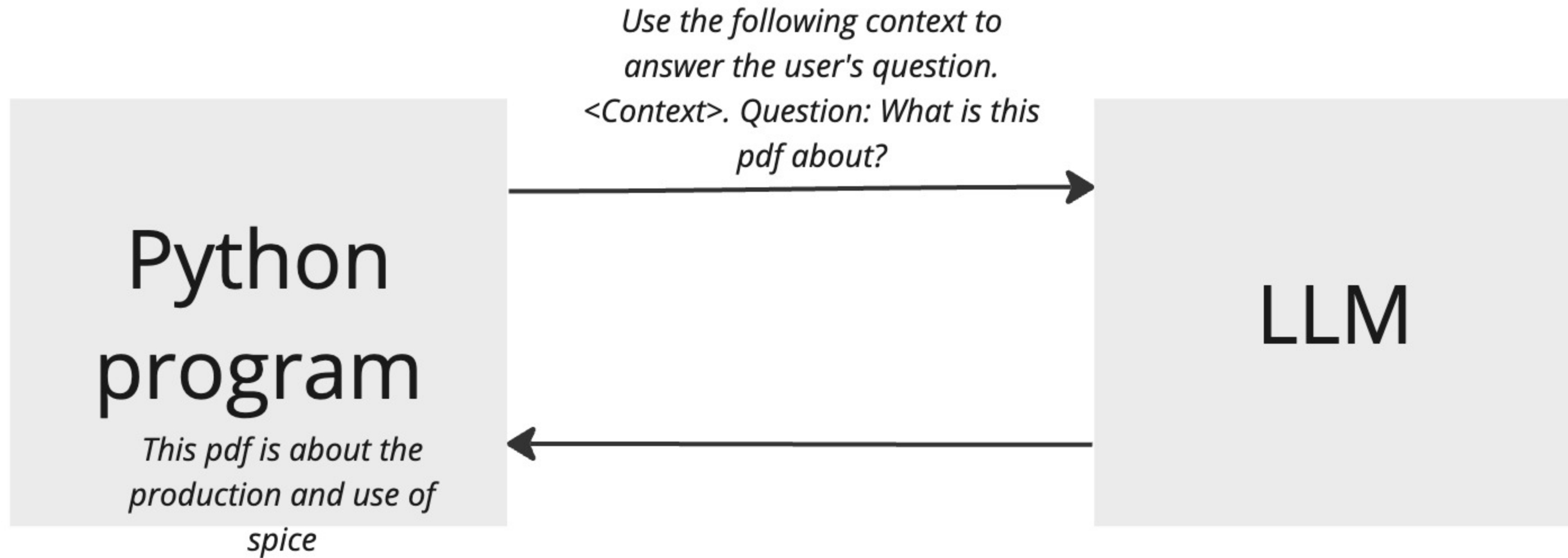
## **build\_chat()**

Builds and returns a Conversational Retrieval Chain to answer a user's question

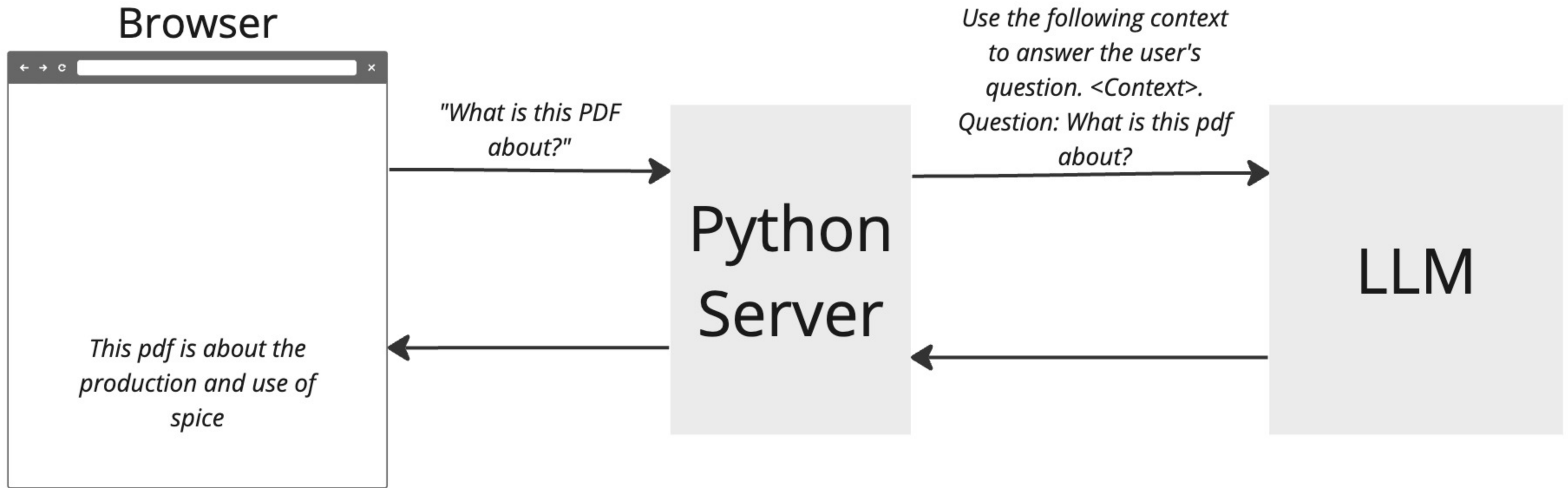
The chain should support streaming text generation

