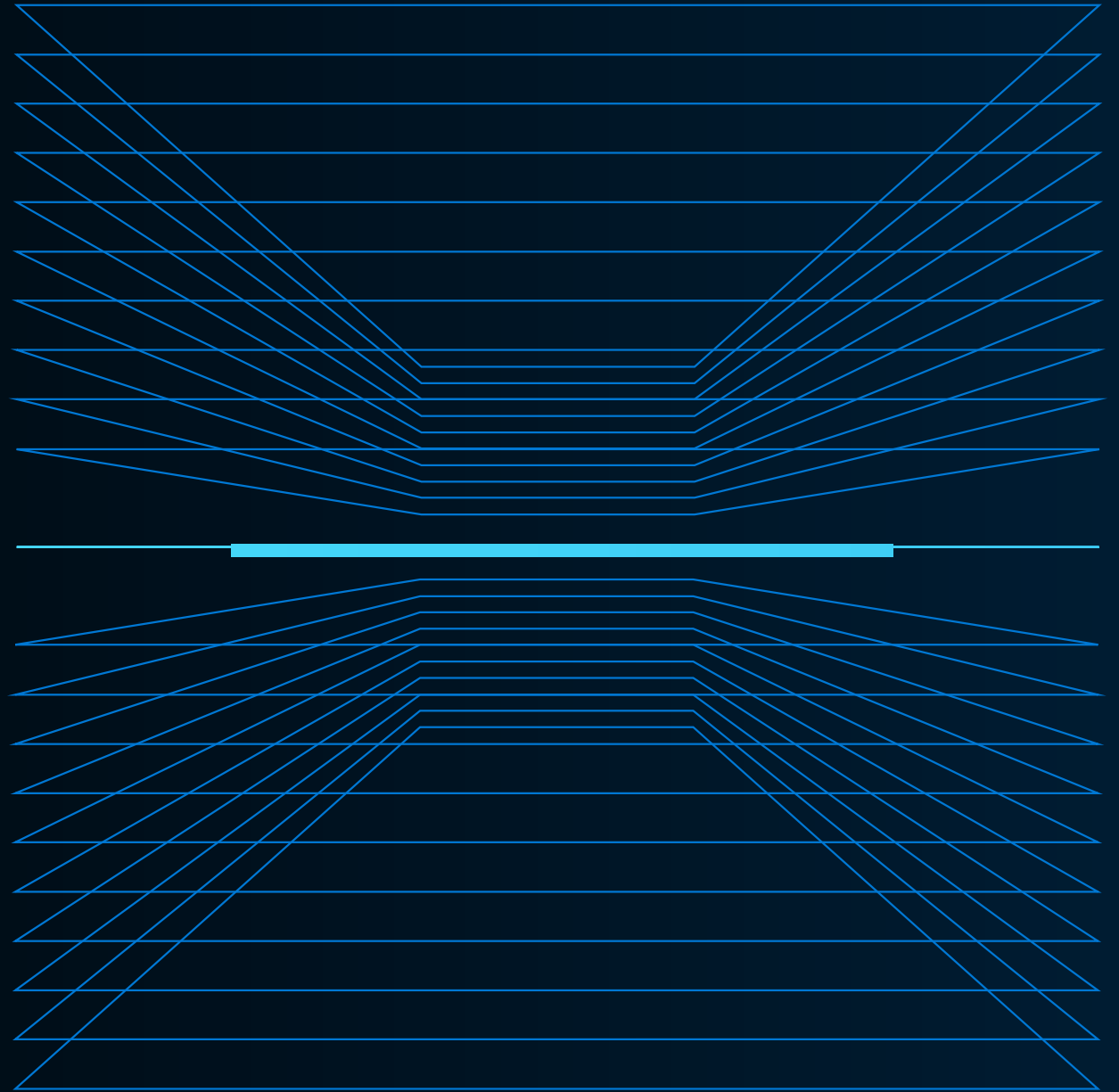Microsoft

Azure OpenAI Service
Deep-Dive Workshop

Goal for today

- Theory
  - Azure OpenAI Service
  - Completions
  - Embeddings
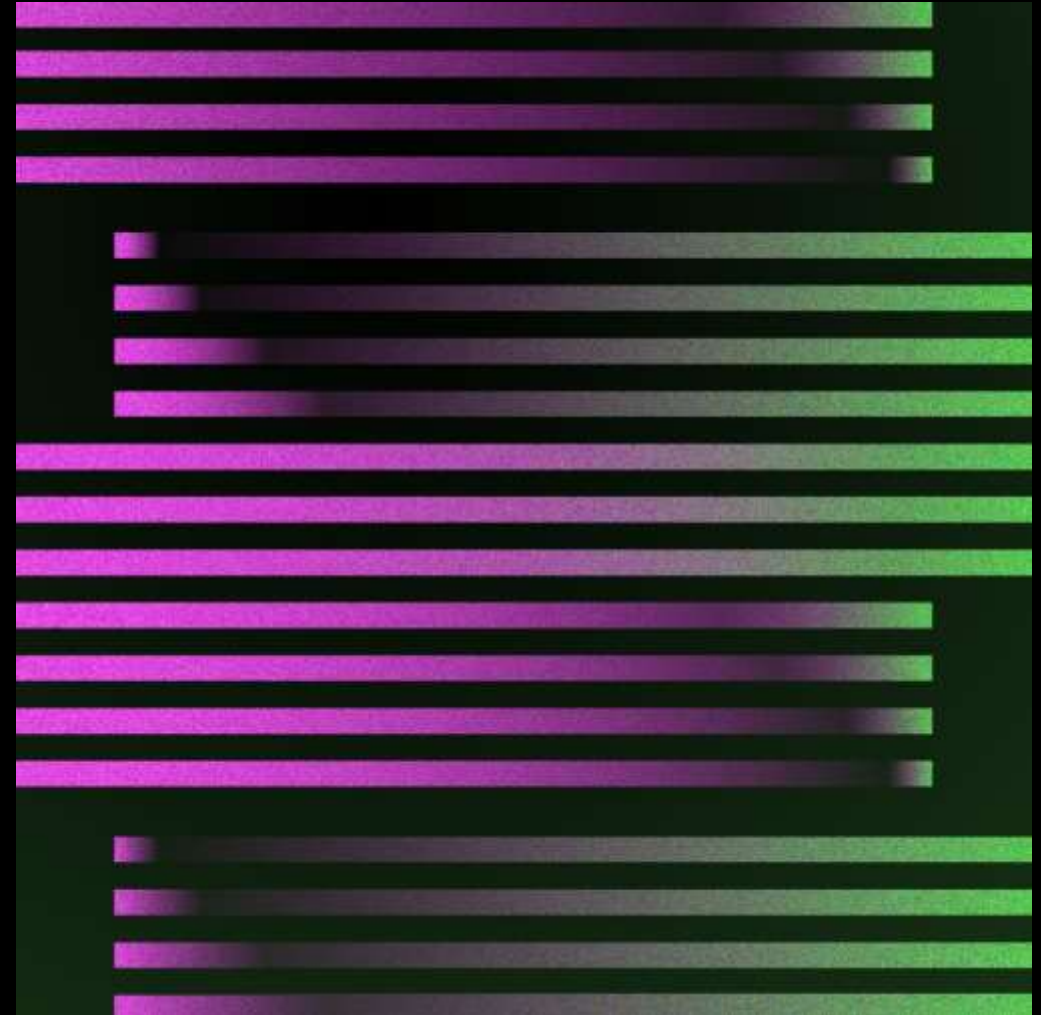 - Few-shot learning & fine-tuning
  - MVP Development
- Hands-on
  - 10+ prompt engineering exercises
  - Q&A with embeddings
  - Movies unsupervised classification & recommendation system
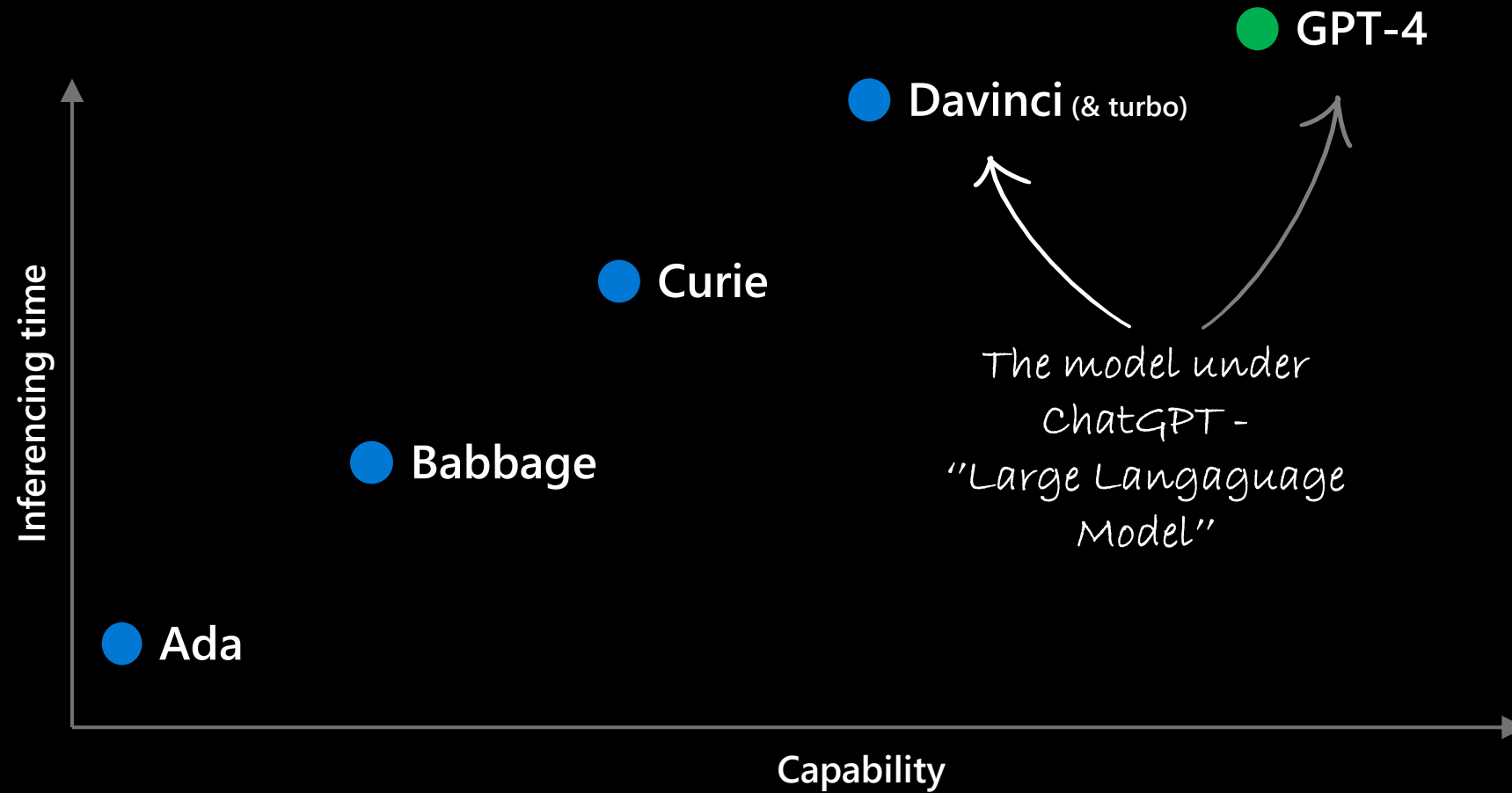
Plan for today

