An Introduction to OpenAI

2023.08

Jeff Fattic
App Innovation Lead | Microsoft

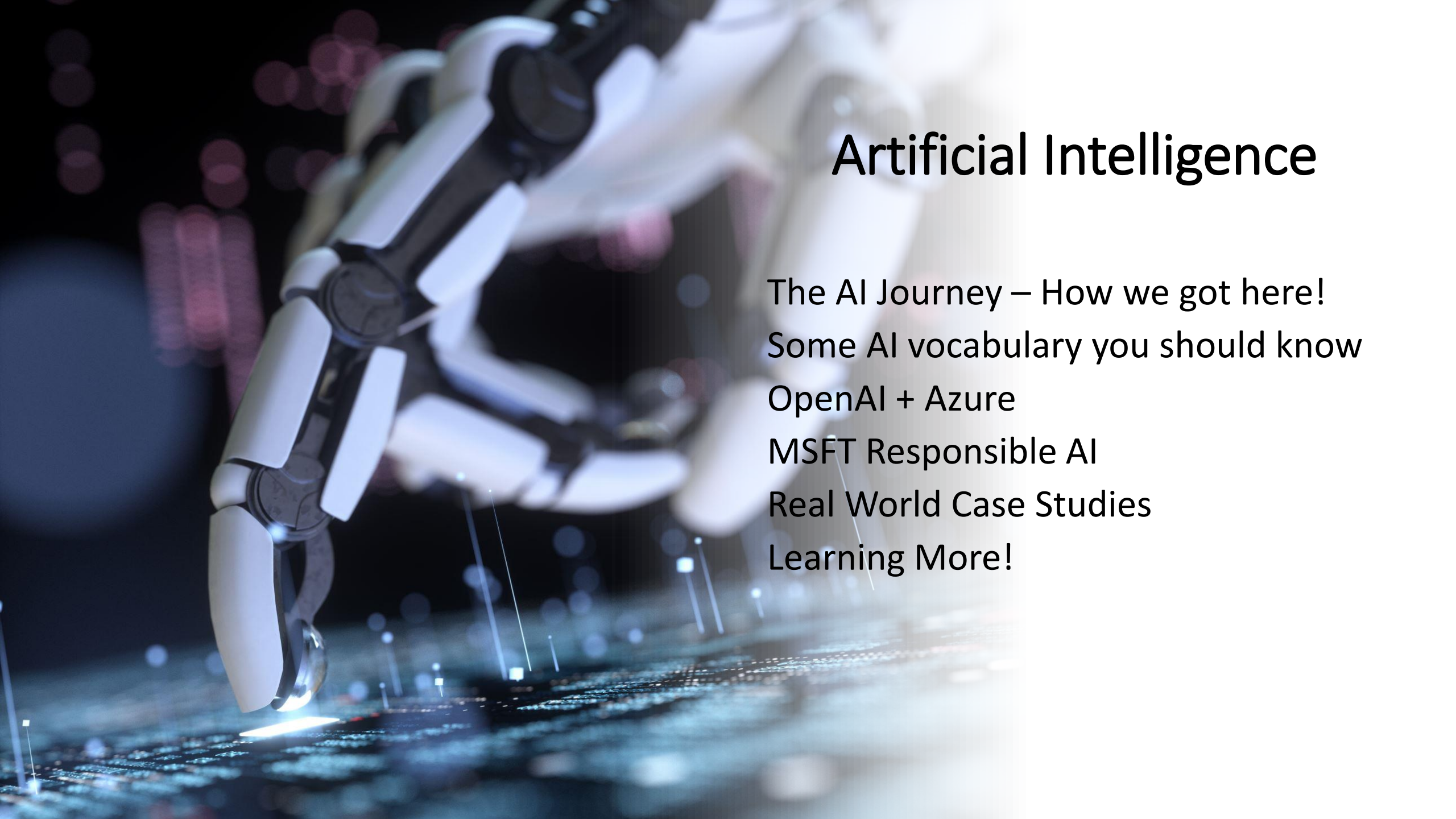
Randy Pagels
DevOps Architect | Xpirit USA

https://github.com/microsoft/globalopenaihack

SemanticKernel

[OPTIONAL]

1. **Clone this repo**
2. **Create Settings.json in Tutorial00**



Artificial Intelligence

The AI Journey – How we got here!

