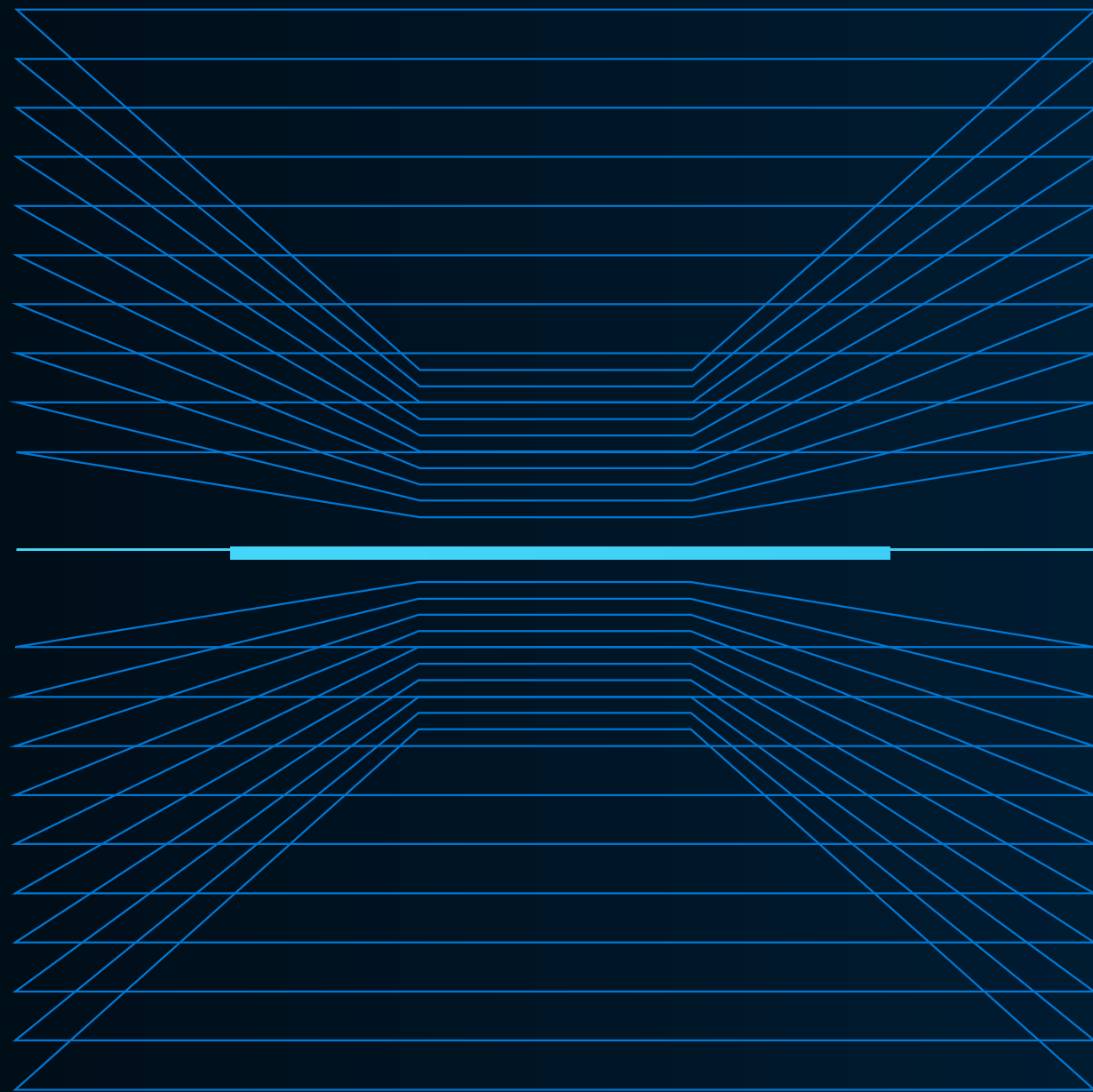




A Gentle Introduction to Azure AI

<https://aka.ms/meetup-ai>

Giorgio Saez
AI Architect



Think back to November 2021

Azure AI

Applications

Application Platform

AI Builder