Time (CET)	Topic
09:00 – 09:45	Intro Azure OpenAI Service
09:45 – 10:15	Azure OpenAI Studio
10:15 – 10:30	Break
10:30 – 12:00	Prompt Engineering Exercises
12:00 – 13:00	Lunch Break
13:00 – 13:30	Use Cases and Common Architectures
13:30 – 14:15	Using Azure OpenAI Service to access company data
14:15 – 14:45	MVP Development
14:45 – 15:00	Break
15:00 – 16:30	Hands-on lab on two exemplary use-cases

Why so much hype around ChatGPT?

- ChatGPT
  - 100m users in 2.5 months
  - Most "human-like" interaction
- What's new about it?
  - Can generate and do "things"
  - Answer knowledge questions directly
  - Answer coding questions
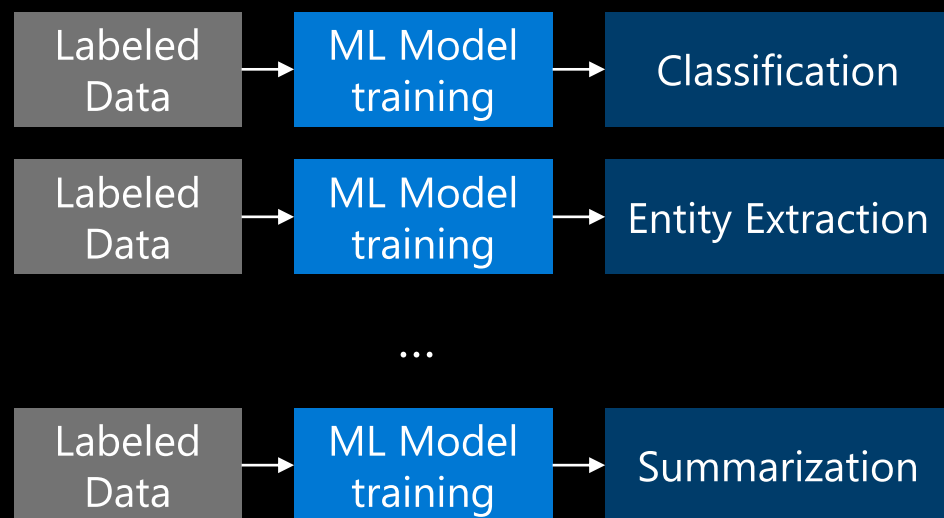 - Can correct itself, if you ask it
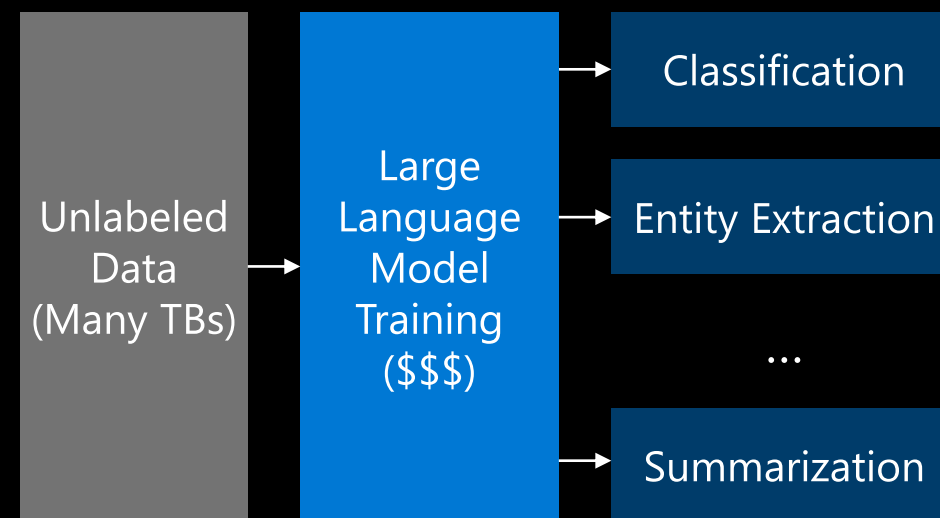- Prompt examples here

GPT-3 & GPT-4 Models

GPT-4

Davinci (& turbo)

Inferencing time

Curie

Babbage

The model under
ChatGPT -
"Large Langaguage
Model"

Ada

Capability

Large Language Models

Typical ML for NLP

| Labeled Data | → | ML Model training | → | Classification |

| Labeled Data | → | ML Model training | → | Entity Extraction |

...

| Labeled Data | → | ML Model training | → | Summarization |

One model per capability
Labeled data to train
Highly optimized for use case

Large Language Models

Unlabeled Data (Many TBs)	→	Large Language Model Training ($$$)	→	Classification
			→	Entity Extraction
				...
			→	Summarization

Single model for all use cases
Describe in natural language what it should do

LLMs can also "generate things"

"Classify this document"

"Write me a letter explaining why the markets are down, given these reasons: ..."