Some AI vocabulary you should know

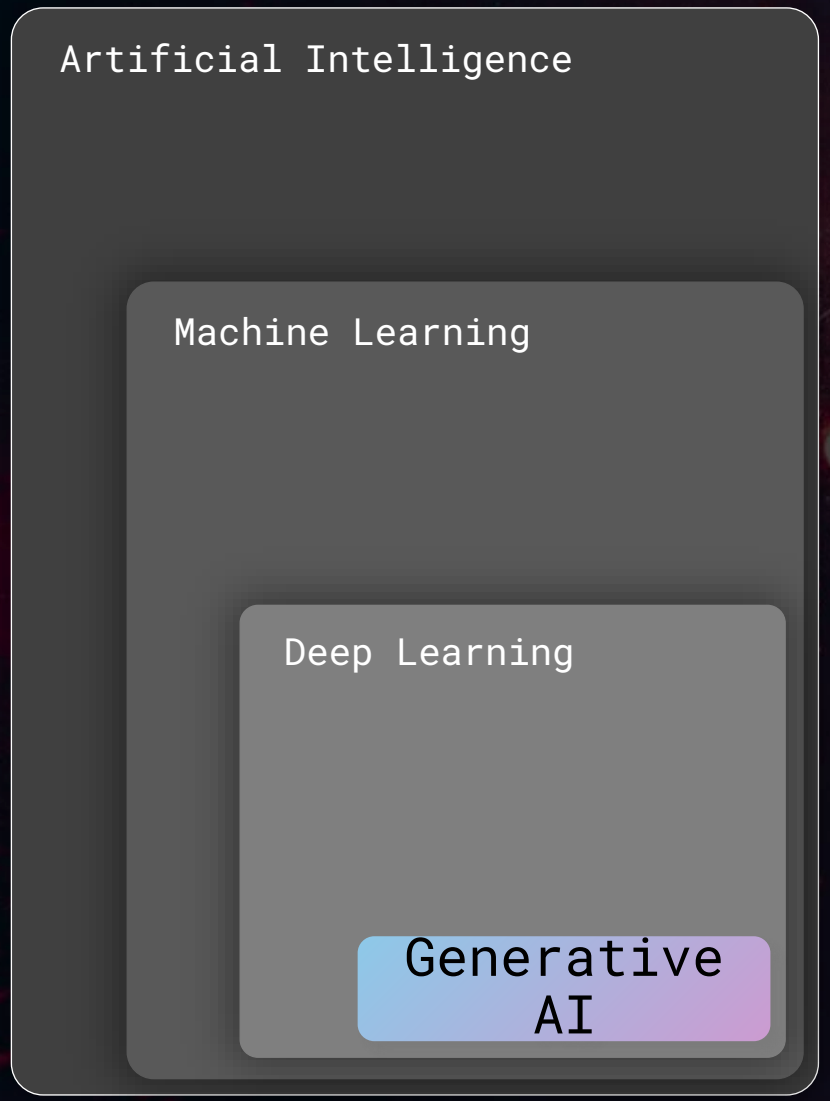
OpenAI + Azure

MSFT Responsible AI

Real World Case Studies

Learning More!