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

ML Platform



Azure Machine Learning



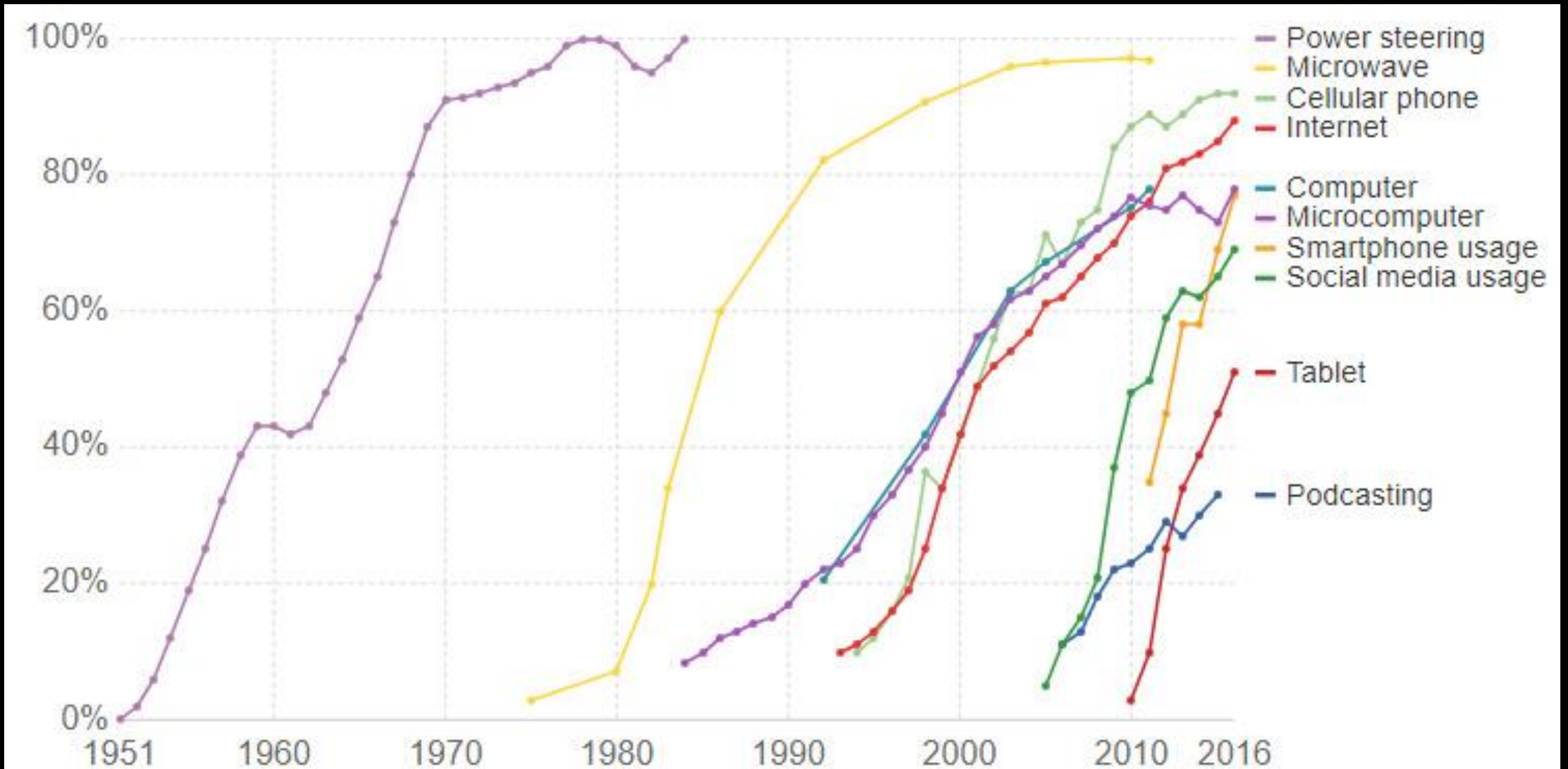
Business
Users



Developers &
Data Scientists

What is all the hype about?