Large Language Model

Large Language Model

Class: Insurance document

New document, generated by the LLM
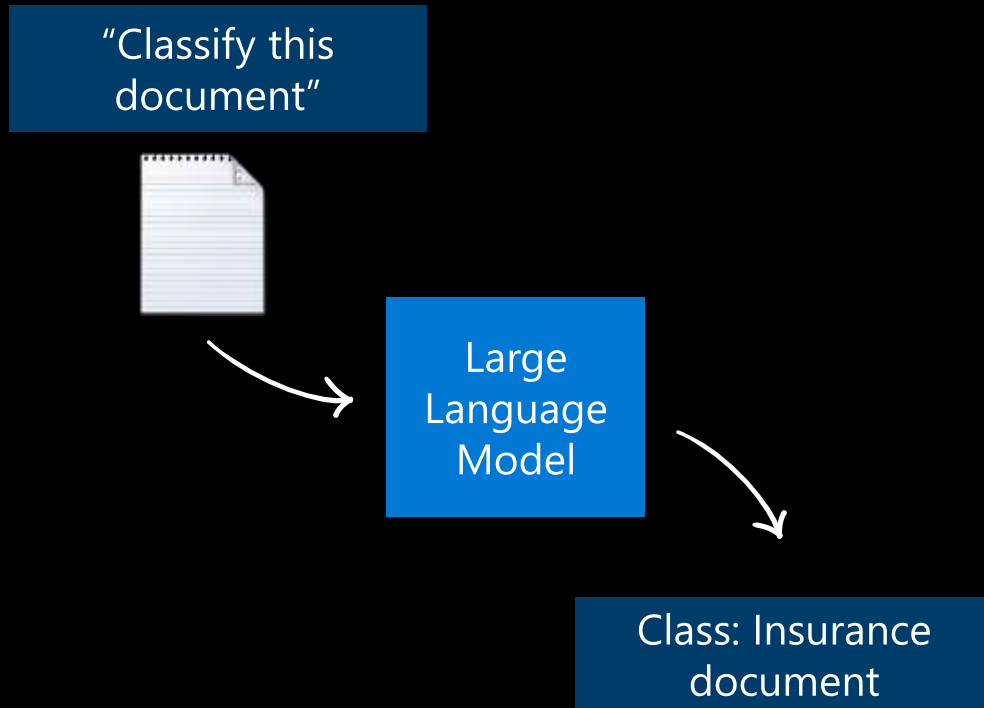
LLMs can do typical ML tasks easily (often zero-shot learned)
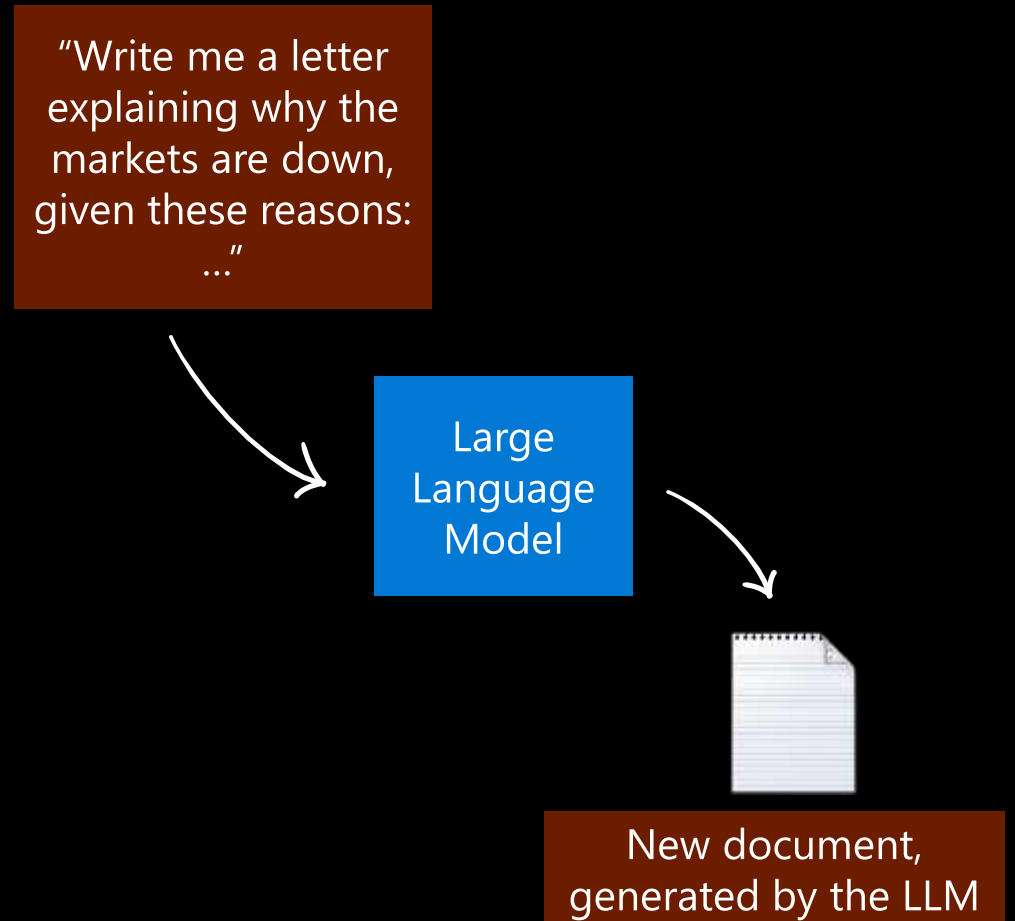
LLMs can also machine-read (understand) and generate new documents

Artificial Intelligence

Machine Learning

Deep Learning

Generative AI

1950's — **Artificial Intelligence**
the field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence

«10x» improvement

1990's — **Machine Learning**
subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions

