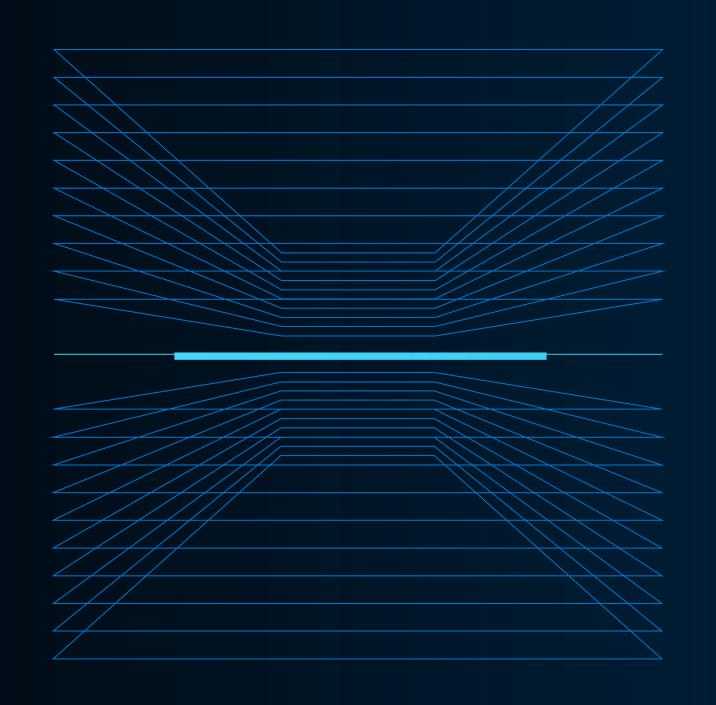


Getting start with Azure OpenAl

Sergio Zenatti Filho Sr Cloud Solution Architect



Microsoft



Sergio Zenatti Filho

Senior Cloud Solution Architect at Microsoft

Sergio has over 20 years of experience designing and delivering Data and Analytics Solutions. He has extensive experience in the Microsoft Data and Analytics Platform in the cloud and also on-premises. Sergio is passionate about learning new technology and helping customers to define the best solution for their business.

Connect



in /sergiozenatti



Agenda

- Azure OpenAl Overview
- Why Azure OpenAl
- Concepts
- Demos
- Use Cases
- Next Steps
- Q&A