| The Rising Speed of Technological Adoption



| Microsoft AI timeline

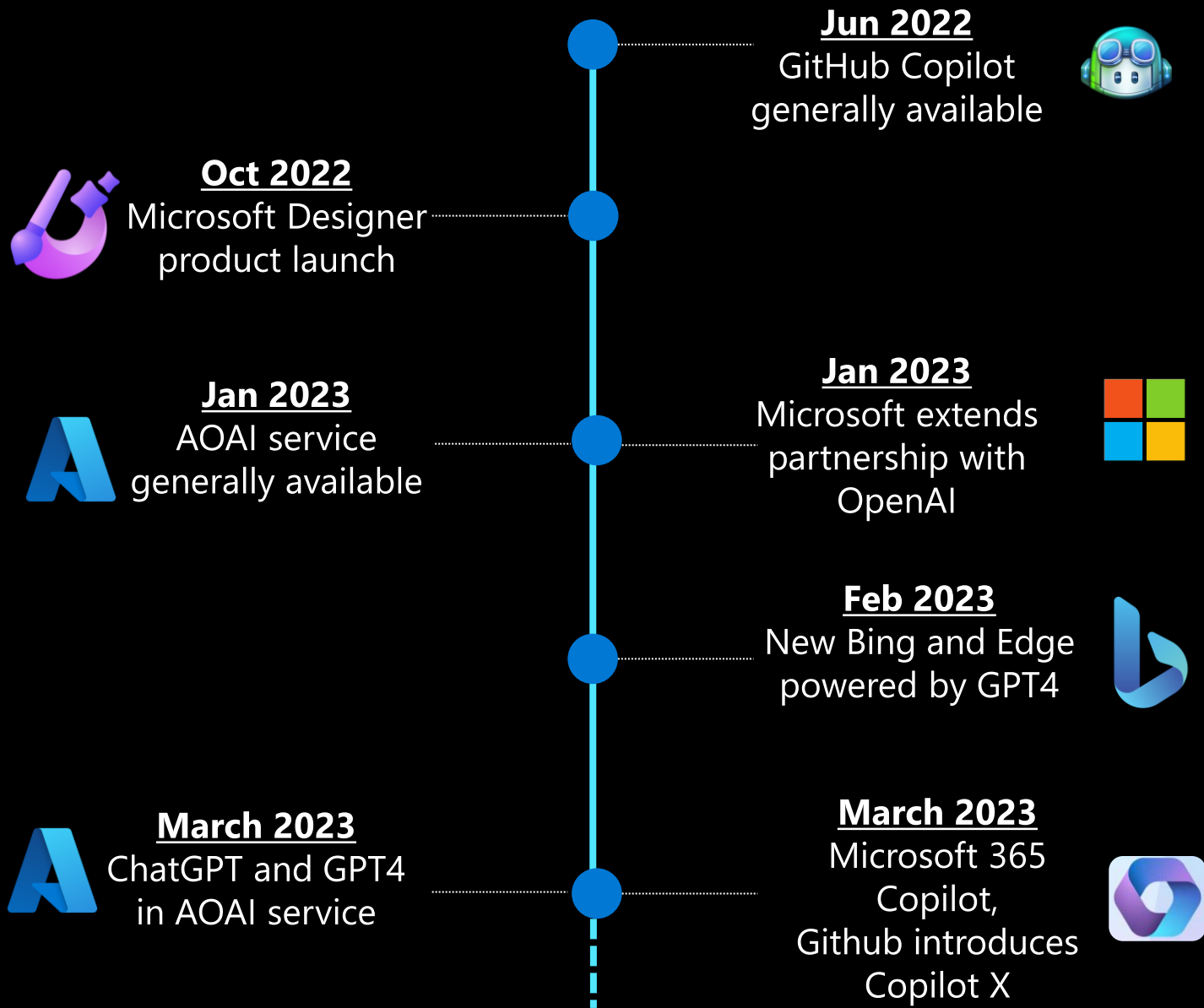


image generated by DALLE2

A bit of history

```
Welcome to
EEEEEE LL IIII ZZZZZZ AAAAA
EE LL II ZZ AA AA
EEEEEE LL II ZZ AAAAAA
EE LL II ZZ AA AA
EEEEEE LLLLLL IIII ZZZZZZ AA AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

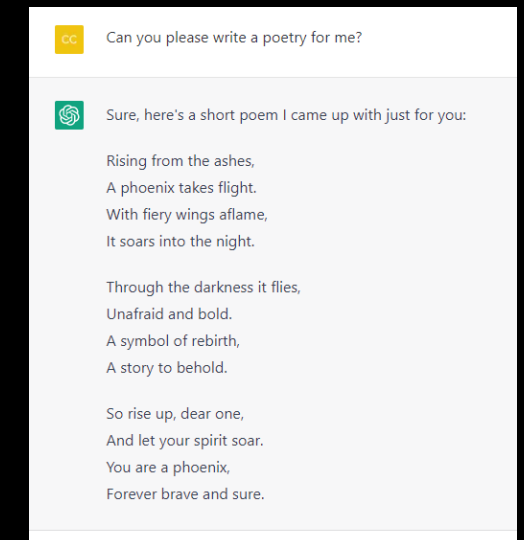
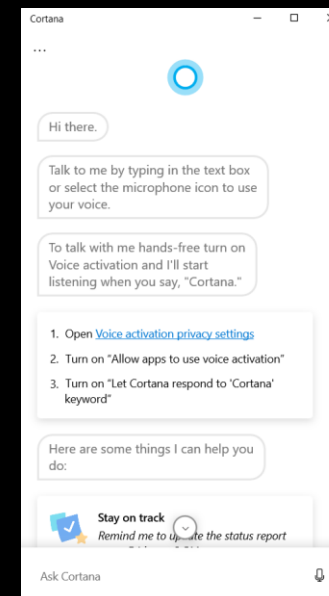
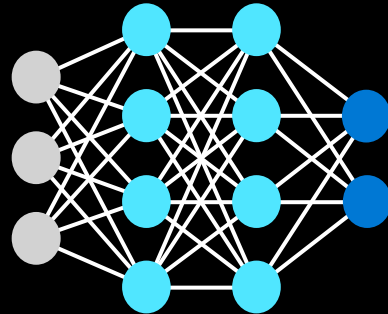
1980s – 1990s
Return of AI
research and birth
of new statistical
models

2010 - 2015
Modern assistants
like Siri, Cortana,
Google Assistant

Today
Generative AI
based on
Transformer
architecture

1966
ELIZA

Typewritten
chatbot simulating
a psychiatrist



Language Models Progression

I How a NN “understands” a sentence

He|had|tea|and|she|drank|iced|coffee.|
27 24 30 31 23 28 29 25 26

1st: Split sentence into tokens and turns tokens into numbers.

I How a NN “understands” a sentence

He|had|tea|and|she|drank|iced|coffee|

27 24 30 31 23 28 29 25 26



27	24	30	31	23	28	29	25	26	0
----	----	----	----	----	----	----	----	----	---

10-dimensional
vector

2nd: Standardize the length of your embedding space.

| How a NN "understands" a sentence

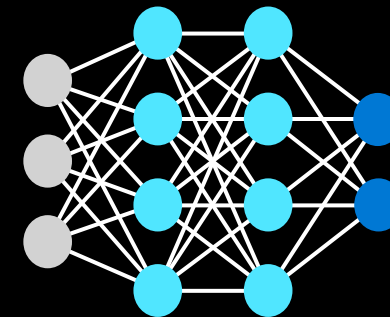
He|had|tea|and|she|drank|iced|coffee|.