«10x» improvement

2010's — **Deep Learning**
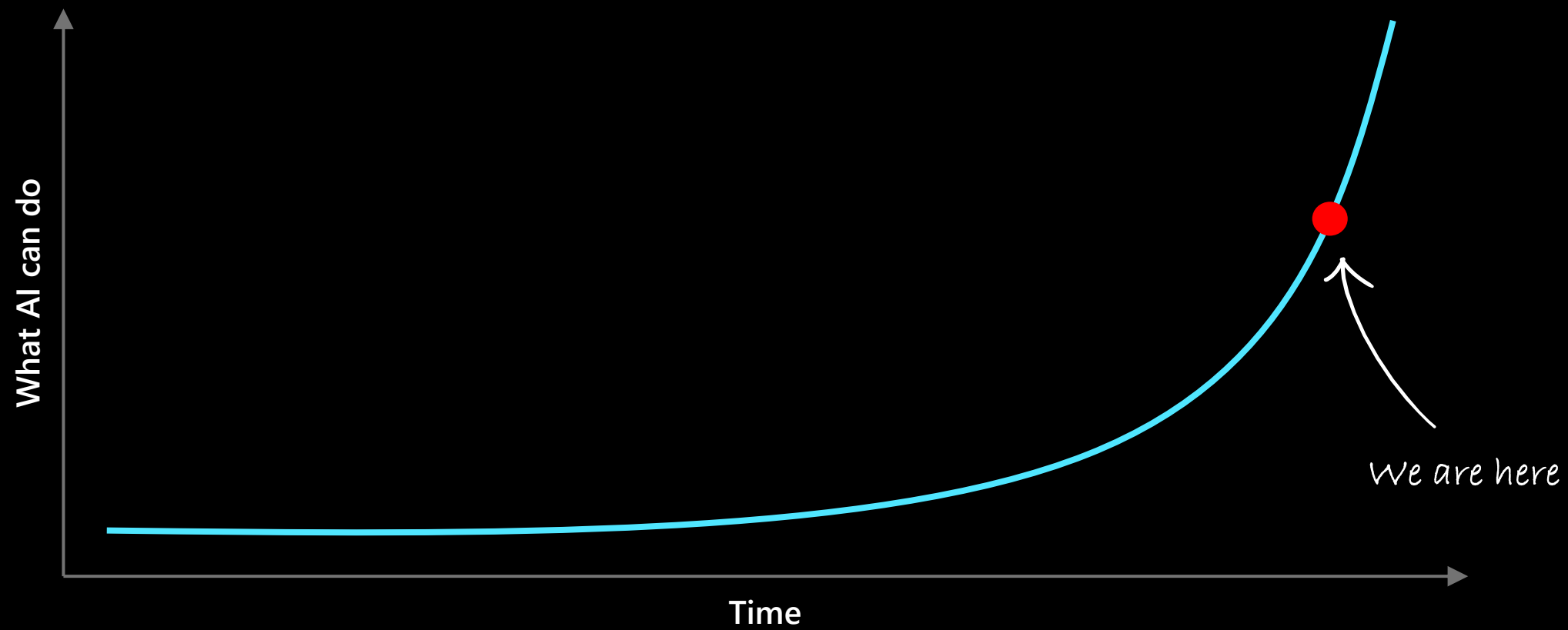a machine learning technique in which layers of neural networks are used to process data and make decisions

«10x» improvement

2020's — **Generative AI**
Create new written, visual, and auditory content given prompts or existing data.

Tell me about Azure OpenAI Service from Microsoft!

Azure OpenAI Service is a product that allows you to access powerful language models from OpenAI, such as GPT-3, Codex and DALL-E, through a REST API [1]. It is part of Azure Cognitive Services and runs on Azure global infrastructure [2]. What are you interested in learning more about Azure OpenAI Service?

Learn more: 1. learn.microsoft.com   2. azure.microsoft.com   3. learn.microsoft.com   +3 more   1 of 6 ●

How does Microsoft cooperate with OpenAI?

Microsoft and OpenAI have a long-term partnership that involves several aspects [1] [2]. Microsoft invested $1 billion in OpenAI in 2019 [1] [3] and became its exclusive cloud provider [4] [3]. The two companies also collaborate on developing AI supercomputing technologies [4] [2] and licensing some of OpenAI's technology to Microsoft for commercialization [4]. Azure OpenAI co-develops the APIs with OpenAI, ensuring compatibility and a smooth transition from one to the other [5]. What else would you like to know about Microsoft and OpenAI?

Learn more: 1. forbes.com   2. blogs.microsoft.com   3. openai.com   4. forbes.com   +2 more   2 of 6 ●

Did Microsoft give more funding in 2023?

Yes, Microsoft confirmed a new multi-year, multi-billion dollar (reported to be $10 billion) investment in OpenAI on January 23, 2023 [1] [2]. This was the third phase of their long-term partnership and followed their previous investments in 2019 and 2021 [1] [2]. Why are you curious about Microsoft's funding for OpenAI?

Learn more: 1. en.wikipedia.org   2. blogs.microsoft.com   3. en.wikipedia.org   +2 more   3 of 6 ●

Ask me anything...

OpenAI

Ensure that artificial general intelligence (AGI) benefits humanity.

Groundbreaking research
Latest models
Fast innovation

Microsoft

Empower every person and organization on the planet to achieve more

Enterprise features
Production SLAs
Data privacy

GPT-3
Generate and Understand Text

Codex
Generate and Understand Code

DALL·E *preview*
Generate images from text prompts

ChatGPT *preview*
Chat version of GPT-3

Azure AI

Applications

Microsoft 365   Microsoft Dynamics 365   **Partner Solutions**

Application Platform
AI Builder

Power BI   Power Apps   Power Automate   Power Virtual Agents

Scenario-Based Services
Applied AI Services

Bot Service   Cognitive Search   Form Recognizer   Video Indexer   Metrics Advisor   Immersive Reader

Customizable AI Models
Cognitive Services

Vision   Speech   Language   Decision   Azure OpenAI Service

ML Platform

Azure Machine Learning

Business Users

Developers & Data Scientists

Azure OpenAI Service

GPT-3	Codex
DALL·E *preview*	**ChatGPT** *preview*

Deployed within your Azure subscription, secured by you, accessed only by you, and tied to your datasets and applications

Large, pretrained AI models to unlock new scenarios

Custom AI models fine-tuned with your data and hyperparameters

Built-in responsible AI to detect and mitigate harmful use

Enterprise-grade security with role-based access control (RBAC) and private networks

Demo Azure OpenAI Studio

Break

Azure OpenAI Service | GPT-3 Prompt Design

Extract the mailing address from this email:

Hi John Doe,
It was great to meet up at Build earlier this week. I thought the AI platform talk was great and I really enjoyed it.

I appreciate the offer for the book. If you are OK, you can mail it to me at home, or 123 Microsoft Way, Bellevue WA 92004.

Regards,
Chris Hoder

Prompt – Text input that provides some context to the engine on what is expecting.

Completion – Output that GPT-3 generates based on the prompt.

Hands-on

- Signup
    - Go to Azure OpenAI Studio - Microsoft Azure
    - Walk through Text Completion tutorial (if needed)
        - https://beta.openai.com/docs/guides/completion
- Then do the exercises on the following slides
    - Each slide has some text and a desired completion (output)
    - Try to write a prompt, that achieves that completion
    - You'll find the solution for each exercise in the speaker notes
    - You'll also find the exercises here:
        - azure-openai-in-a-day-workshop/exercises.md at main · microsoft/azure-openai-in-a-day-workshop (github.com)

Exercise 1 – German translation

Exercise	Write a prompt that generates the expected completion
Input text	`I was enjoying the sun, but then a huge cloud came and covered the sky.`
Expected completion	`Ich genoss die Sonne, aber dann kam eine riesige Wolke und bedeckte den Himmel.`

Exercise 2 – Negation

Exercise	Write a prompt that generates the expected completion
Input text	`I was enjoying the sun, but then a huge cloud came and covered the sky.`
Expected completion	`I was not enjoying the sun, and then a huge cloud did not come and cover the sky.`

Exercise 3 – Classification

Exercise

Write a prompt that generates the expected completion

Input text

"Not much to write about here, but it does exactly what it's supposed to. filters out the pop sounds. now my recordings are much more crisp. it is one of the lowest prices pop filters on amazon so might as well buy it, they honestly work the same despite their pricing"

Expected completion
(or similar)

```
Positive: 0.75
Neutral: 0.20
Negative: 0.05
```

Exercise 4 – E-Mail Summarization

Exercise	Write a prompt that generates the expected completion
Input text	`Your own long email thread`

Expected completion (or similar)	`Summary: XYZ`
	`Open Questions: XYZ`
	`Action Items: XYZ`

Exercise 5 – Rewriting

Exercise	Write a prompt that generates the expected completion
| Input text | ```
I was enjoying the sun, but then a huge cloud came and
covered the sky.
``` |
| Expected completion | ```
She was enjoying the sun, but then a huge cloud came and
covered the sky.
``` |