Brief history of AI



1956

Artificial Intelligence

The field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence

1997

Machine Learning

Subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions

2017

Deep Learning

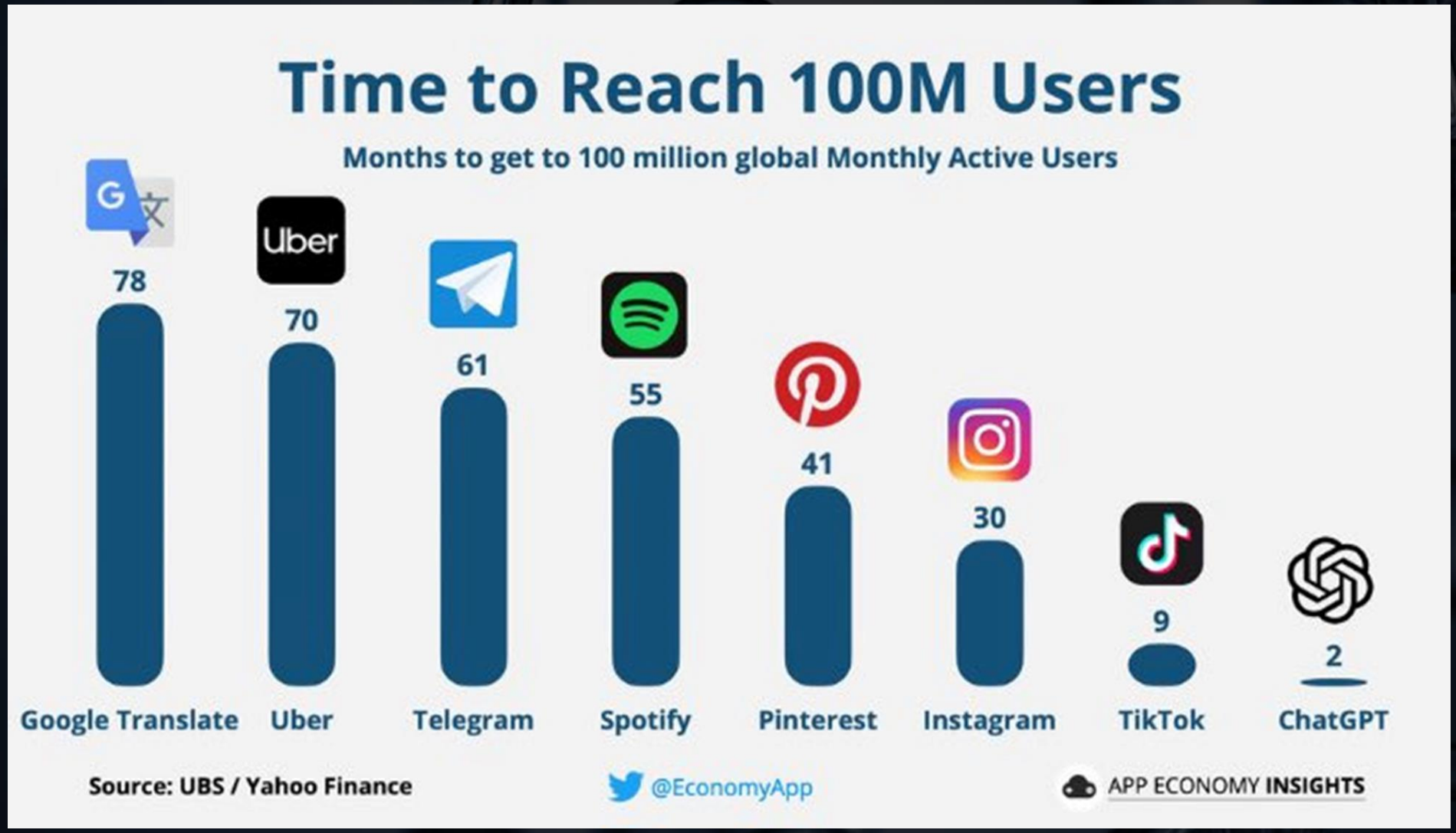
A machine learning technique in which layers of neural networks are used to process data and make decisions

2021

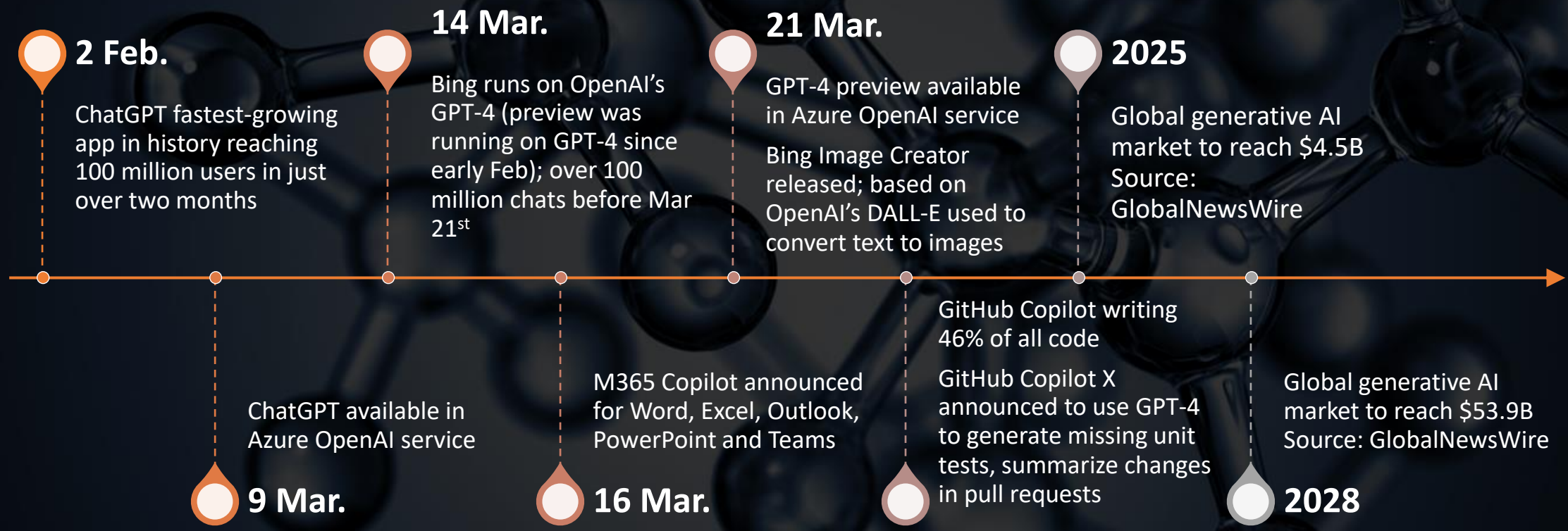
Generative AI

Create new written, visual, and auditory content given prompts or existing data

And then ChatGPT…



And more recently…



Some AI Vocabulary

OpenAI

ChatGPT

Prompts

Prompt Engineering

Completions

Tokens

Fine-Tuning

Embeddings

Copilot

Semantic Kernel

What is Azure OpenAI Service?

