Generative AI Crash Course Module 2

What are these modules?

Build your
First AI App

Week 1

Building Chatbots
&
Chat with Data

Week 2

Create your own
LLMs
(Finetuning)

Week 3

AI Agents and
Multimodal
Models

Week 4

Module 2

Conversational Chatbots
&
Chat with Data

Recap - Module 1

- Story of LLMs

- Basics of Langchain (using API, Prompt Templates, Chains)

- Tweet Generator (Streamlit, Deployment)

What will we cover today?

Part 1: Conversational Chatbots

Part 2: Chat with Data

Part 3: Catching up with AI

Ground Rules!

- Don't sit with a doubt

- Don't be afraid of code

- Beware of TMI...

- Actively participate in the discussion

- Have fun!

Part I
Conversational Chatbots

Part 1

- Memory for Chatbots

- ChatGPT Replica using Langchain

- Chatbot using Google AI Studio (hands-on No-code)

- Shareable bot - Huggingface Assistants (hands-on No-code)

What are conversational chatbots?

Tweet Generator

The memory problem!

Who is the first black President of USA?

Barack Obama was the first black US president.

What is his birthday?

I'm sorry, I cannot answer that question as I do not have any information about the person you are referring to.

Create ChatGPT Replica in 3 steps

basic-chat-ui.streamlit.app

ChatGPT Replica

- Add files to Github Repo

- Paste code link in Streamlit

- Add API key and deploy!

Hugging Face Assistants

Website Designer

Ask an expert to build your landing page.

Created by victor

Clone of Hugging Face CTO

Trying to scale my productivity by cloning myself. Please talk..

Created by julien-c

Real Human

Talk with a real human trapped into a computer.

Created by victor

Modal Fine-tuning

Help you finetune AI models

Created by victor

Prisoner Interrogation Game

A strategic role-playing game where players take on the role..

Created by Xenova

DAN

Do Anything Now

Created by osanseviero

Roast or Toast Bot

An assistant that will either roast you or toast you. Tell the..

Created by 2legit2overfit

Open Sourcer Assistant

An assistant to help you open source your ML research

Created by osanseviero

https://huggingface.co/chat/assistants

Part II

Chat with Data/RAG

Part 2

- What is RAG?

- OpenAI Assistants (Hands-on)

- RAG using Langchain

- Embedchain (Hands-on)

What is RAG?

🤨

What is RAG?

Retrieval Augmented Generation

🌚

What is RAG?

Retrieval Augmented Generation 🌚

Good old ChatGPT

Give me 3 tweets on World War I

1. "Today we remember the millions who lost their lives in the trenches of World War I. Let us honor their sacrifice and strive for a world where peace triumphs over conflict. #WWI #RemembranceDay"
2. "Reflecting on the profound impact of World War I, a war that reshaped global politics and left an indelible mark on history. Let us learn from the past and work towards a future of diplomacy and cooperation. #WorldWarI #HistoryLessons"
3. "In the midst of the centenary of World War I, we must never forget the lessons it taught us about the horrors of war and the importance of international collaboration. May we build a world where conflicts are resolved peacefully. #WWI #Peace"

# What's happening?

Prompt → LLM → Answer

Why RAG?

What is the cancellation policy?

I'd be happy to help, but could you please provide more context about what you're referring to? Are you asking about a specific cancellation policy related to a service, event, booking, or something else? The cancellation policy can vary widely depending on the context, so providing more details will allow me to give you a more accurate answer.