Exercise 6 – Multiple Tasks

Exercise	Write a prompt that generates the expected completion
Input text	I was enjoying the sun, but then a huge cloud came and covered the sky.
Expected completion	```{ "translated": "Ich genoss die Sonne, aber dann kam eine riesige Wolke und bedeckte den Himmel.", "negated": "I was not enjoying the sun, and no huge cloud came and covered the sky.", "third_person": "She was enjoying the sun, but then a huge cloud came and covered the sky." }```

Exercise 7 – Data extraction to JSON

Exercise

Write a prompt that generates the expected completion

Input text

Hello, my name is Mateo Gomez. I lost my Credit card on August 17th, and I would like to request its cancellation. The last purchase I made was of a Chicken parmigiana dish at Contoso Restaurant, located near the Hollywood Museum, for $40. Below is my personal information for validation:

Profession: Accountant
Social Security number is 123-45-6789
Date of birth: 9-9-1989
Phone number: 949-555-0110
Personal address: 1234 Hollywood Boulevard Los Angeles CA
Linked email account: mateo@contosorestaurant.com
Swift code: CHASUS33XXX

Expected completion

```
{
    "reason": "Lost card",
    "classified_reason": "lost_card",
    "name": "Mateo Gomez",
    "ssn": "123-45-6789",
    "dob": "09/09/1989"
}
```

Exercise 8 – Fashion product description

Exercise

Write a prompt that generates the expected completion

Input text

```
Season: Winter
Style: Sweater
Gender: Female
Target group: Teenager
Material: Cotton
```

Expected completion
(or similar)

```
Stay warm and stylish this winter with our cozy cotton
sweaters, perfect for the fashion-forward teenager. Refresh
your wardrobe with the latest winter styles from our
collection.
```

Exercise 9 – Write a Blog

Exercise

Write a blog post about a topic of your choice

Input text

```
Step 1: I want you to act as a social media manager. You will
be helping me to brainstorm blog post outline ideas for the
topic <Topic 1>:
Step 2: Write 3 engaging and informative paragraphs about <Idea
1 description>
Step 3: Write 3 engaging and informative paragraphs about <Idea
2 description>
Step 4: Tags <List of relevant #hashtags>
```

Exercise 10 – Call Center

Exercise   Analyze a call center conversation

Input text

Employee: "Hello, this is Julia Schreider from Contoso Company. How can I help you today?"
Customer: "Hi, I am Carsten Mueller. I ordered a package 10 days ago, on February 10th, and it was supposed to arrive in maximum 5 business days. I have called three times already and nobody could provide any more information. I want to know where the package is and I want the problem to be solved immediately. This is the worst service I had for a long time!"
Employee: "I apologize for the inconvenience, Mr. Mueller. I understand your frustration and I'm here to help. Can you please provide me with your order number so I can look into this for you?"
Customer: "Yes, it's ACZ456789."
Employee: "Thank you. I'm looking into it now. Can you please hold for a few minutes while I check the status of your package?"
Customer: "Okay."
Employee: "Thank you for your patience. I am sorry to inform you that I am unable to find the status of your package. It appears to have left the sending address, but no up-to-date status on the current location. I will further investigate your case and get back to you as soon as possible via phone call. Could you please provide me your contact information?"
Customer: "Ah not again. Anyway, my phone number is +4911112223344."
Employee: "I apologize again for the inconvenience. Is there anything else I can help you with today?"
Customer: "No."
Employee: "Thank you. Have a great day!"

Expected Completion

```
{
    "classified_reason": "lost_package",
    "resolve_status": "unresolved",
    "call_summary": "Customer ordered package 10 days ago and has not received it yet.",
    "customer_name": "Carsten Mueller",
    "employee_name": "Julia Schreider",
    "order_number": "ACZ456789",
    "customer_contact_nr": "+4911112223344",
    "new_address": "N/A",
    "sentiment_initial": ["angry", "frustrated"],
    "sentiment_final": ["calm"],
    "satisfaction_score_initial": 0,
    "satisfaction_score_final": 5,
    "eta": "N/A",
    "action_item": ["track_package", "inquire_package_status", "contact_customer"]
}
```

Exercise 11 – Few-shot learning

Exercise

Write a few-shot learned prompt that classifies a movie summary.

Data samples

Paul Atreides, a brilliant and gifted young man born into a great destiny beyond his understanding, must travel to the most dangerous planet in the universe to ensure the future of his family and his people. As malevolent forces explode into conflict over the planet's exclusive supply of the most precious resource in existence-a commodity capable of unlocking humanity's greatest potential-only those who can conquer their fear will survive.
['Action', 'Adventure', 'Science Fiction']

A botched store robbery places Wonder Woman in a global battle against a powerful and mysterious ancient force that puts her powers in jeopardy.
['Action', 'Adventure', 'Fantasy']

After the devastating events of Avengers: Infinity War, the universe is in ruins due to the efforts of the Mad Titan, Thanos. With the help of remaining allies, the Avengers must assemble once more in order to undo Thanos' actions and restore order to the universe once and for all, no matter what consequences may be in store.
['Adventure', 'Science Fiction', 'Action']

A widowed new dad copes with doubts, fears, heartache and dirty diapers as he sets out to raise his daughter on his own. Inspired by a true story.
['Drama', 'Family', 'Comedy']

New data:
Harry, Ron and Hermione walk away from their last year at Hogwarts to find and destroy the remaining Horcruxes, putting an end to Voldemort's bid for immortality. But with Harry's beloved Dumbledore dead and Voldemort's unscrupulous Death Eaters on the loose, the world is more dangerous than ever.

Exercise 12 – NL to SQL with Codex

Exercise | Write a prompt that generates the expected SQL statement

Table details
```
Table: customer // Columns: firstname, name, customer_id, address
Table: orders // Columns: order_id, customer_id, product_id, product_amount
Table: products // Columns: product_id, price, name, description
```

Query
```
A query that returns the top 10 orders and show the customer name
```

Break

Top 3 Azure OpenAI Service Use Cases

Contact Center

Summarization, information extraction from call center calls and chat conversations

Proof: Telco, banking, insurance

ChatGPT for my Enterprise

Index knowledge base and give generative answers to arbitrary questions

Proof: Nearly every industry

Document Intelligence

Information extraction, classification, summarization

Extract complex entities from documents and their relationship (e.g., in media or research)

Proof: Nearly every industry

Azure OpenAI Service – Implemented Use Case

- ## Contact Center & Customer Support
  - Summarization, information extraction from call center calls and chat conversations
  - **Proof:** Several telcos

- ## Question & Answering (Q&A)
  - Index knowledge base and give generative answers to arbitrary questions
  - **Proof:** Several insurances

- ## Product Description Generation
  - Generate product taglines from metadata
  - **Proof:** Several fashion retailers

- ## Code generation
  - Code creation, summarization, explanation
  - **Proof:** Several Professional Service Companies

- ## Document Intelligence
  - Information extraction, classification, summarization on documents, e.g., maintenance logs
  - Extract complex entities from documents and their relationship (e.g., in media or research)
  - **Proof:** Largely used in Energy sector