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

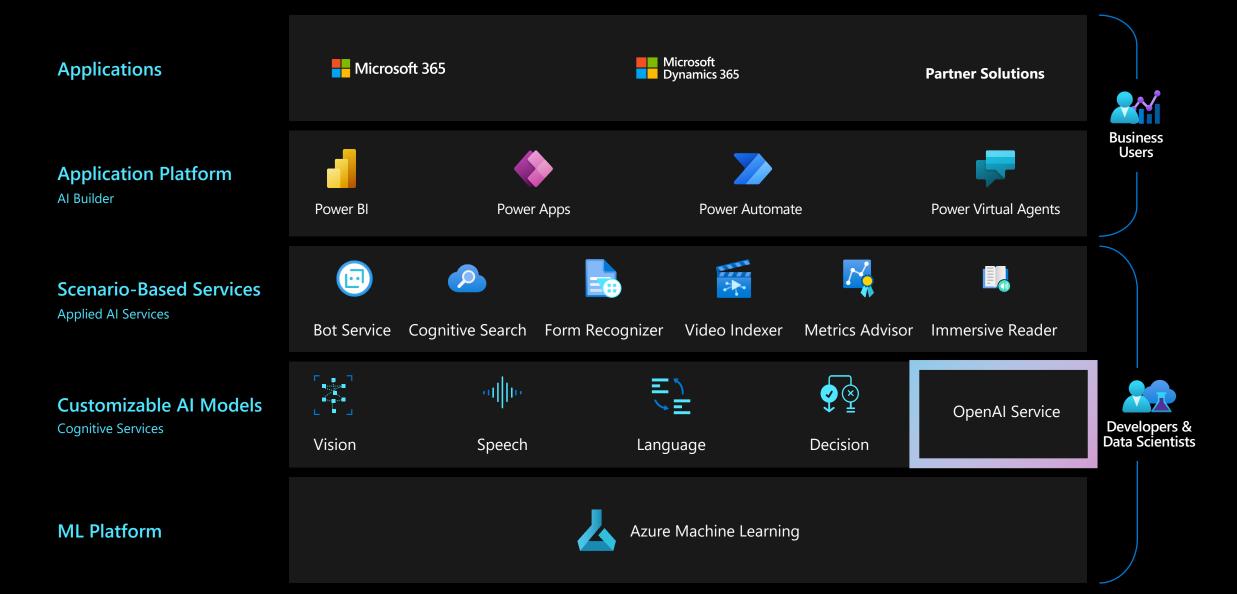
Codex

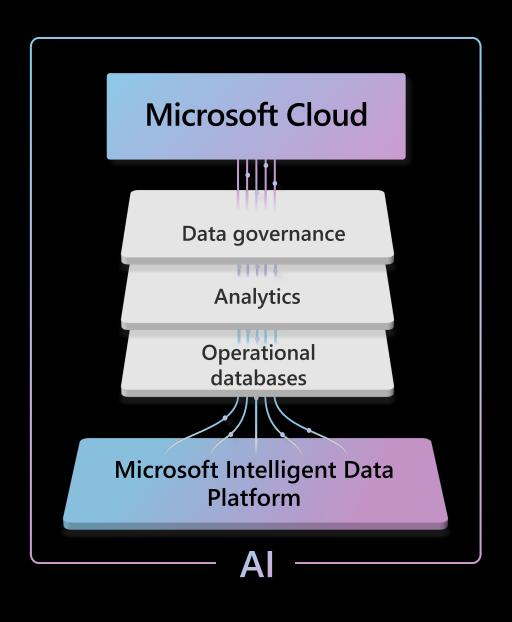
Code

DALL·E 2

Images

Azure Al







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

Generative Al Content creation by API

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



Prompt and Completion

A short piece of text used to initiate an AI language model to generate a longer piece of text.

In computer science, the time complexity is the computational complexity that describes the amount of computer time it takes to run an algorithm. Time complexity is commonly estimated by counting the number of elementary operations performed by the algorithm, supposing that each elementary operation takes a fixed amount of time to perform. Thus, the amount of time taken and the number of elementary operations performed by the algorithm are taken to be related by a constant factor [...]

Time complexity is a measure of the amount of time it takes for an algorithm to run, and is usually estimated by counting the number of elementary operations performed. It is usually expressed as a function of the size of the input, and the asymptotic behavior of the complexity is often focused on.

Best Practices

Be Specific

Be Descriptive

Double Down

Order Matters

Alternative Path

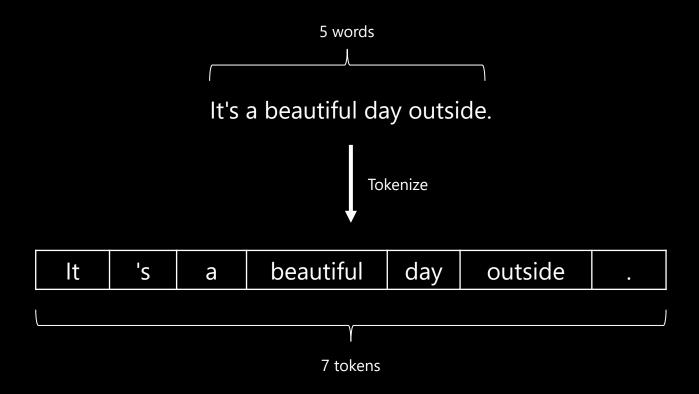
Summarize the previous text in a 2 sentences:

Prompt

Completion

Tokens

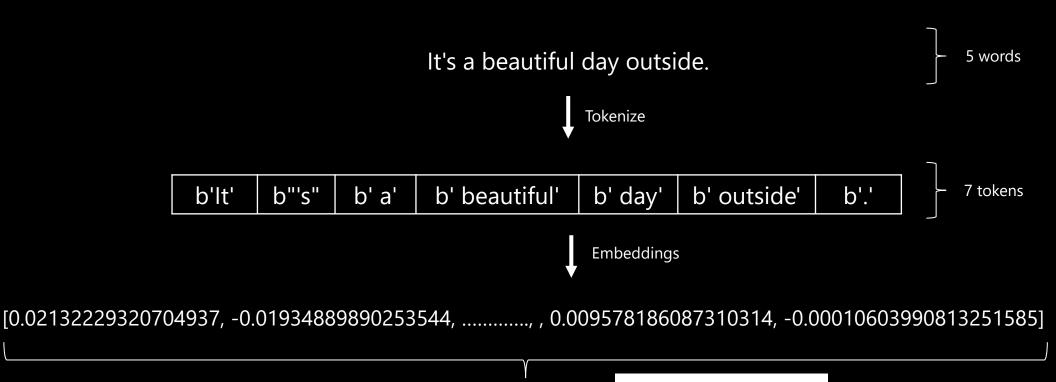
Smaller chunks of text created by splitting input text.





Embeddings

Representation of tokens as numerical vectors for machine learning algorithms to process.



Cosine Similarity

Metric used to measure similarity between two documents / text contents.

It's a beautiful day outside.