- One of several Azure AI services. Azure OpenAI Service allows developers to use large language generative AI models for enterprise-grade applications.
- With built-in responsible AI and enterprise-grade security, the service is designed to detect and mitigate harmful use.

GPT-3 / GPT-4

Generate & understand text

Codex (deprecated)

Generate & understand code

DALL-E 2

Generate images from text prompts

ChatGPT

Personalize conversations

Azure OpenAI Service Capabilities

- Use cases hitting every industry: education, healthcare, entertainment, financial services, retail, manufacturing, and more!

Extracting insights

Code generation

Semantic search

Summarizing text

Language Translation

Writing assistance

Virtual Assistants

Dialog agents

Creative Ideation

Essay outlines

Subject Research

Code documentation

Code refactoring

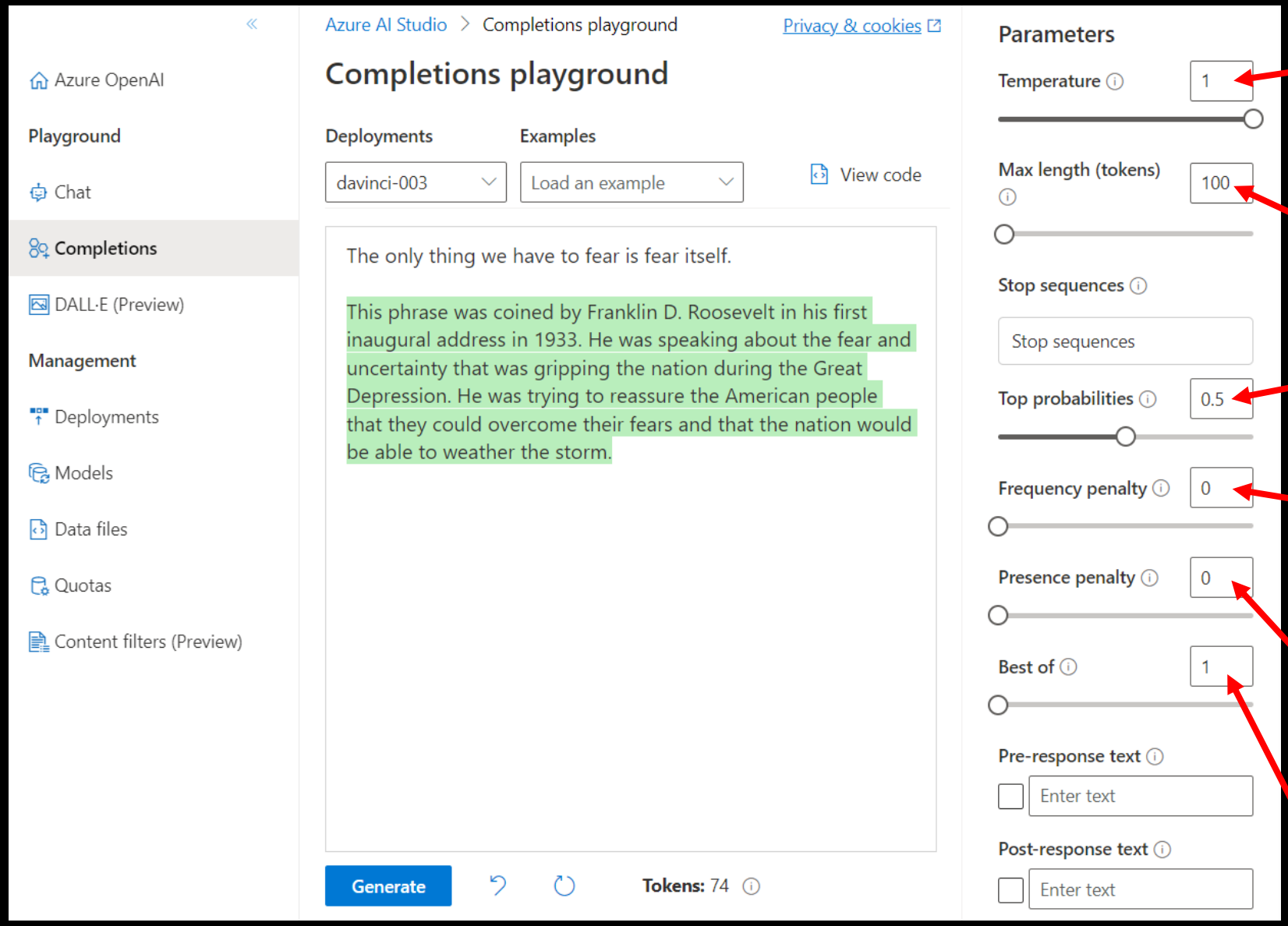
Image Generation

Classifying text

Answering questions

Unit test generation

Anatomy of Azure OpenAI Service (Completions)



Temperature
Controls randomness
Generate more creative completions.