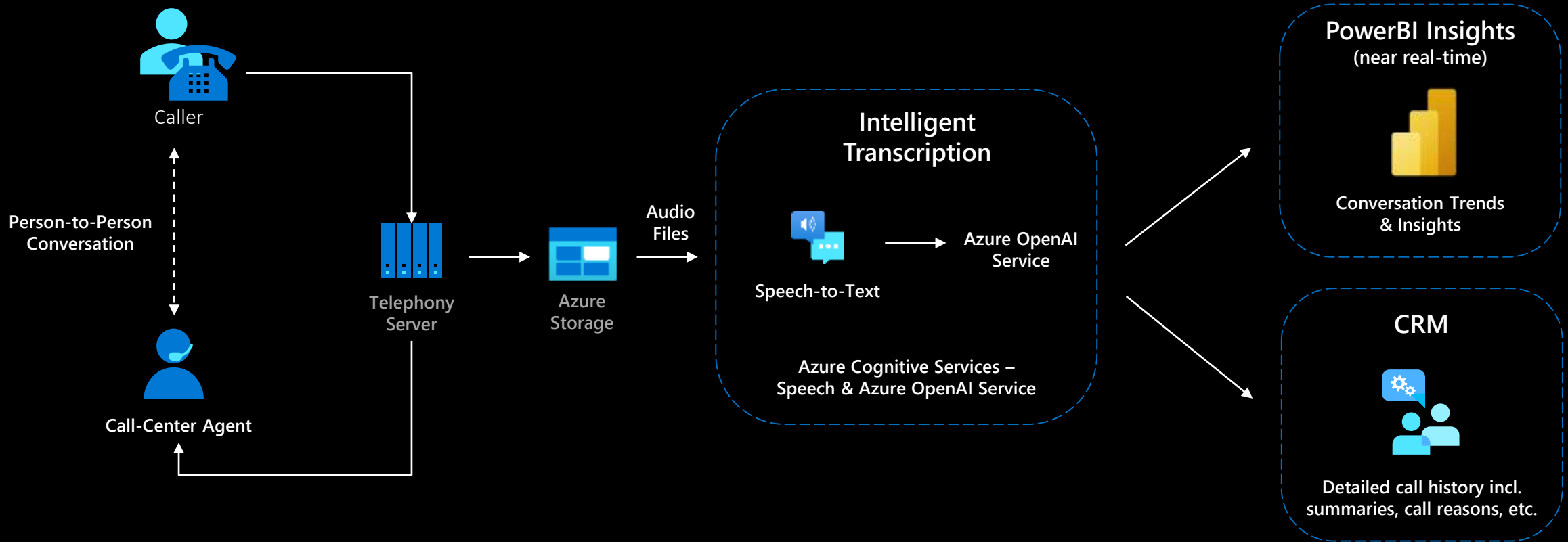- ## Text Rewriting / Paraphrasing
  - News article rewriting for SEO optimization
  - **Proof:** Several Media Companies

- ## Event Summarization
  - Post-game summaries from transcripts & automated reviews
  - Meeting summarization & action item extraction
  - **Proof:** Several Sports Companies

Contact Center Analytics using Speech API & Azure OpenAI Service

Extract rich insights from call transcripts

Azure OpenAI Service | Sample Use Cases

Playground

Deployments

text-davinci-002

Examples

Summarize Text

Code View

Azure OpenAI Service Instruction in English

You must extract the following information from the phone conversation below:

1. Call reason (key: reason)
2. Cause of the incident (key: cause)
3. Names of all drivers as an array (key: driver_names)
4. Insurance number (key: insurance_number)
5. Accident location (key: location)
6. Car damages as an array (key: damages)
7. A short, yet detailed summary (key: summary)

Make sure fields 1 to 6 are answered very short, e.g. for location just say the location name. Please answer in JSON machine-readable format, using the keys from above. Format the ouput as JSON object called "results". Pretty print the JSON and make sure that is properly closed at the end.

Customer conversation in English (or other language)

Phone conversation:
Hi I just had a car accident and wanted to report it. OK, I hope you're alright, what happened? I was driving on the I-18 and I hit another car. It all happened so quick. I got a bit shaken up. That's understandable. Can you give me your full name? Sure, it's Sarah standl. Do you know what caused the accident? I think I might have hit a pothole. OK, where did the accident take place? On the I-18 freeway. Was anyone else injured? I don't think so. But I'm not sure. OK, well we'll need to do an investigation. Can you give me the other drivers information? Sure, his name is John Radley. And your insurance number. OK. Give me a minute. OK, it's 546452. OK, what type of damages has the car? Headlights are broken and the airbags went off. Are you going to be able to drive it? I don't know. I'm going to have to have it towed. Well, we'll need to get it inspected. I'll go ahead and start the claim and we'll get everything sorted out. Thank you.

Output as JSON

```
results = {
    "reason": "car accident",
    "cause": "hit a pothole",
    "driver_names": ["Sarah Standl", "John Radley"],
    "insurance_number": 546452,
    "location": "I-18 freeway",
    "damages": ["headlights broken", "airbags went off"],
    "summary": "Sarah Standl was driving on the I-18 when she hit another car. She thinks she hit a pothole. John Radley was the other driver. There were no injuries but both cars have damage."
}
```

Call Center Automation

Example:
Azure OpenAI Servcice Prompt

How do I build something like ChatGPT that uses my own data?