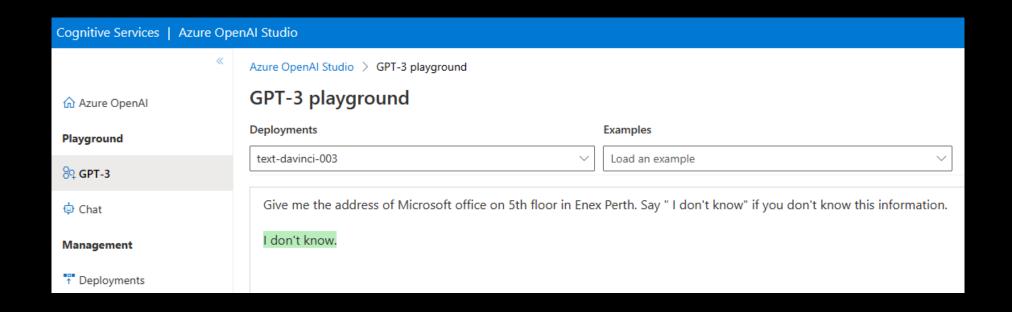
[......]

Cosine Similarity(v1, v2) = 0.92

The sun is shining brightly, and it is a perfect day for taking a stroll in the park.

Hallucinations

A confident response that does not seem to be justified by the model's training data - Wikipedia



OpenAl Models As of March 2023

Different pre-trained models, grouped by family and capability.

More accurate, difficult problems	Understand and generate natural language	Generate code	Create embeddings
GPT-4 (Preview)	<u>GPT-3</u>	<u>Codex</u>	<u>Embeddings</u>
gpt-4 gpt-4-32k	text-davinci-003 text-curie-001 text-babbage-001 text-ada-001 gpt-35-turbo (Preview)	code-davinci-002 code-cushman-001	text-similarity-ada-001 text-similarity-babbage-001 text-similarity-curie-001 text-similarity-davinci-001 text-search-ada-doc-001 text-search-ada-query-001 text-search-babbage-doc-001 code-search-ada-code-001 code-search-ada-text-001

Model Parameters

User configurable settings that influence the model output.

Temperature, Top probabilities

Regulate the level of determinism in the model's response generation process.

Max response

No. of tokens to return as model response.

Frequency penalty

Control the probability of generating identical text in the model's response.

Presence penalty

Control token repetition to promote new topic introduction in responses.

Best of

Produce several responses and select the one with the highest overall probability of its tokens for display.

Zero-shot, Few Shot Learning

methods of using a model without any fine-tuning or additional training.

Zero-shot

Convert movie titles into emoji:

Spider-Man: Into the Spider-Verse

Result

Spider-Man: 💥 🔼

Few-shot

Convert movie titles into emoji.

Back to the Future: 😨 🐵 🚜 🕒









Batman: 👗 🏄





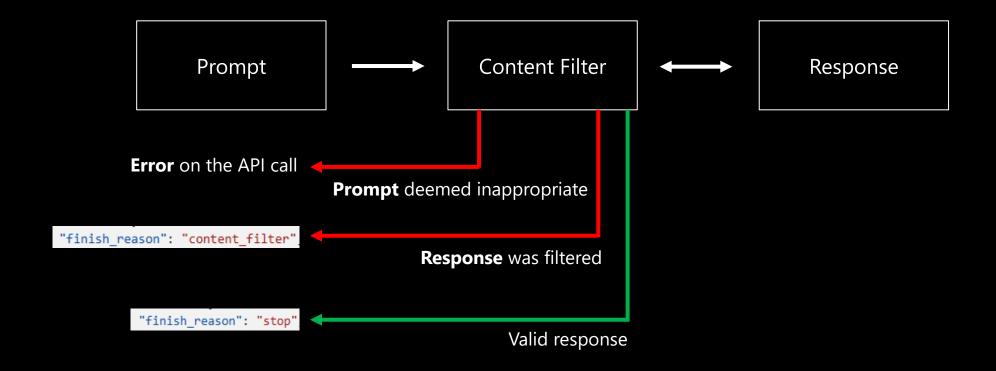
Transformers: 🚜 😨





Content filtering

Task of detecting misuse by analyzing prompt and response through classification models.