Tokens
Limits number of tokens per model.
A token approximately 4 characters.

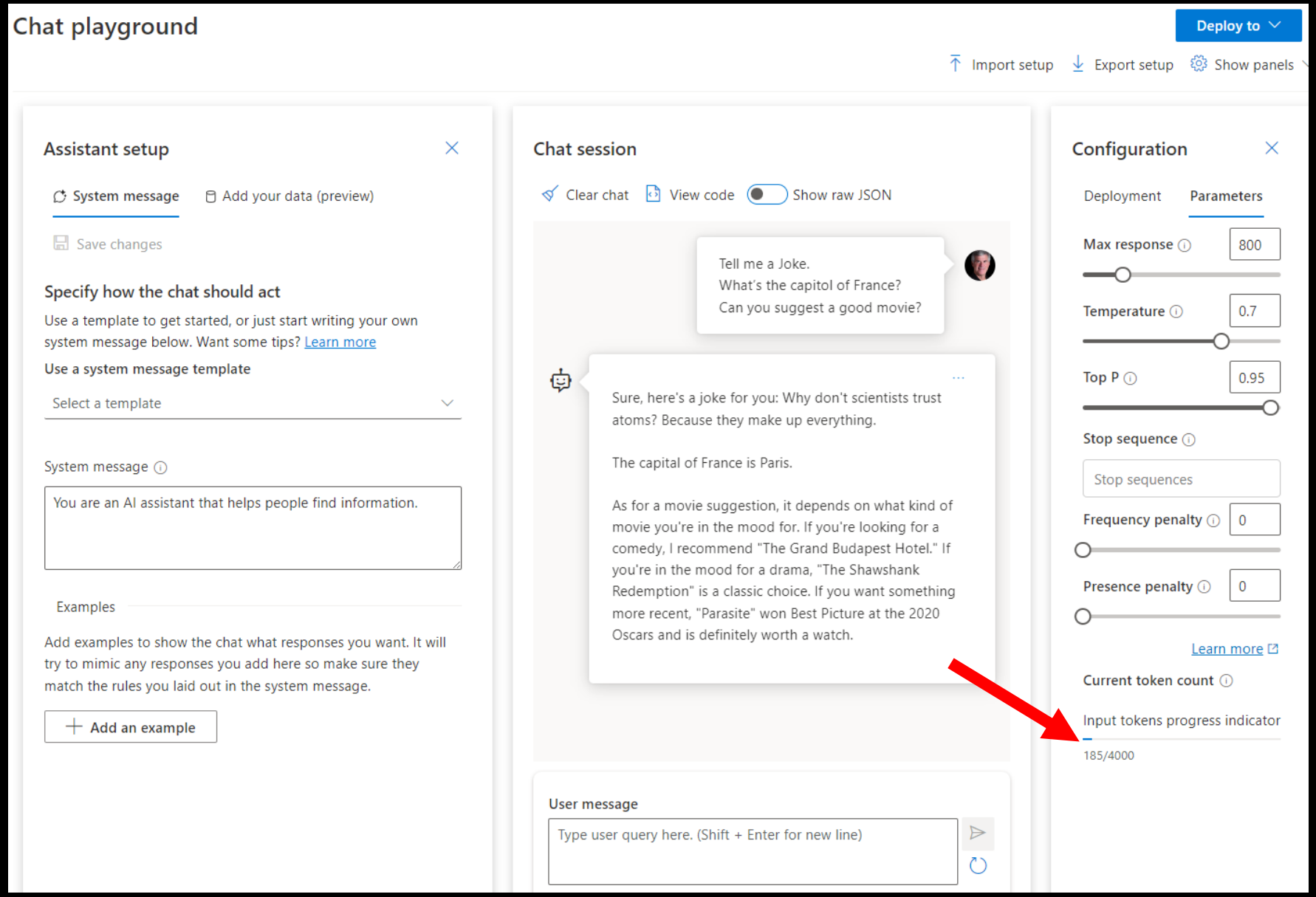
Top probabilities
Controls randomness

Frequency penalty
Decreases likelihood of repeating exact same text in a response.

Presence penalty
Increases likelihood of introducing new topics in a response.