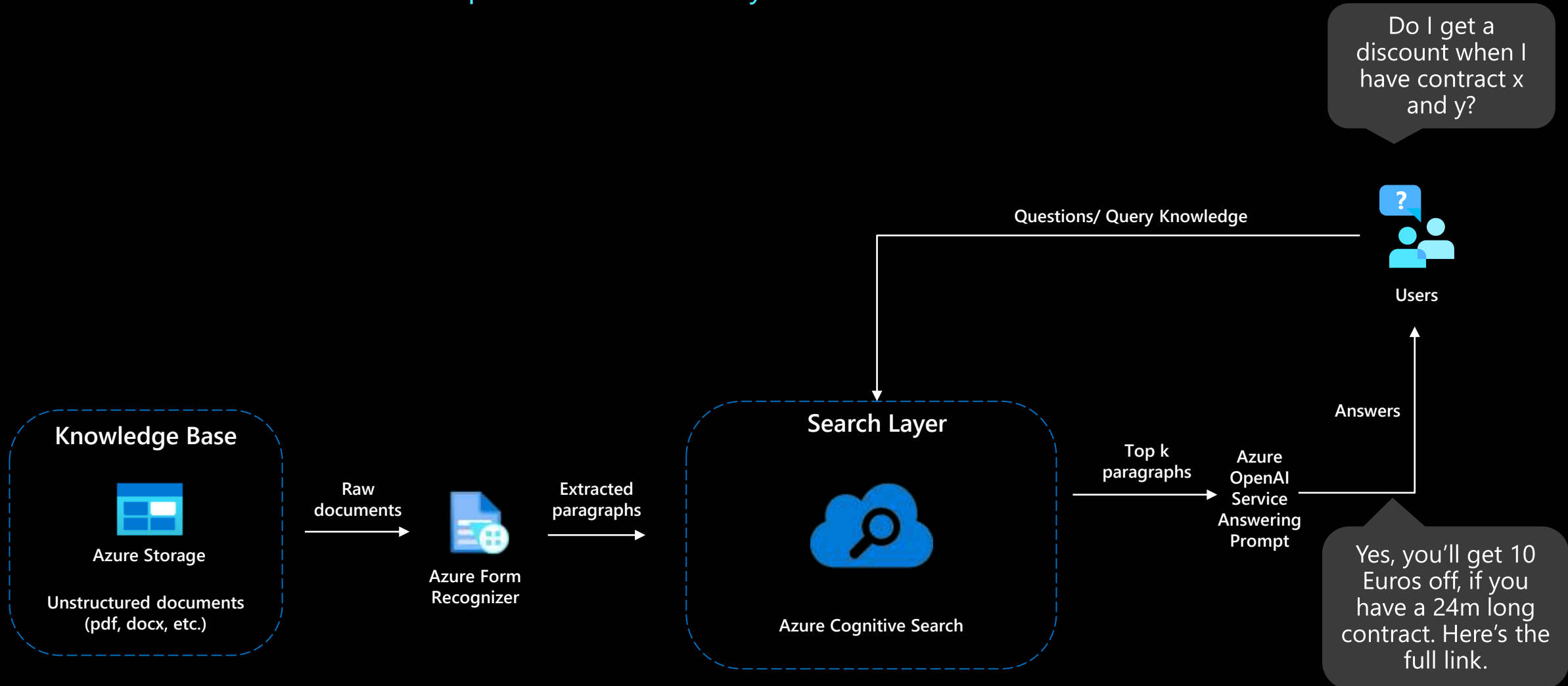**1** You can use ChatGPT out of the box

2 You have to fine-tune the model with your own data

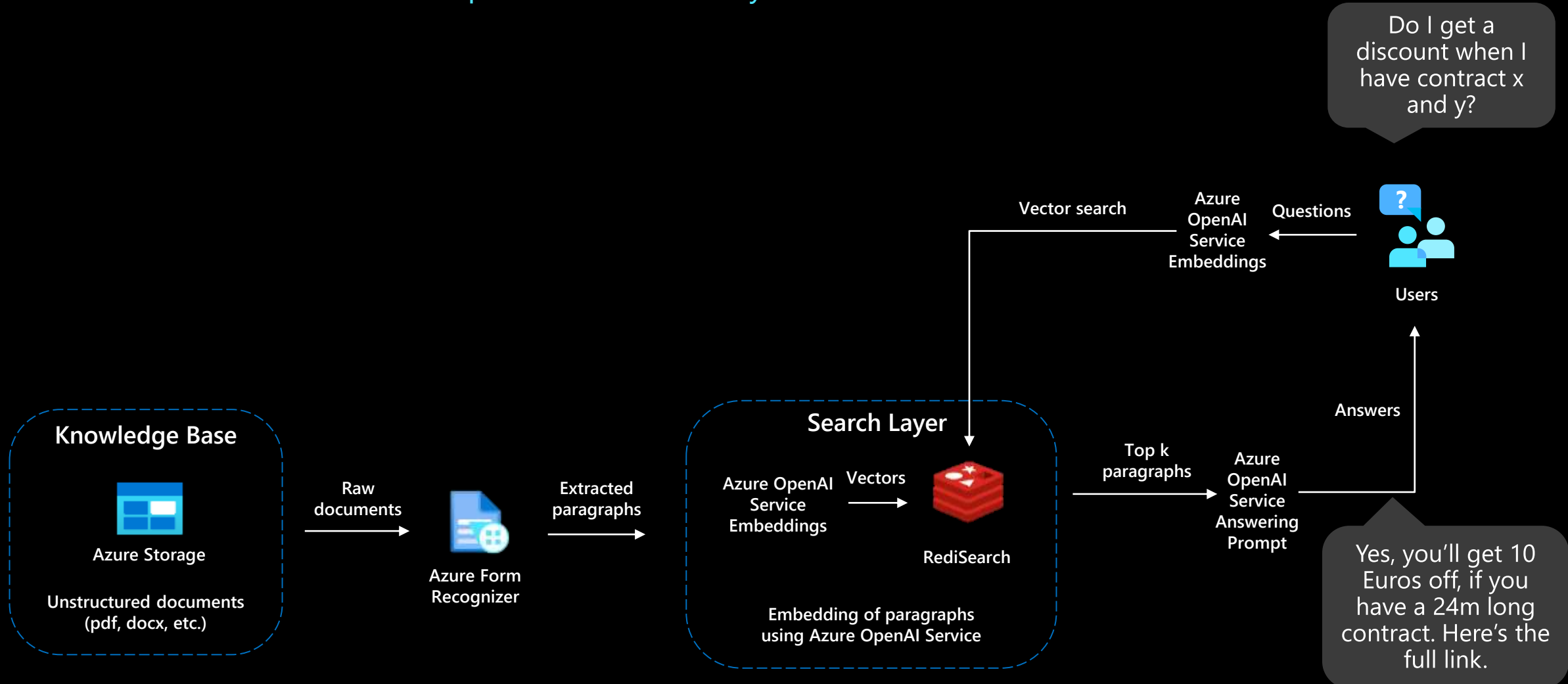**3** Inject the relevant information into ChatGPT along with the question

ChatGPT with Enterprise Data

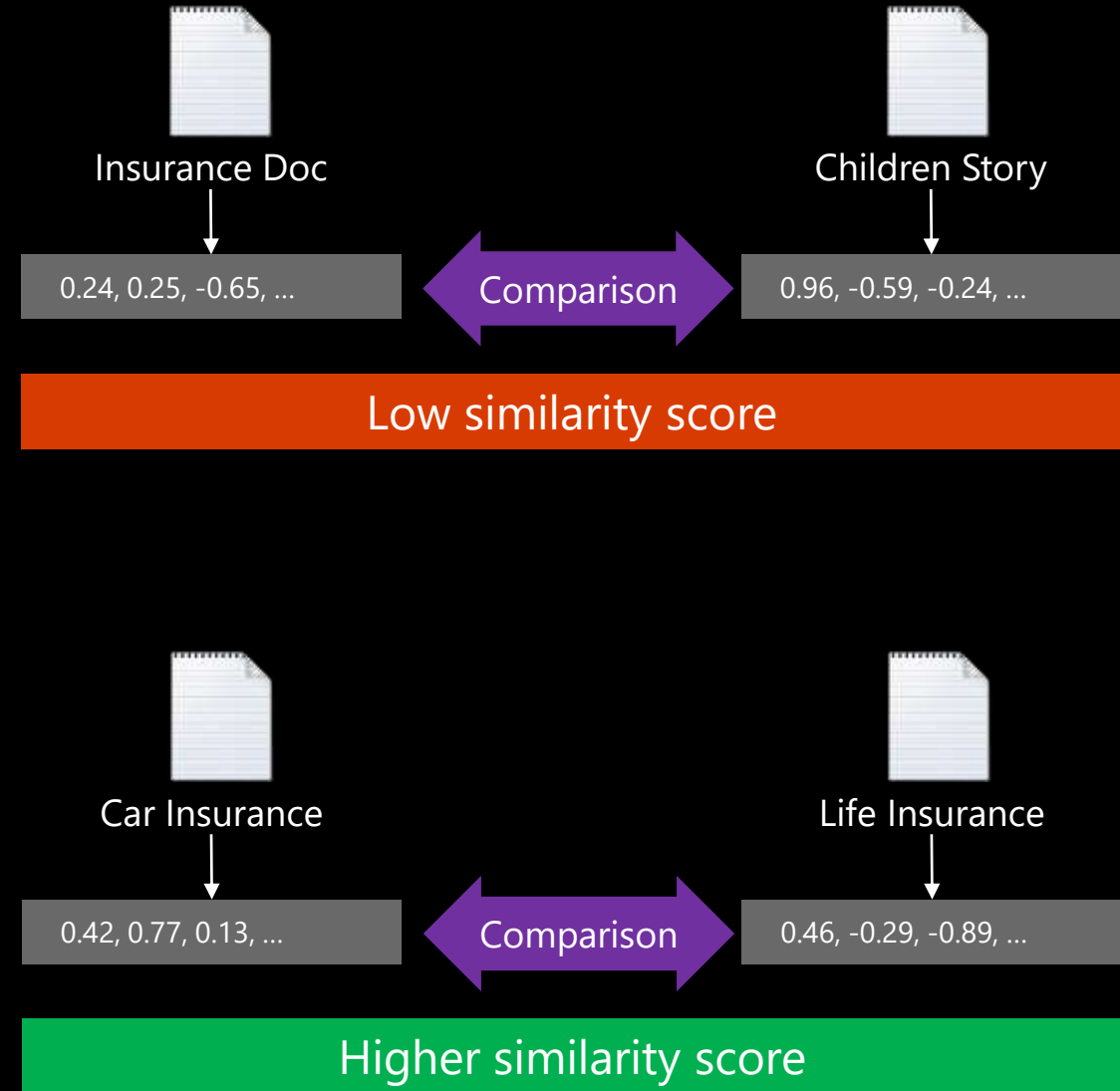## Extract precise answers from your unstructured documents

Do I get a discount when I have contract x and y?

Vector search

Azure OpenAI Service Embeddings

Questions

?

Users

Answers

Knowledge Base

Search Layer

Raw documents

Azure Storage

Unstructured documents (pdf, docx, etc.)