The chain should use random variations of component parts

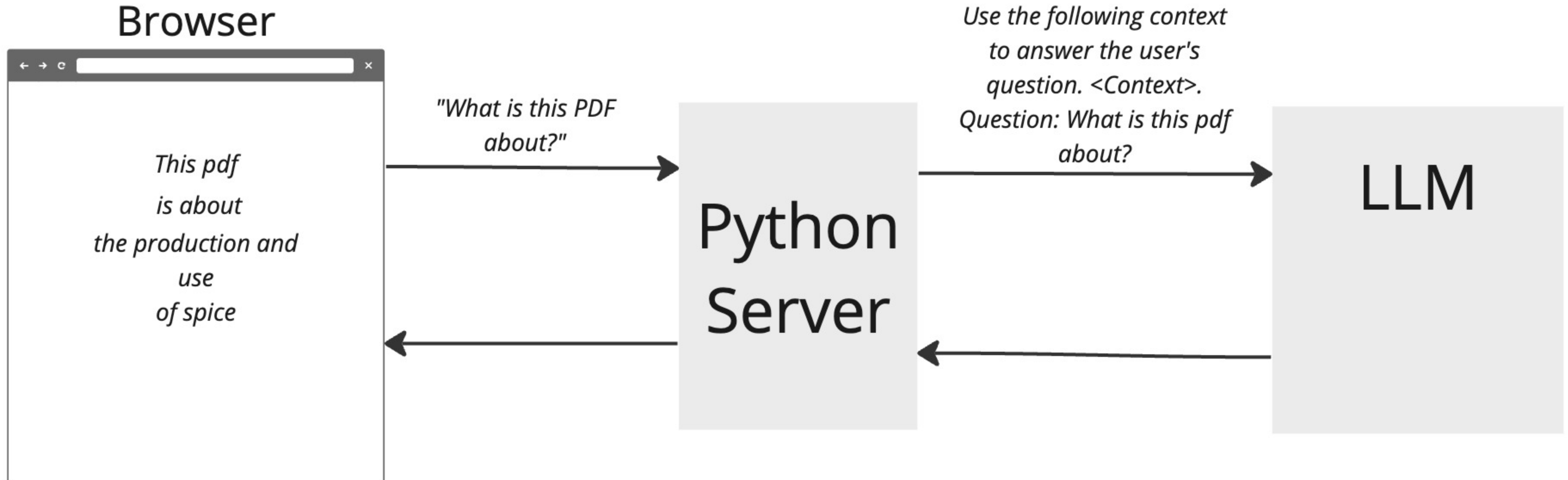
# Bulk Text Generation



# Bulk Text Generation



# Streaming Text Generation





# RetrievalChain

*Retriever*

*Memory*

*LLM*

## Retrievers

*retriever that fetches 2  
relevant documents*

*retriever that fetches 3  
relevant documents*

## LLM

*GPT-3.5-turbo*

*GPT-4*

## Memory

*Memory that remembers **all**  
previous questions about a pdf*

*Memory that doesn't remember  
any previous conversations*

*What is this PDF about?*

### RetrievalChain

*retriever that fetches 2 relevant documents*

*Memory that remembers all previous questions about a pdf*

**GPT-3.5-turbo**

**Bad Answer**

This PDF is, like, about stuff you put on food or something.