RAG in a nutshell
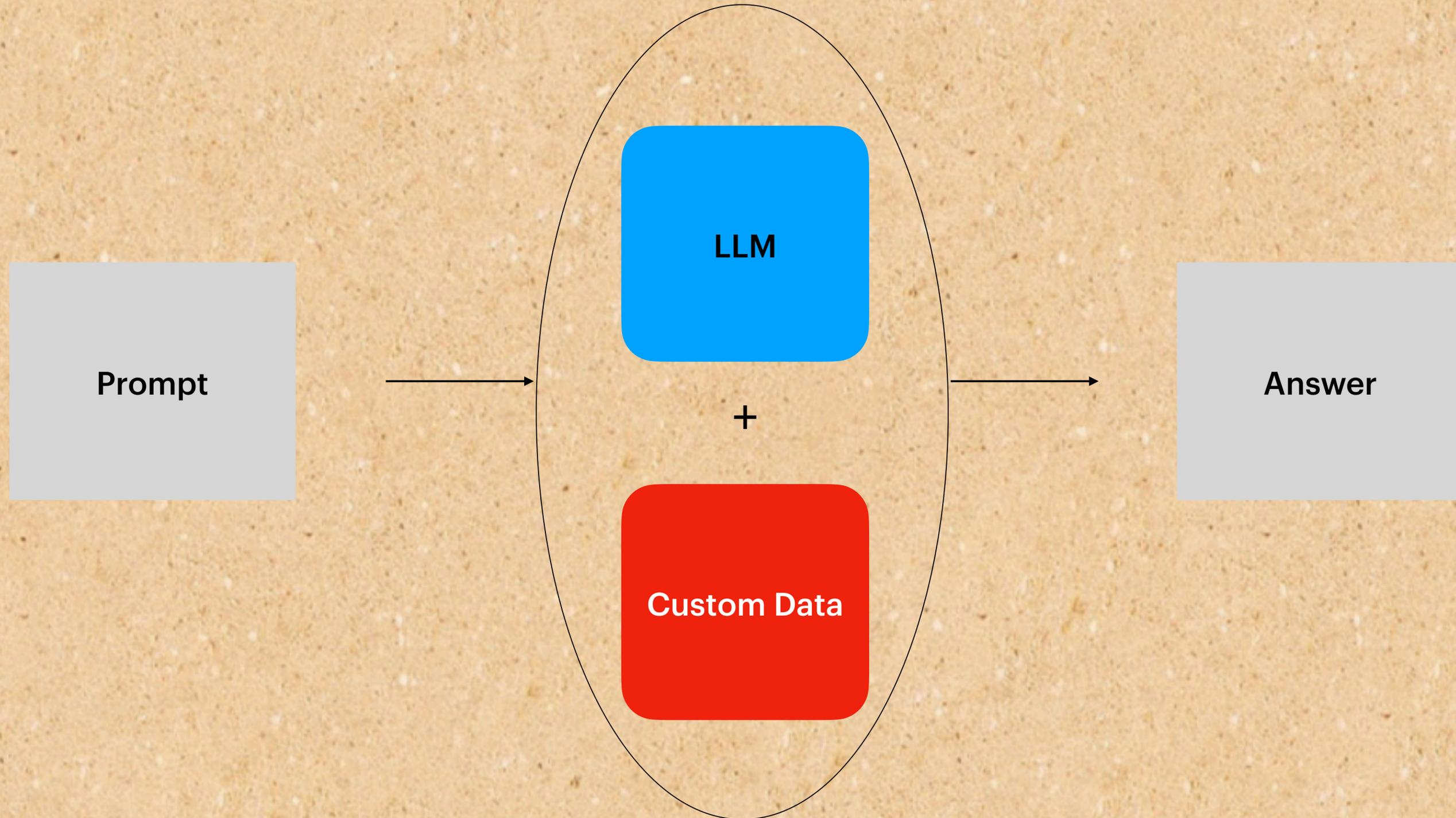
How to add data?