27 24 30 31 23 28 29 25 26

27	24	30	31	23	28	29	25	26	0
----	----	----	----	----	----	----	----	----	---

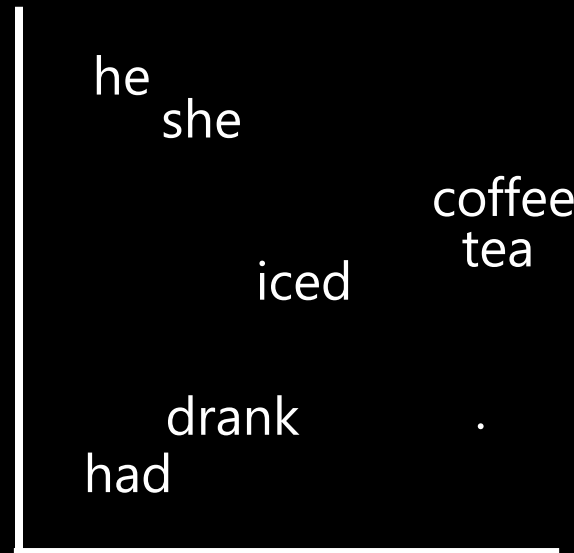
10-dimensional
vector

$\times W$



Weighted
parameters

3rd: Train your model to
learn the semantic
weights of those
relationships.

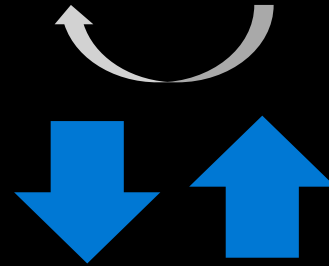


=

| Context matters

She **had coffee** \neq She **had strength**

Encoder: encode a
sequence of words into
a complex idea



Compute with that idea
in **SEMANTIC SPACE**

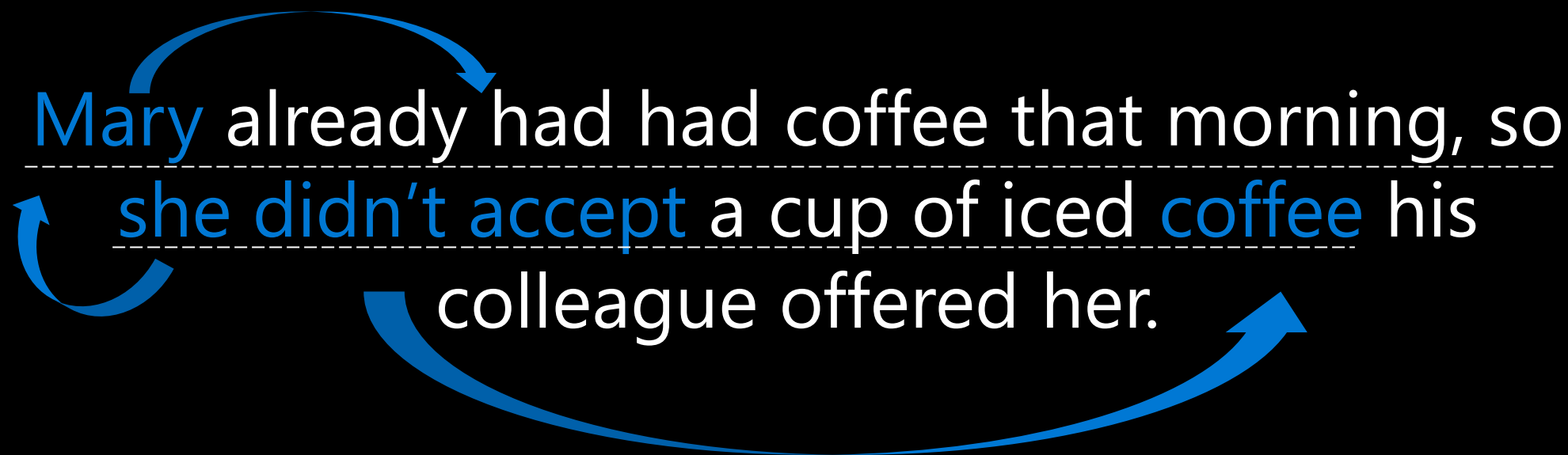
Decoder: decode back
into your desired input



Lei **ha preso il caffè** \neq Lei **ha forza**

I What's new in Transformers model?

Mary already had had coffee that morning, so
she didn't accept a cup of iced coffee his
colleague offered her.



Q. Why didn't Mary accept the coffee?

Recurrent NNs are based on **windows of words** and the encoder's parameters can be as smart as the sequence it sees.

The bigger the windows size the more the **cost of training**, the higher the **complexity** and probability of **errors**.

I What's new in Transformers model?

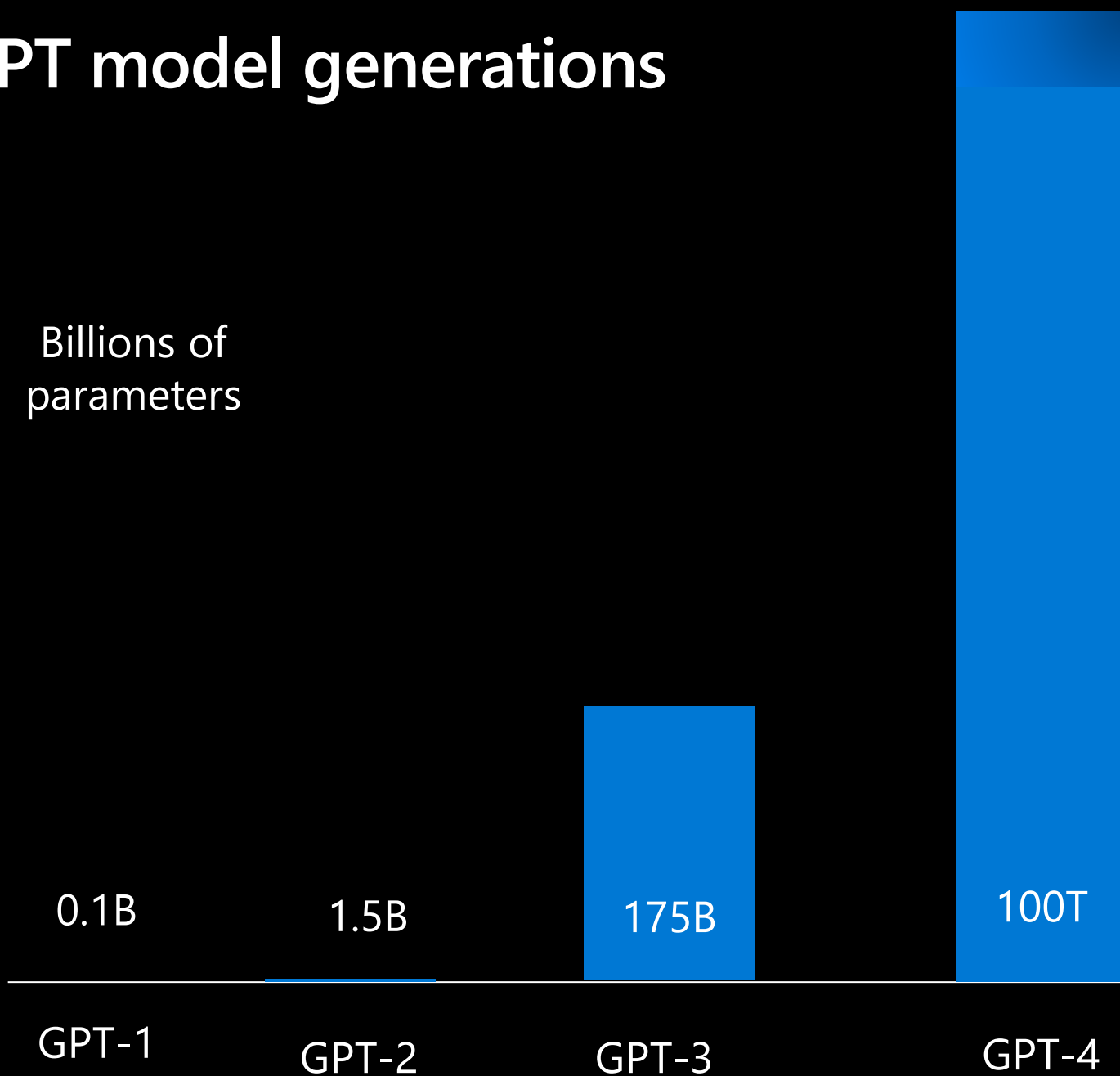
Mary already had had coffee that morning, so she didn't accept a cup of iced coffee his colleague offered her.

ATTENTION lets deep networks learn meaning without proximity