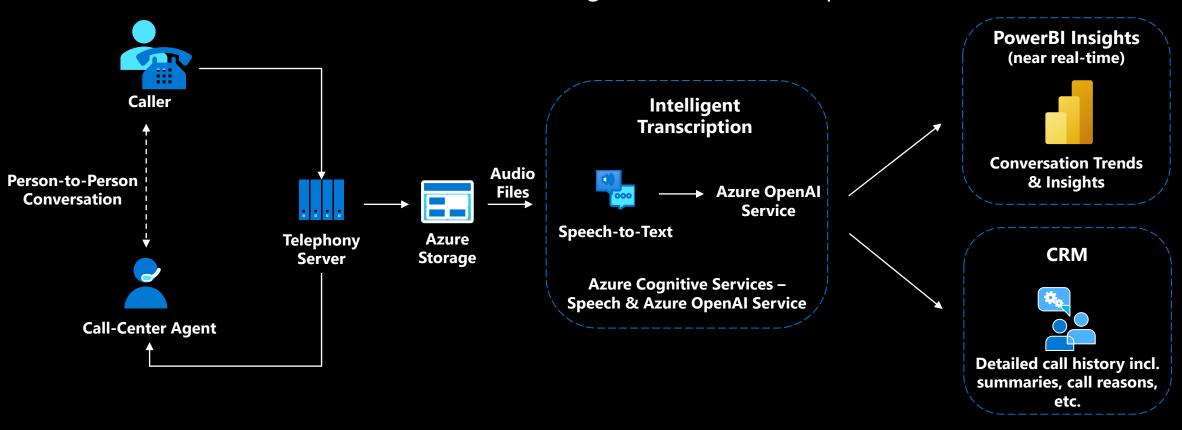




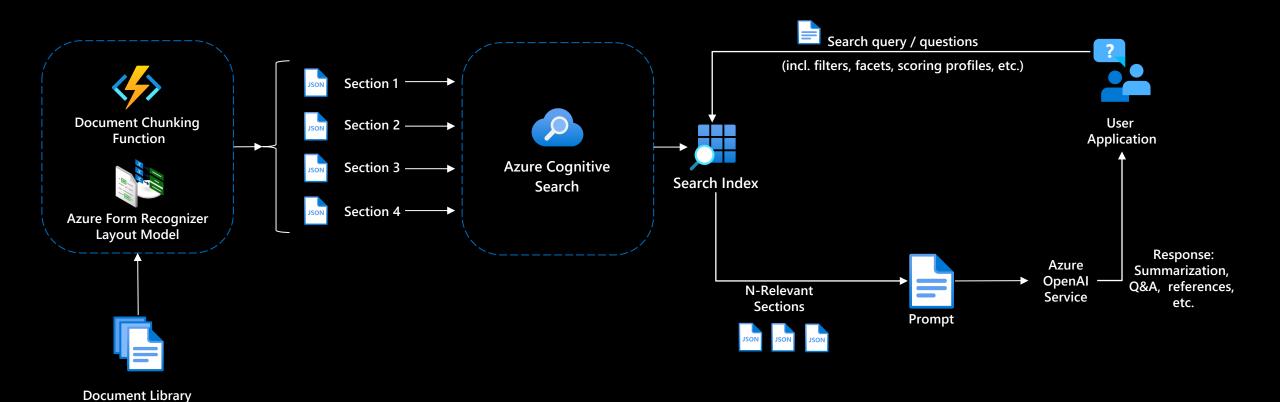
Demo – Azure OpenAl

Contact Center Analytics using Speech API & Azure OpenAI Service

Extract rich insights from call transcripts



Al-Powered Q&A over Enterprise Data Sources



Azure OpenAl Service Learning Guide

- Create an Azure subscription
- Apply for access to the Azure OpenAl Service: https://aka.ms/oai/access
- Watch the video: <u>An Introduction to Azure OpenAl</u>
- Get Familiarized with the product content: <u>Azure OpenAl Page</u>
- Review how other <u>customers are using Azure Al</u>

- | How-to guides: Generate embeddings and Fine-tune
- Complete embeddings and document search tutorial
- How your/your customer data is processed, used and stored in Azure OpenAI:
 Data, privacy and security
- Review the Enterprise Data with ChatGPT tech blog, and work through the accompanying GitHub repository.
- Become familiar with the APIs

Basic







Advanced

Start Here!

Intermediate

- Understand "What is Azure OpenAl?": <u>compare Azure OpenAl and OpenAl.</u> Review key concepts.
- Work through Introduction to Azure OpenAl Service training module
- Explore key <u>Responsible Al</u> guidelines and principles

- Learn more in the <u>Azure OpenAl Workshop</u>
- See examples in the Open Al Cookbook
- Start experimenting and developing your MVP
- Get support and help

In Seattle: May 23-25, 2023 | Online: May 23-24, 2023 PDT

Join us in Seattle for Microsoft Build!

President & De LIVE at Microsoft President & De LIVE at Microsoft Build! Be there to explore how Microsoft and OpenAI's full-stack AI platform empowers developers and get ready to make AI work for you.

Register now

Showing 18 speakers













Q&A

Thank you!

Connect