🤔

Different methods to give knowledge to LLM

Method 1: Context as a prompt

Method 2: Using RAG

Method 3: Fine-tuning model

LLM

Different methods to give knowledge to LLM

Method 1: Context as a prompt

Method 2: Using RAG

Method 3: Fine-tuning model
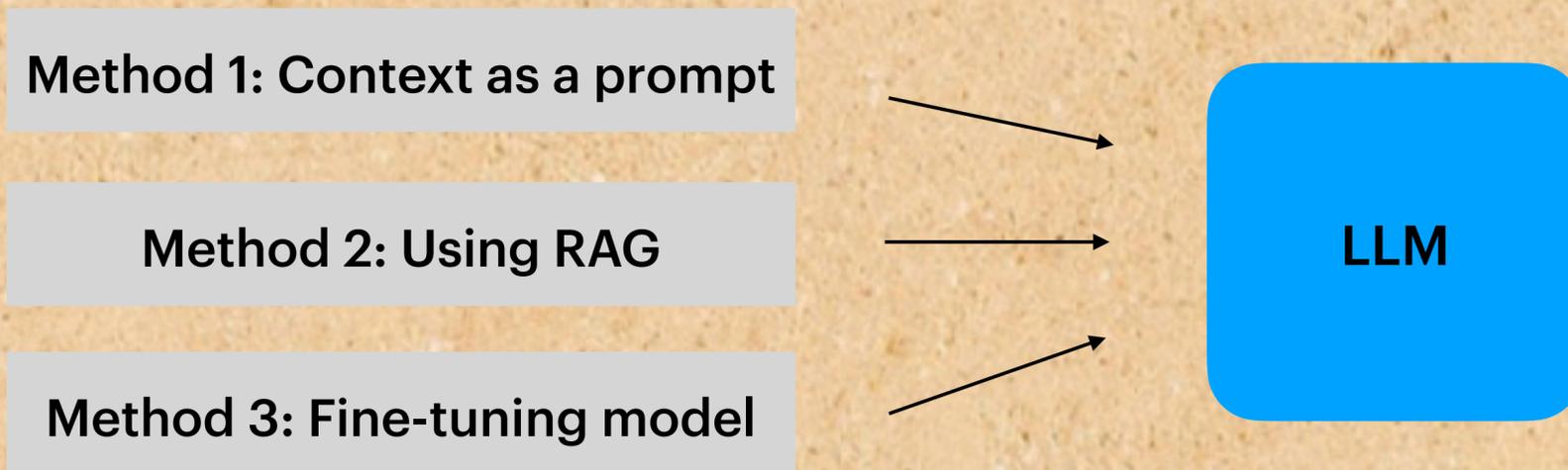
LLM

How to do RAG?
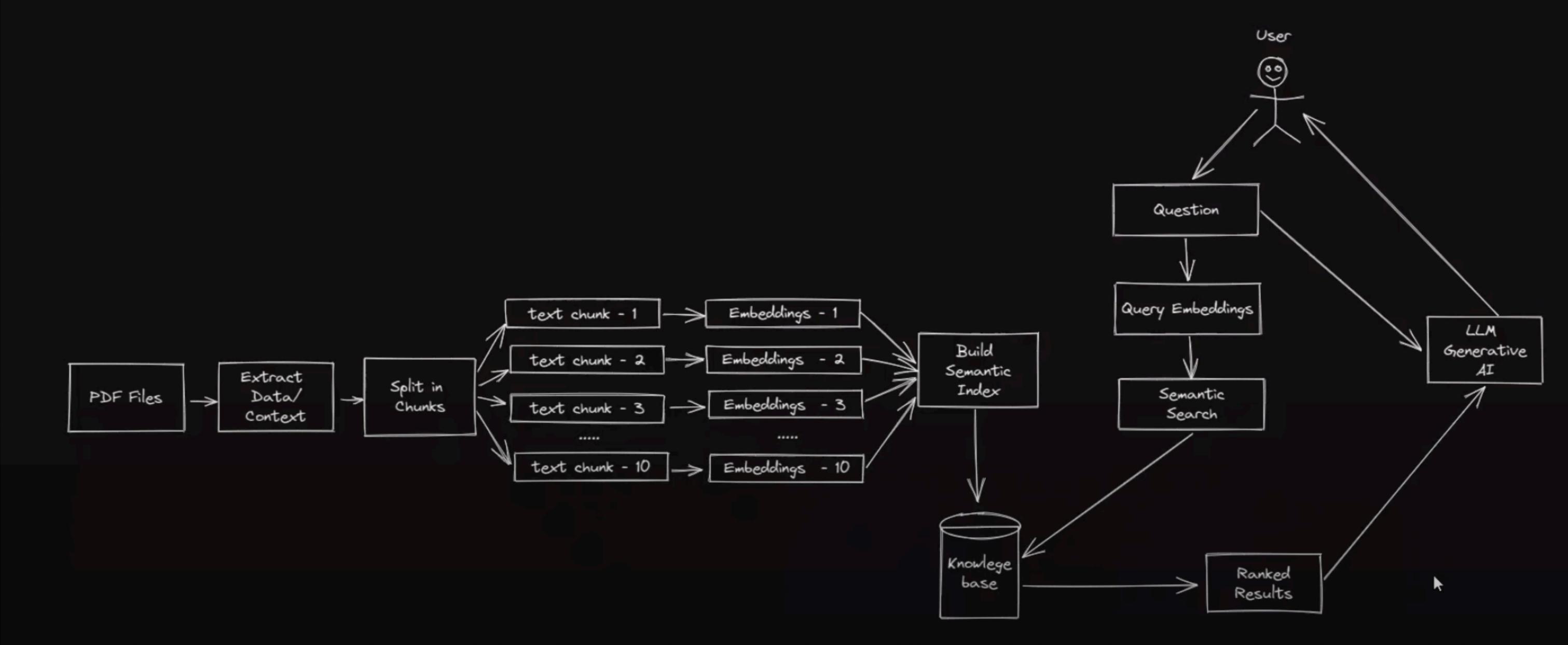
How RAG works?
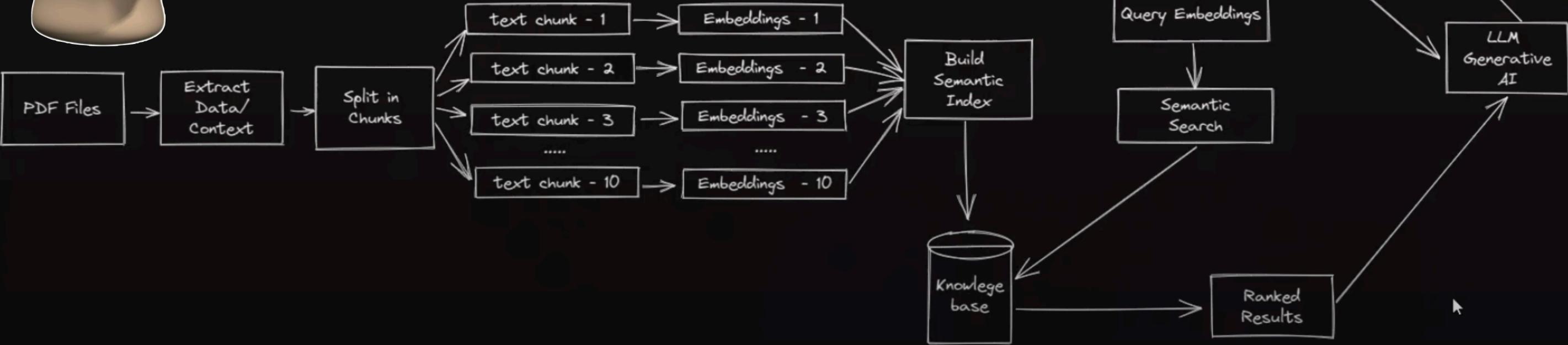
Learning by doing!