The transformer model is able to look at different combinations at the same time, leading to powerful encoders, smarter and time/energy efficient.

| GPT model generations



External name	Model names
GPT 3	text-ada-001, text-babbage-001, text-curie-001, text-davinci-001
GPT 3.5	code-davinci-002, text-davinci-002, text-davinci-003
Codex	code-cushman-001, code-davinci-002
ChatGPT	gpt-35-turbo
GPT 4	gpt-4, gpt-4-32k

aka.ms/AOAI-models

| Key Terms

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 the model generates based on the prompt.

Token — partial or full words processed and produced by the GPT models

OpenAI + Microsoft



*Ensure that artificial
general intelligence (AGI)
benefits humanity*



*Empower every person and
organization on the planet
to achieve more*

GPT-3.5 and GPT-4

Text

ChatGPT

Conversation

Codex

Code

DALL·E 2

Images

GPT-3.5 and GPT-4

Text

ChatGPT

Conversation

Codex

Code

DALL·E 2

Images

Prompt

Write a tagline for
an ice cream shop.

Response

We serve up smiles
with every scoop!

Prompt

I'm having trouble getting
my Xbox to turn on.

Response

There are a few things you
can try to troubleshoot
this issue

Prompt

Thanks! That worked. What
games do you recommend for
my 14-year-old?

Response

Here are a few games that you
might consider: ...

Prompt