Azure Form Recognizer

Extracted paragraphs

Azure OpenAI Service Embeddings

Vectors

Top k paragraphs

Azure OpenAI Service Answering Prompt

RediSearch

Embedding of paragraphs using Azure OpenAI Service

Yes, you'll get 10 Euros off, if you have a 24m long contract. Here's the full link.

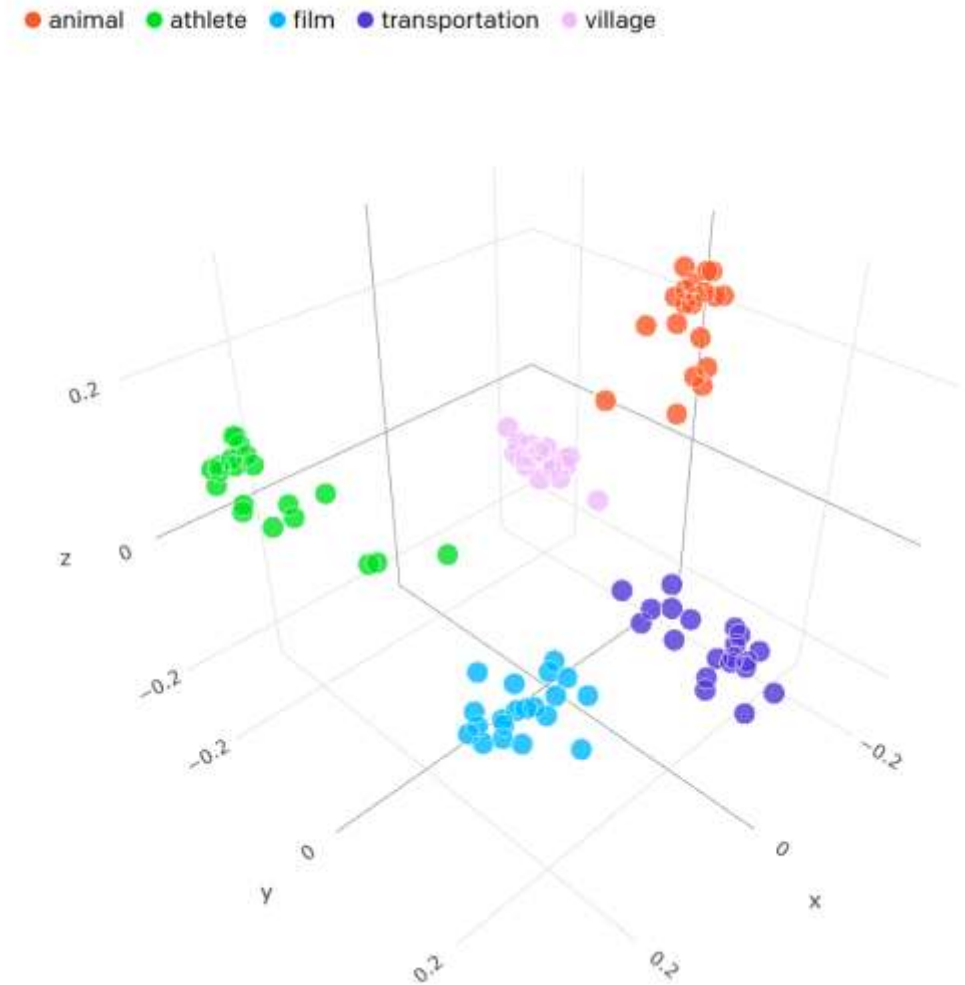Solution Accelerator

Demo Website

Azure OpenAI Service Embeddings

- Dense representation of semantic meaning of a piece of text
  - Allows to compare how similar two pieces of text are
  - Same idea as word2vec and others, but significantly better at semantic capturing

- A tool to built ChatGPT-like experiences on private data
- $1 for 3000-5000 pages of text

Insurance Doc → 0.24, 0.25, -0.65, … ← Comparison → Children Story → 0.96, -0.59, -0.24, …

Low similarity score

Car Insurance → 0.42, 0.77, 0.13, … ← Comparison → Life Insurance → 0.46, -0.29, -0.89, …

Higher similarity score

Use Cases for Embeddings

· Search
  · where results are ranked by relevance to a query string
· Clustering
  · where text strings are grouped by similarity
· Recommendations
  · where items with related text strings are recommended
· Anomaly detection
  · where outliers with little relatedness are identified
· Diversity measurement
  · where similarity distributions are analyzed
· Classification
  · where text strings are classified by their most similar label

Should I **fine-tune GPT-3** with my data?

· Yes, but only for these use cases:
  · Classification
  · Conditional Generation
    · Answering in a specific "style"
  · Both are supervised training tasks!

· Just throwing in your data (unsupervised) likely **won't** work
  · Even if you had a lot of data, it won't change the model's weights enough...

JSON

{"prompt": "<prompt text>", "completion": "<ideal generated text>"}
{"prompt": "<prompt text>", "completion": "<ideal generated text>"}
{"prompt": "<prompt text>", "completion": "<ideal generated text>"}

MVP Development

Partnering with Microsoft to tackle the "Use Case Flood"

1 **Build up Azure OpenAI Service CoE**

- Central point of contact for Azure OpenAI Service use cases
- Helps internal teams to get started

2 **Enablement**

- Microsoft can provide workshops for CoE (incl. train the trainer)
- Scale internal knowledge

3 **Fast evaluation**

- Goal: figure out which of the use cases is a low-hanging fruit & has high ROI
- "Speed dating"
- Microsoft can help

Azure OpenAI Service MVP Development

- Procedure
  - Acquire data for project
    - Typically, mostly evaluation data
    - Training data only required for few use cases
    - Timeline: fully depends on customer
  - Validate idea at small scale as quickly as possible
    - In Azure OpenAI Studio or Jupyter notebook
    - Timeline: 1-5 days (depends on use case)
  - Evaluate idea at larger scale
    - Use annotated data for evaluation or human review
    - Timeline: 5-10 days (depends on use case)
  - Build MVP around it
    - UI, API, security, user access, etc.
    - Timeline: 5-20 days (depends on use case)

- Skills needed
  - Data Scientist/Developer
    - Understanding of Azure OpenAI Service
    - Can write Python
  - Developer for UI
    - Typical skillset for webdev (use streamlit for MVPs)
  - Azure developer
    - Azure service configuration, devops automation
- Best practices
  - Re-use existing code from samples
    - Azure OpenAI Service documentation, Cookbooks
  - Complete idea validation & at scale testing as quickly as possible
    - Fail early if needed

GPT Customization

- Try zero-shot with Davinci (prompt engineering)
  - If it works, great, try smaller models to save $
  - Consider fine-tuning if use case allows it to save $ (might require bigger scale)
- Try few-shots with Davinci (if use case allows it)
  - If it works, great, try smaller models to save $
  - Consider fine-tuning if use case allows it to save $
- Try embeddings or different model for fine-tuning

Reference: How to customize a model with Azure OpenAI - Azure OpenAI | Microsoft Learn

Q&A

Break

Hands-On

- Quick start
  - First, do the short quick start example
  - [azure-openai-in-a-day-workshop/quickstart.ipynb at main · microsoft/azure-openai-in-a-day-workshop (github.com)](#)
- Q&A
  - You'll be completing a simple Q&A example in Python
  - [azure-openai-in-a-day-workshop/qna_with_embeddings_exercise.ipynb at main · microsoft/azure-openai-in-a-day-workshop (github.com)](#)
- Embeddings
  - You'll be building an unsupervised movie clustering example and recommendation engine in Python
  - [azure-openai-in-a-day-workshop/movie_classification_unsupervised_incl_recommendations_exercise.ipynb at main · microsoft/azure-openai-in-a-day-workshop (github.com)](#)

Microsoft

Thank you!

Microsoft AI