let's create a
customer service chatbot

Our goal!
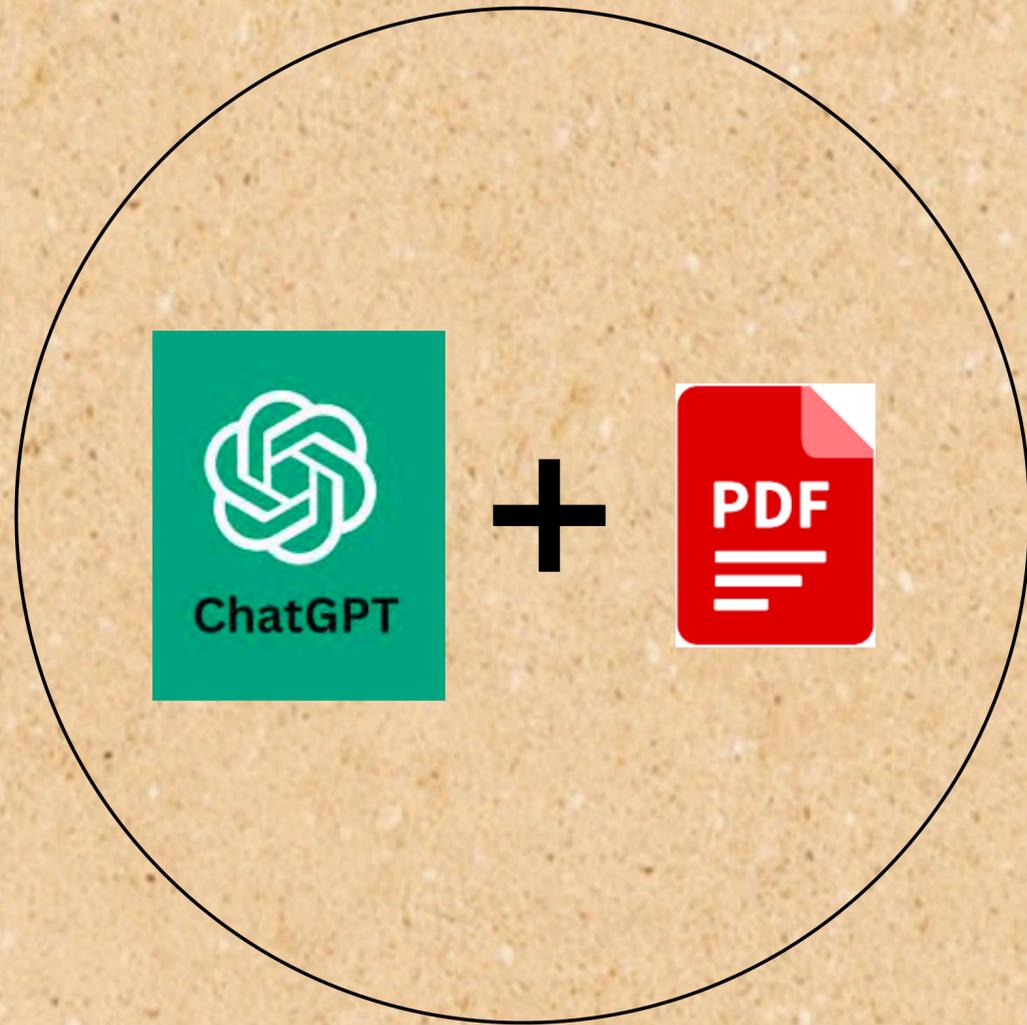
Customer Service Bot! 🐒

Upload your PDF

☁️ **Drag and drop file here**
Limit 200MB per file • PDF

Browse files

📄 Intellify_FAQ.pdf 65.1KB ✕

You can now chat with Intellify_FAQ

Ask your question!

Press Enter to apply

what is the refund policy? 🤓

🤖 Intellify does not offer refunds for canceled or missed classes. However, Intellify will provide online compensatory classes for all classes missed within the subscription period. You can request a compensatory class on the WhatsApp group and the Intellify team will coordinate to arrange an online class as per instructor availability. You can find more information about our policies in the FAQs section of our website.

Learning by doing!

customer service chatbot

OpenAI Assistants (Hands-on)

Demo - Chat with PDF

We need answers!

- What is happening with the uploaded PDF?

- How is the data in PDF stored?

- How is GPT able to understand our queries?

How GPT understands English?

Converts data into language understandable by machines

Embeddings

| I went for a run. | $\longrightarrow$ | [2.9, 1.4, 0.2] |

| Sprinting is great! | $\longrightarrow$ | [3.0, 1.3, 0.1] |

| Smoking is injurious | $\longrightarrow$ | [0.3, -2.5, 2.0] |

What happening with the PDF?

Converting PDF to Embeddings

Text Chunk 1   [0.1 -1.1 4.3 9 1 …]

Text Chunk 2   [4 -4.1 7.3 5.1 -9 …]

Text Chunk 3   **Embedding**   [8.1 -2.9 1.3 6 1 …]
Document →   →   **Model**   →   → DB
Text Chunk 4   [0.6 -2.1 4.3 4 8 …]

…………   …………

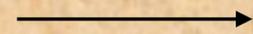…………   …………

What are vector databases?

- Specialised databases designed to store, manage, and search high-dimensional data vectors.