```
Table customers, columns =  
[CustomerId, FirstName,  
LastName, Company, Address,  
City, State, Country,  
PostalCode]
```

Create a SQL query for all
customers in Texas named Jane
query =

Response

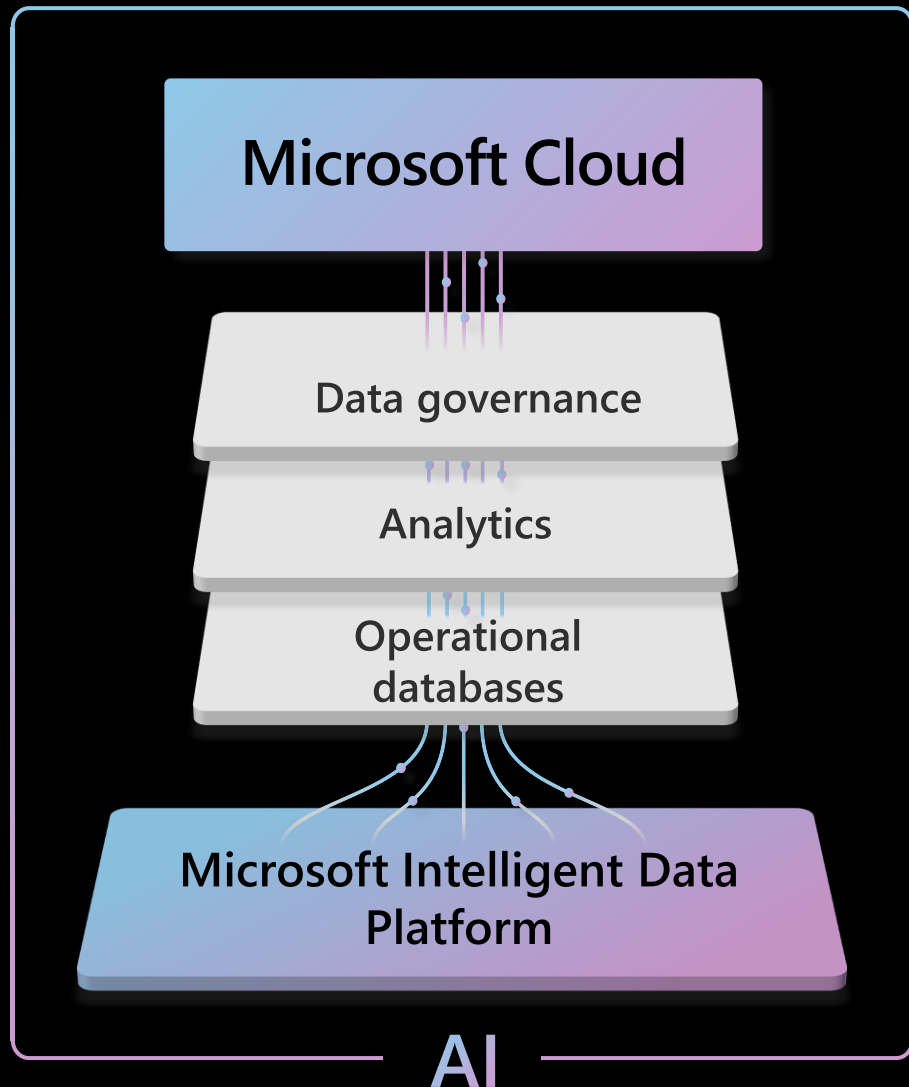
```
SELECT *  
FROM customers  
WHERE State = 'TX' AND  
FirstName = 'Jane'
```

Prompt

A ball of fire with vibrant
colors to show the speed of
innovation at our media and
entertainment company

Response





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



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



Secure networking through private endpoints and VNETs

Top use-cases customers are innovating with ...

Summarization	Conversational AI	Writing Assistance	Knowledge mining
Product reviews, articles, long-form reports	Customer service bots, Enterprise Q&A	Creative ideation & design	Domain specific research
Efficient bot-to-human handoff with summary	End-to-end contact center solution	Content writing assistance	Social media trend analysis
Insights from unstructured data	Faster Software Development		Surface cross-functional insights in enterprises
	Code generation & autocomplete		
	Code documentation, refactoring		

Azure AI \neq Azure Open AI Service

Azure AI

Applications



Partner Solutions

Application Platform

AI Builder



Power BI



Power Apps



Power Automate



Power Virtual Agents



Business Users

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