Best of
Generate multiple responses

Anatomy of Azure OpenAI Service (Chat)



Token Count

132/4000

The API supports a maximum of 4000 tokens shared:

- Prompt
- system message
- Examples
- message history
- user query
- model's response

Microsoft Cloud Runs on Trust

- Your data is **your** data
  - Data is stored encrypted in your Azure subscription
- Your data is **not** used to train the foundational AI models
  - Azure OpenAI Service provisioned in your Azure subscription
  - Model fine tuning stays in your Azure subscription and never moves into the foundation AI models
- Your data is **protected** by the most comprehensive enterprise compliance and security controls
  - Encrypted with Customer Managed Keys
  - Private virtual networks and role-based access control
  - Soc2, ISO, HIPPA, CSA STAR compliant

Microsoft Responsible AI

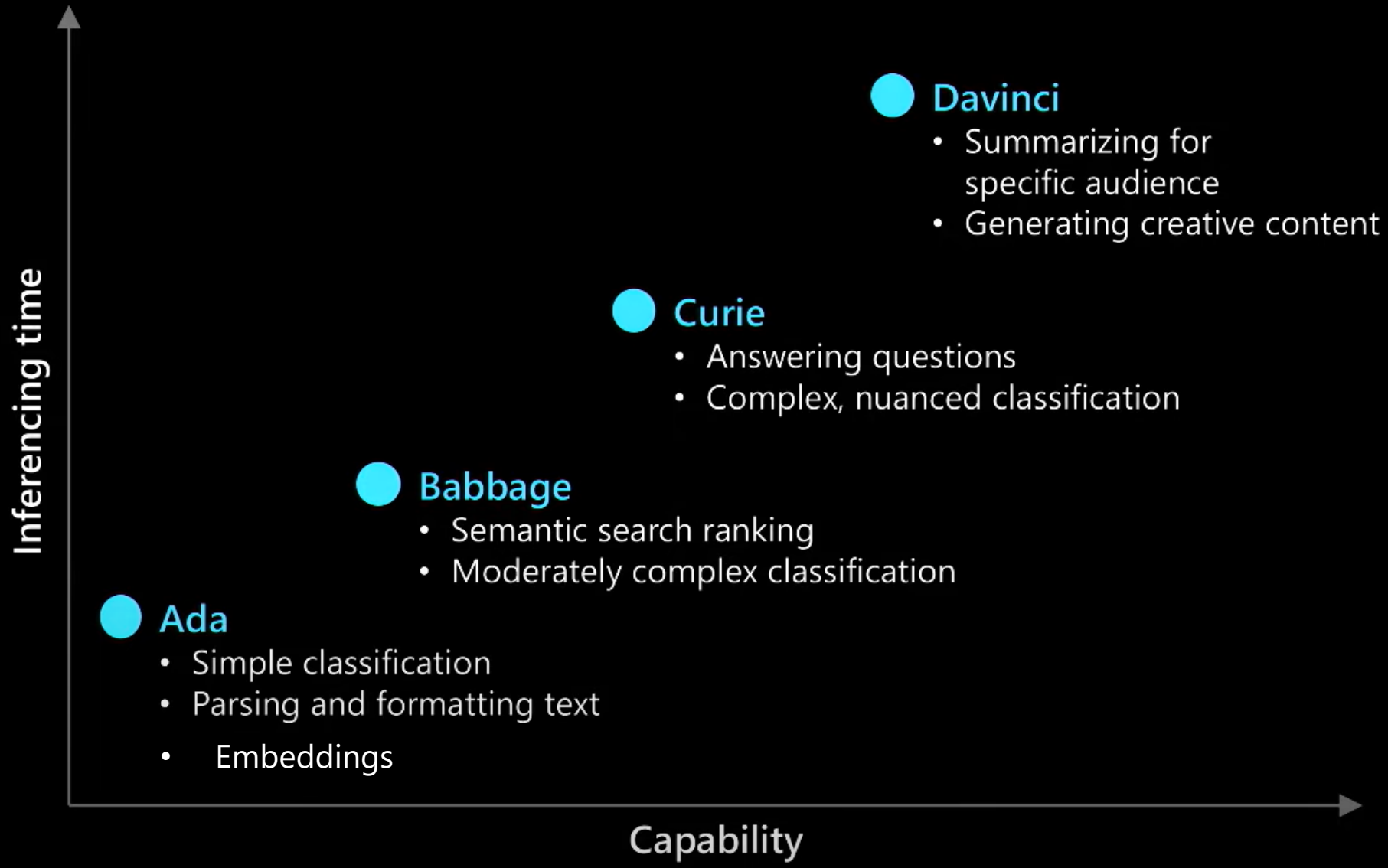
[Responsible AI principles from Microsoft]

Fairness	AI systems should **treat people fairly**
Reliability & Safety	AI systems should **perform reliably** and safely
Privacy & Security	AI systems should be secure and **respect privacy**
Inclusiveness	AI systems should **empower everyone** and engage people
Transparency	AI systems should **be understandable**
Accountability	**People** should **be accountable** for AI systems