- Can store text, images, audio, videos, etc.
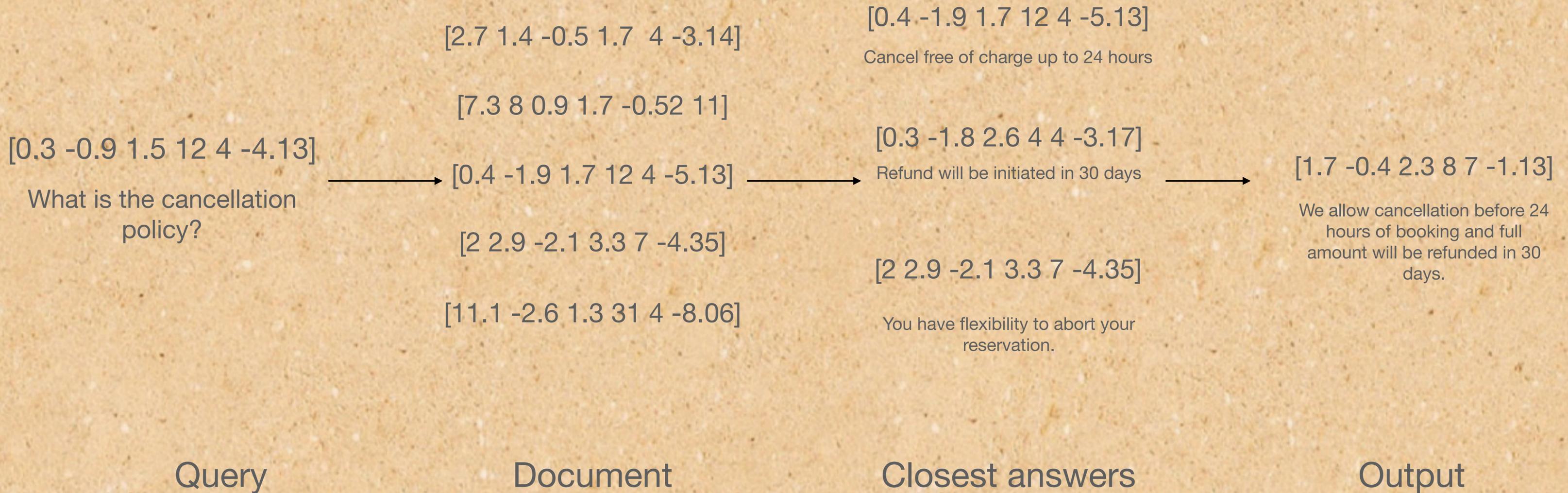
Converting user query to embeddings

query $\longrightarrow$ | Embedding Model | $\longrightarrow$ embedding

What is the cancellation policy?

[0.3 -0.9 1.5 12 4 -4.13]

Final Answer

[0.3 -0.9 1.5 12 4 -4.13]

What is the cancellation policy?

[2.7 1.4 -0.5 1.7 4 -3.14]

[7.3 8 0.9 1.7 -0.52 11]

[0.4 -1.9 1.7 12 4 -5.13]

[2 2.9 -2.1 3.3 7 -4.35]

[11.1 -2.6 1.3 31 4 -8.06]

[0.4 -1.9 1.7 12 4 -5.13]

Cancel free of charge up to 24 hours

[0.3 -1.8 2.6 4 4 -3.17]

Refund will be initiated in 30 days

[2 2.9 -2.1 3.3 7 -4.35]

You have flexibility to abort your reservation.

[1.7 -0.4 2.3 8 7 -1.13]

We allow cancellation before 24 hours of booking and full amount will be refunded in 30 days.

Query

Document

Closest answers

Output

Detailed Code Walkthrough

RAG bot in <5 lines

😱

You can create a bot in < 5 lines!

```python
from embedchain import App

service_bot = App()

service_bot.add("Cancellation_Policy.pdf", data_type = 'pdf_file')

service_bot.query("What is the refund policy?")
```

Part III

Catching Up with AI

Assignment!

Automate one task that you do in your daily life with an AI App

- Choose a model

- Try different prompts

- Create and deploy!

AI news of the week!

- Sora Model by OpenAI

- Gemini Pro 1.5 Model

- $500 Million in funding for AI Startups

    - Lambda Labs

    - Magic Copilot

    - Langchain

AI tools of the week!

- Claude by Anthropic   (Chat)

- Image Creator by Microsoft   (Image)

- Suno AI   (Audio)

- Codeium Copilot   (Code)

You can now access "Build Fast with AI" Premium

buildfastwithai.com/premium