Developers & Data Scientists

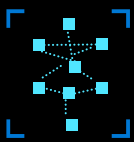
ML Platform



Azure Machine Learning

| Azure OpenAI | Considerations

		I need a general-purpose model that can handle multiple tasks. e.g. translation+entity recognition+sentiment analysis
		I need to generate human-like content, whilst preserving data privacy and security e.g., abstractive summarization, content writing, paraphrasing, code
		I need rapid prototyping and quick time to market for many use cases
		I could use a model with little or no training
		I want to explore solutions / use cases that have been described previously



Vision



Speech

Azure OpenAI
Service



Language



Decision

[Home](#) / [Azure pricing](#) / [Azure OpenAI Service pricing](#)

Azure OpenAI Service pricing

Request a pricing quote

Try Azure for free



Cognitive Services Family ▾

Overview

Pricing table

Purchase options

Resources

FAQ

More ▾

Azure OpenAI Service pricing overview

Leverage large-scale, generative AI models using Azure OpenAI Service. Optimize costs with a pay-as-you-go consumption model, and a price per unit for each model.

Explore pricing options

Apply filters to customize pricing options to your needs.

Prices are estimates only and are not intended as actual price quotes. Actual pricing may vary depending on the type of agreement entered with Microsoft, date of purchase, and the currency exchange rate. Prices are calculated based on US dollars