Comparison of GPT versions

- **GPT-3.5** (4,000 Tokens)
  - Use-case specific models to optimize inference time and performance
  - Suitable for a large range of use cases
- **ChatGPT**
  - Should be first choice for most use cases
  - Most economical GPT model in Azure OpenAI Service
  - For all workloads, not just chat
- **GPT-4** (32,000 Tokens)
  - Improved problem solving and reasoning capabilities by 10x
  - Iterative refinement:
    - Paste in code errors & GPT-4 will fix for you
    - Iterate on stories
  - Increased token limit – works well for long content

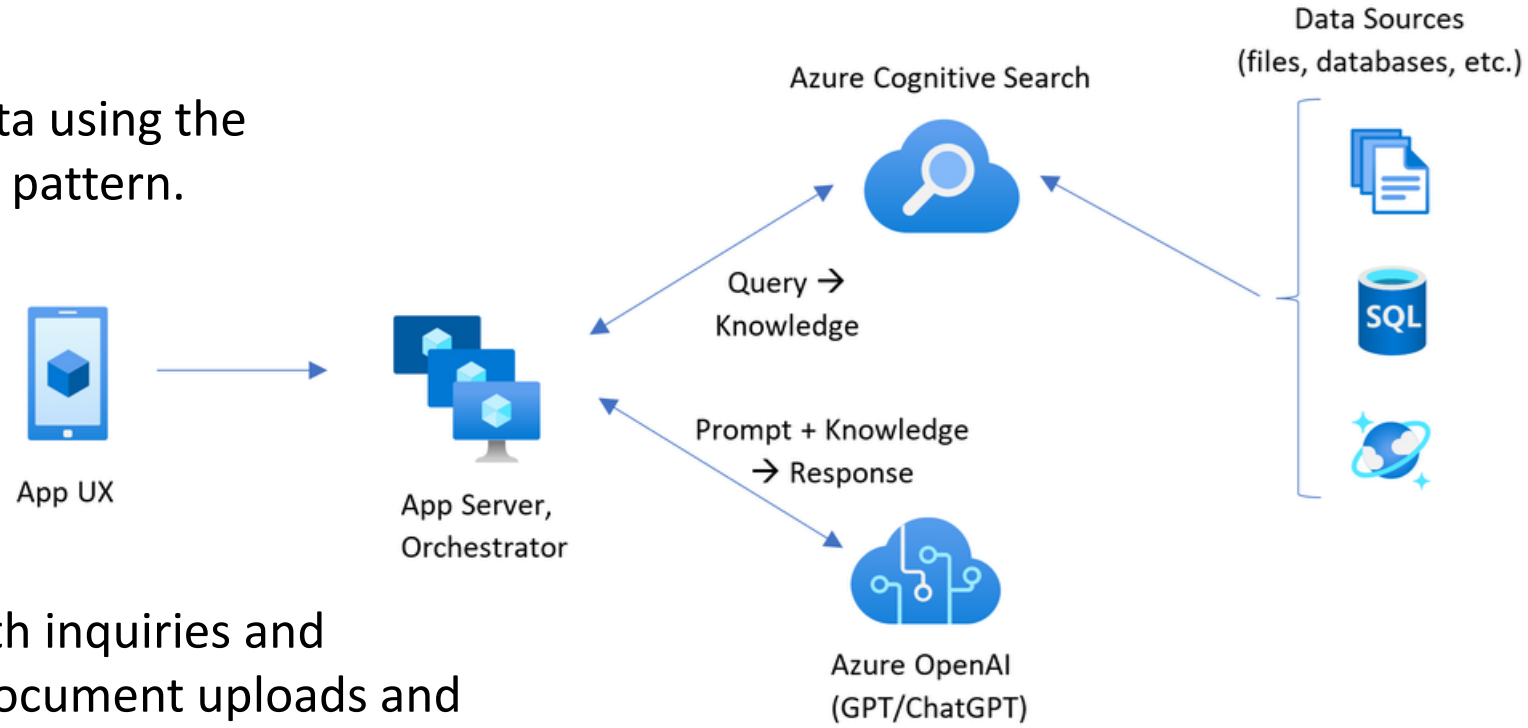
GPT-3 models


Inferencing time
Capability
Davinci
• Summarizing for specific audience
• Generating creative content
Curie
• Answering questions
• Complex, nuanced classification
Babbage
• Semantic search ranking
• Moderately complex classification
Ada
• Simple classification
• Parsing and formatting text
• Embeddings

Example: Retrieval Augmented Generation pattern

Uses Azure OpenAI Service to access the ChatGPT model (gpt-35-turbo), and Azure Cognitive Search for data indexing and retrieval.