*What is this PDF about?*

### RetrievalChain

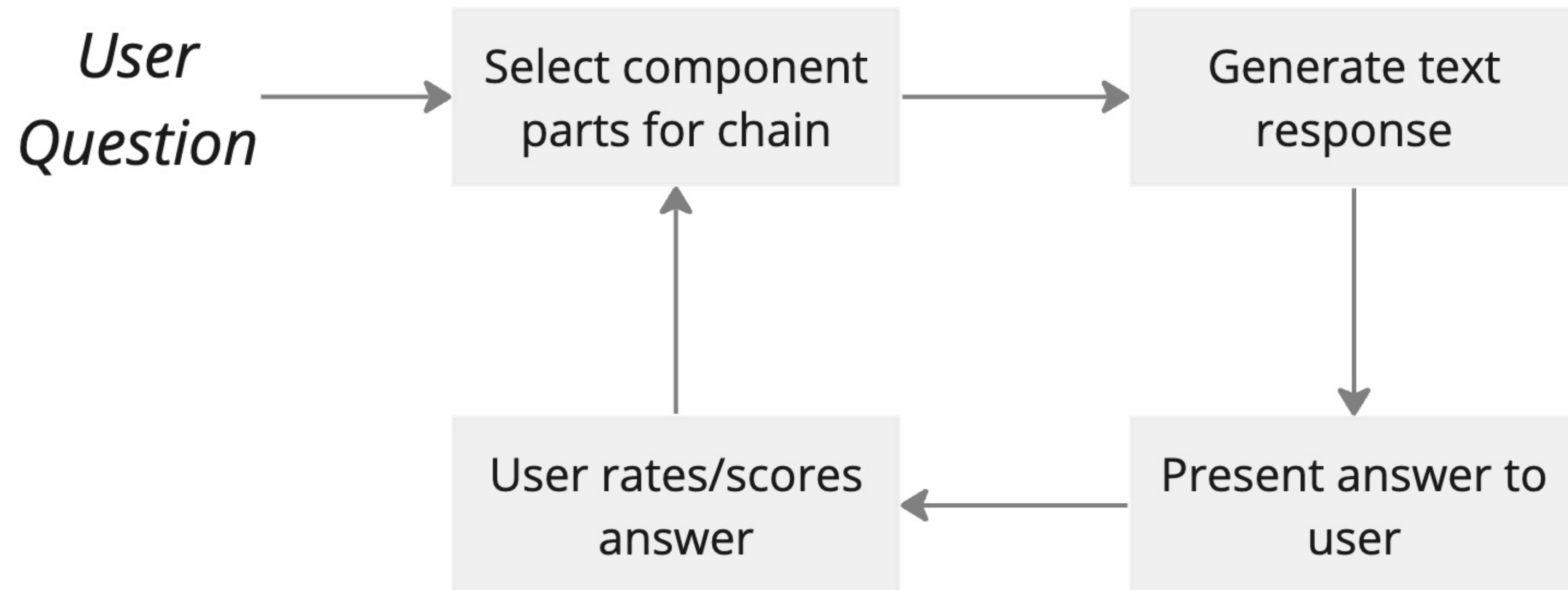
*retriever that fetches 3 relevant documents*

*Memory that remembers all previous questions about a pdf*

**GPT-4**

**Good Answer**

This PDF is about the production and usage of spice. It also focuses on the origins and top producers of spice.





# Chat Panel

What is this PDF about?

This pdf is about the production and usage of spice



How is spice made?

Spice is made by chopping or grinding up plant matter



How do we keep track of user's "liking" or "disliking" an answer?

How do we keep track of what combination of parts was used to generate an answer?

# Pinecone





spice.pdf



Pinecone ID	vector	metadata		
		page	source	text
be17bac7	[ ... ]	3	/tmp/tmp.pdf	"spice is..."
ccb58e26	[ ... ]	5	/tmp/tmp.pdf	"spice is..."
3a495ca2	[ ... ]	1	/tmp/tmp.pdf	"spice is..."
9f78186e	[ ... ]	2	/tmp/tmp.pdf	"spice is..."

# Pinecone

 **spice.pdf**

 **transistors.pdf**

Pinecone ID	vector	metadata		
		page	source	text
be17bac7	[ ... ]	3	/tmp/tmp.pdf	"spice is..."
ccb58e26	[ ... ]	5	/tmp/tmp.pdf	"spice is..."
3a495ca2	[ ... ]	1	/tmp/tmp.pdf	"spice is..."
9f78186e	[ ... ]	2	/tmp/tmp.pdf	"spice is..."
cb58e26	[ ... ]	3	/tmp/tmp.pdf	"transistors are..."
a495ca2	[ ... ]	1	/tmp/tmp.pdf	"transistors are..."
f78186e	[ ... ]	2	/tmp/tmp.pdf	"transistors are..."

*User's Question*

**What is this pdf  
about?**




**[ -.01, .34, .92, ...]**


*embedding*



Pinecone ID	vector	metadata		
		page	source	text
be17bac7	[ ... ]	3	/tmp/tmp.pdf	"spice is..."
ccb58e26	[ ... ]	5	/tmp/tmp.pdf	"spice is..."
3a495ca2	[ ... ]	1	/tmp/tmp.pdf	"spice is..."
9f78186e	[ ... ]	2	/tmp/tmp.pdf	"spice is..."
cb58e26	[ ... ]	3	/tmp/tmp.pdf	"transistors are..."
a495ca2	[ ... ]	1	/tmp/tmp.pdf	"transistors are..."
f78186e	[ ... ]	2	/tmp/tmp.pdf	"transistors are..."

# Pinecone

 **spice.pdf**

 **transistors.pdf**

Pinecone ID	vector	metadata			
		page		text	pdf_id
be17bac7	[ ... ]	3		"spice is..."	123
ccb58e26	[ ... ]	5		"spice is..."	123
3a495ca2	[ ... ]	1		"spice is..."	123
9f78186e	[ ... ]	2		"spice is..."	123
cb58e26	[ ... ]	3		"transistors are..."	456
a495ca2	[ ... ]	1		"transistors are..."	456
f78186e	[ ... ]	2		"transistors are..."	456