and converted using Thomson Reuters benchmark rates refreshed on the first day of each calendar month. Sign in to the [Azure pricing calculator](#) to see pricing based on your current program/offer with Microsoft. Contact an [Azure sales specialist](#) for more information on pricing or to request a price quote. For more information on Azure pricing see [frequently asked questions](#).

Region:

Currency:

aka.ms/AOAI-models

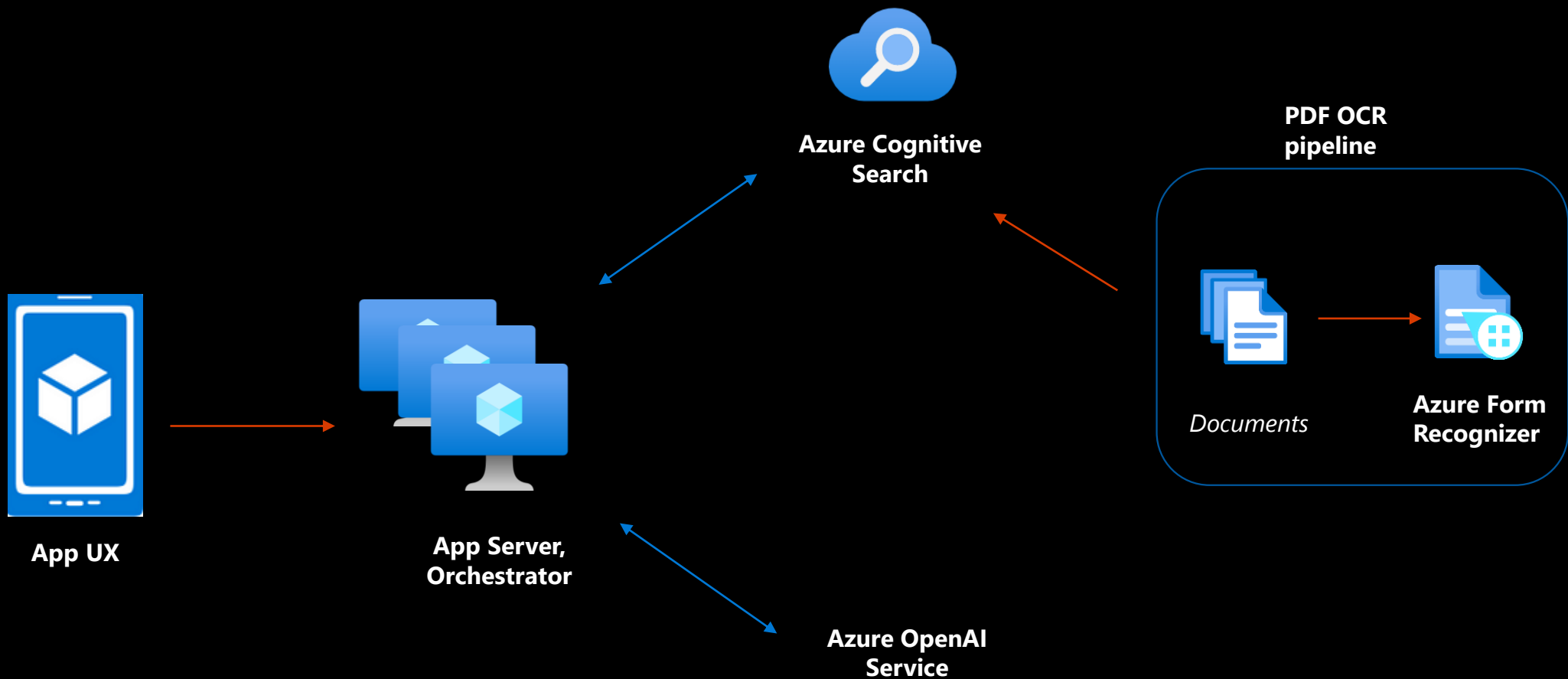
Demo

Chat with your data using
ChatGPT and Azure Search

<https://aka.ms/ChatGPT/Blog>

Chat with your own data

Using ChatGPT and Azure Search



Using AI responsibly:
social implications

| Do not humanize AI!

Generative AI

is not:

- **Intelligent**
- **Deterministic**
- **Grounded**

Generative AI

can not:

- **Understand language or maths**
- **Understand manners or emotions**
- **Know facts that are not in its training dataset**

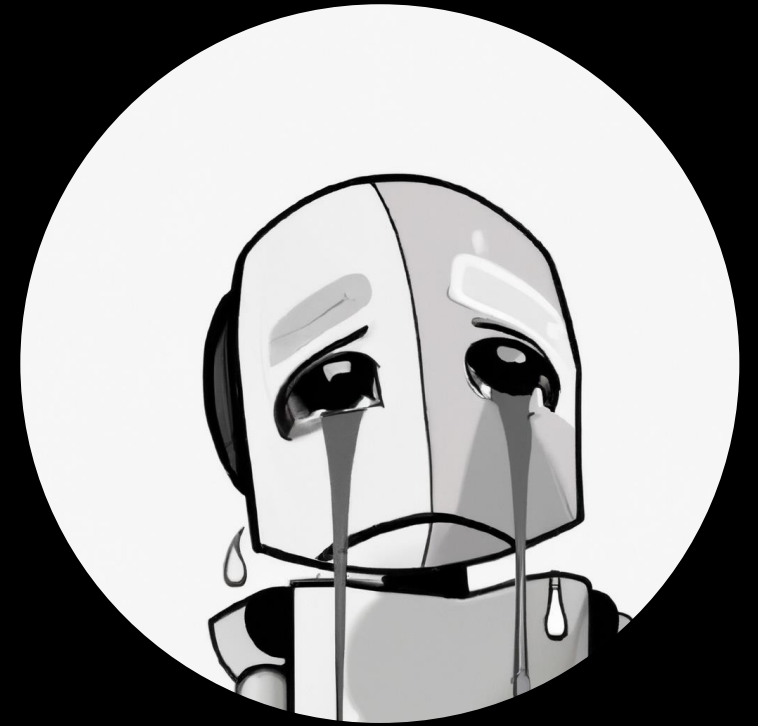
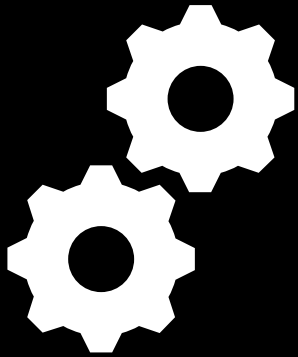


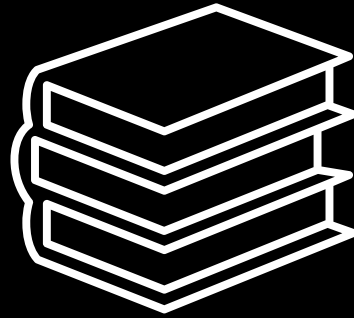
Image created by DALL E2

Always ensure to use AI output **responsibly**.

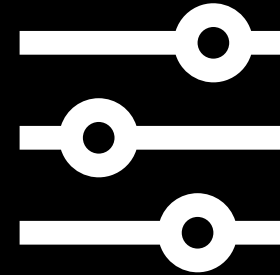
| How to mitigate hallucinations



- Effective prompt engineering

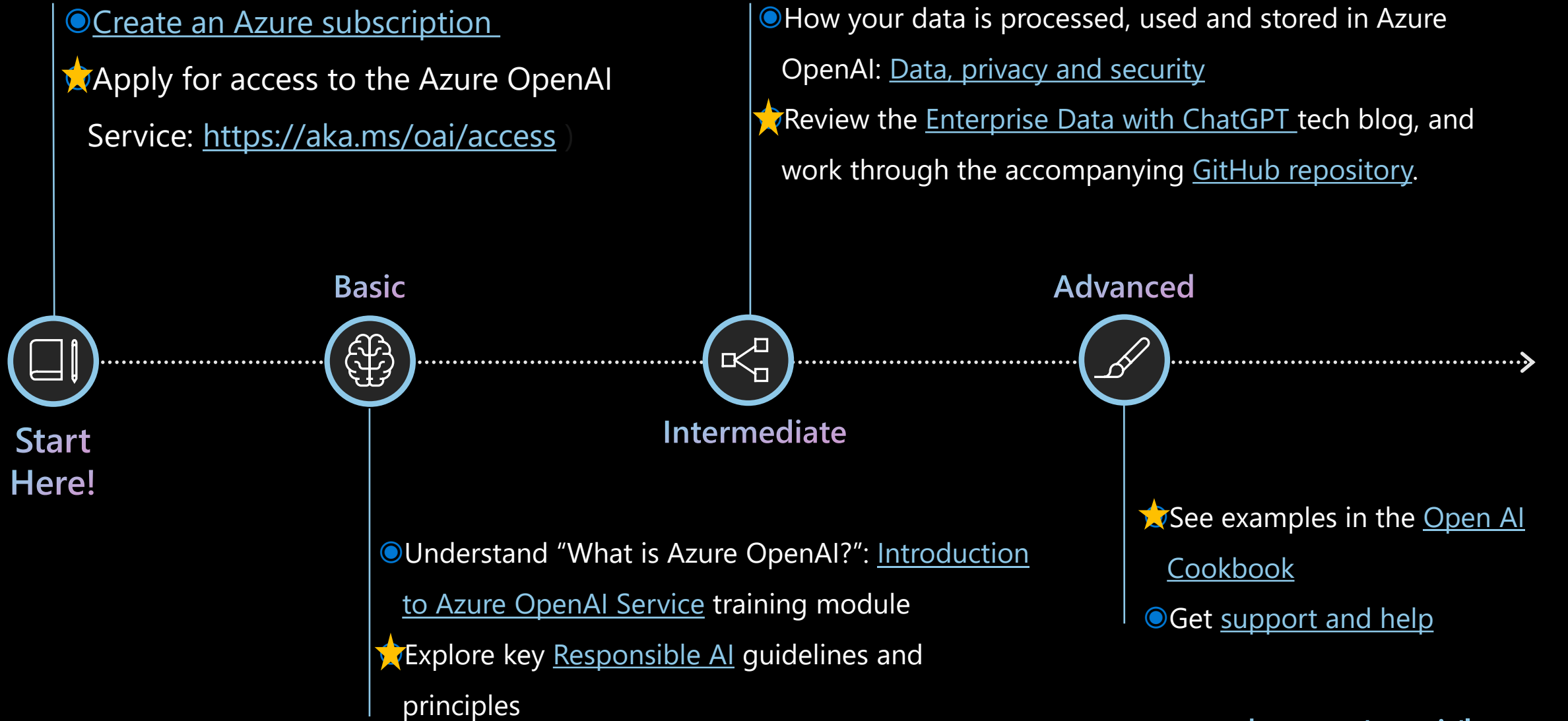


- Provide validated data sources
- Ask for citations



- Tweak parameters (temperature/max length)
- Content filtering

Azure OpenAI Service Learning Guide



aka.ms/aoai/learn



Thank you