ChatGPT-like experiences over data using the **Retrieval Augmented Generation** pattern.



Example
Applications to empower staff with inquiries and patients with interactive health document uploads and simplification.

https://github.com/Azure-Samples/azure-search-openai-demo

Learning: Azure OpenAI Service

Start Here!

- Create an Azure subscription
- ★ Apply for access to the Azure OpenAI Service: https://aka.ms/oai/access)
- Watch the video: An Introduction to Azure OpenAI
- Get Familiarized with the product content: Azure OpenAI Page
- Review how other customers are using Azure AI

Basic

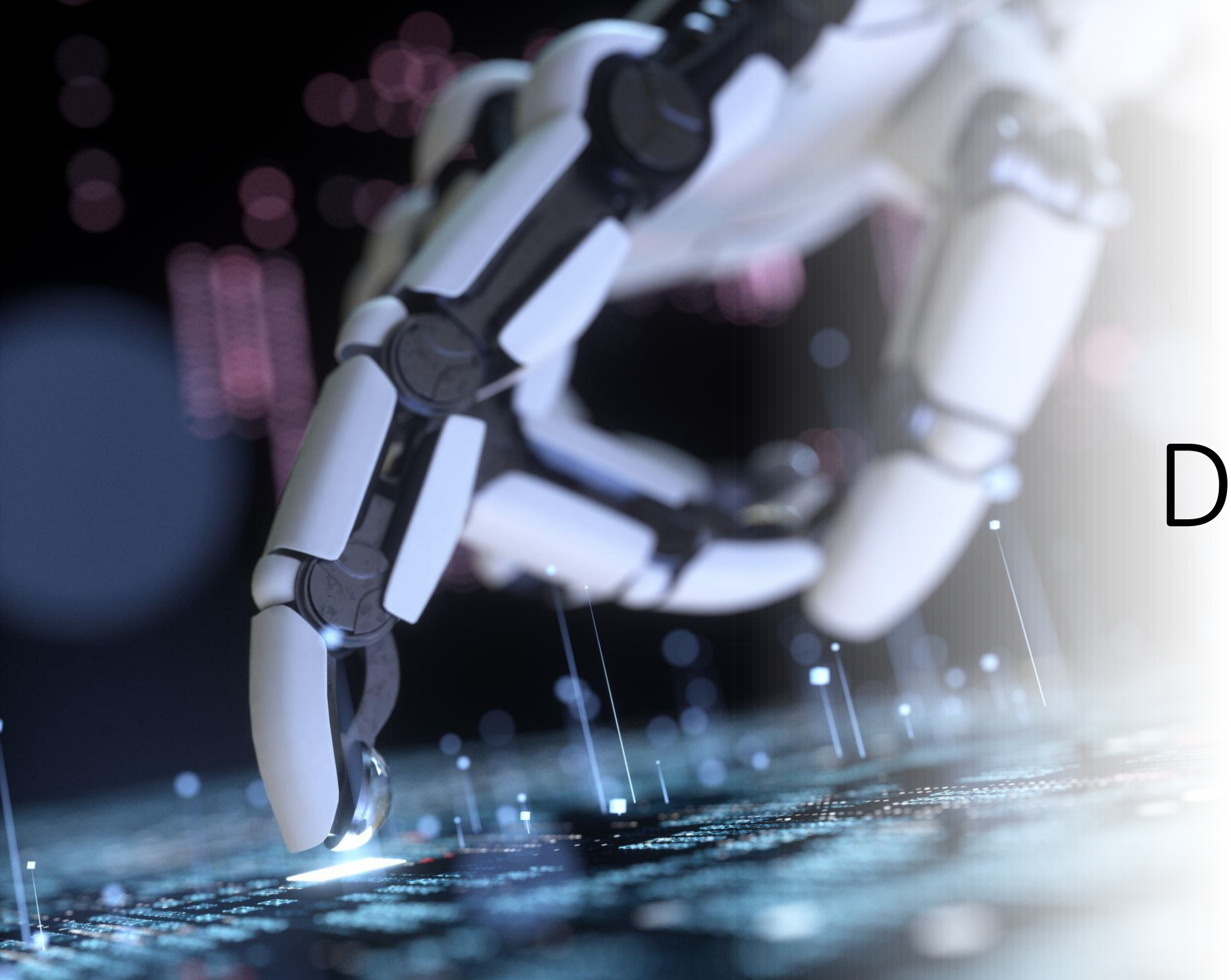
- Understand "What is Azure OpenAI?": compare Azure OpenAI and OpenAI. Review key concepts.
- ★ Work through Introduction to Azure OpenAI Service training module
- Explore key Responsible AI guidelines and principles

Intermediate

- 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

Advanced

- ★



Demos

https://github.com/microsoft/globalopenaihack

SemanticKernel