Microsoft

Nash Tech.

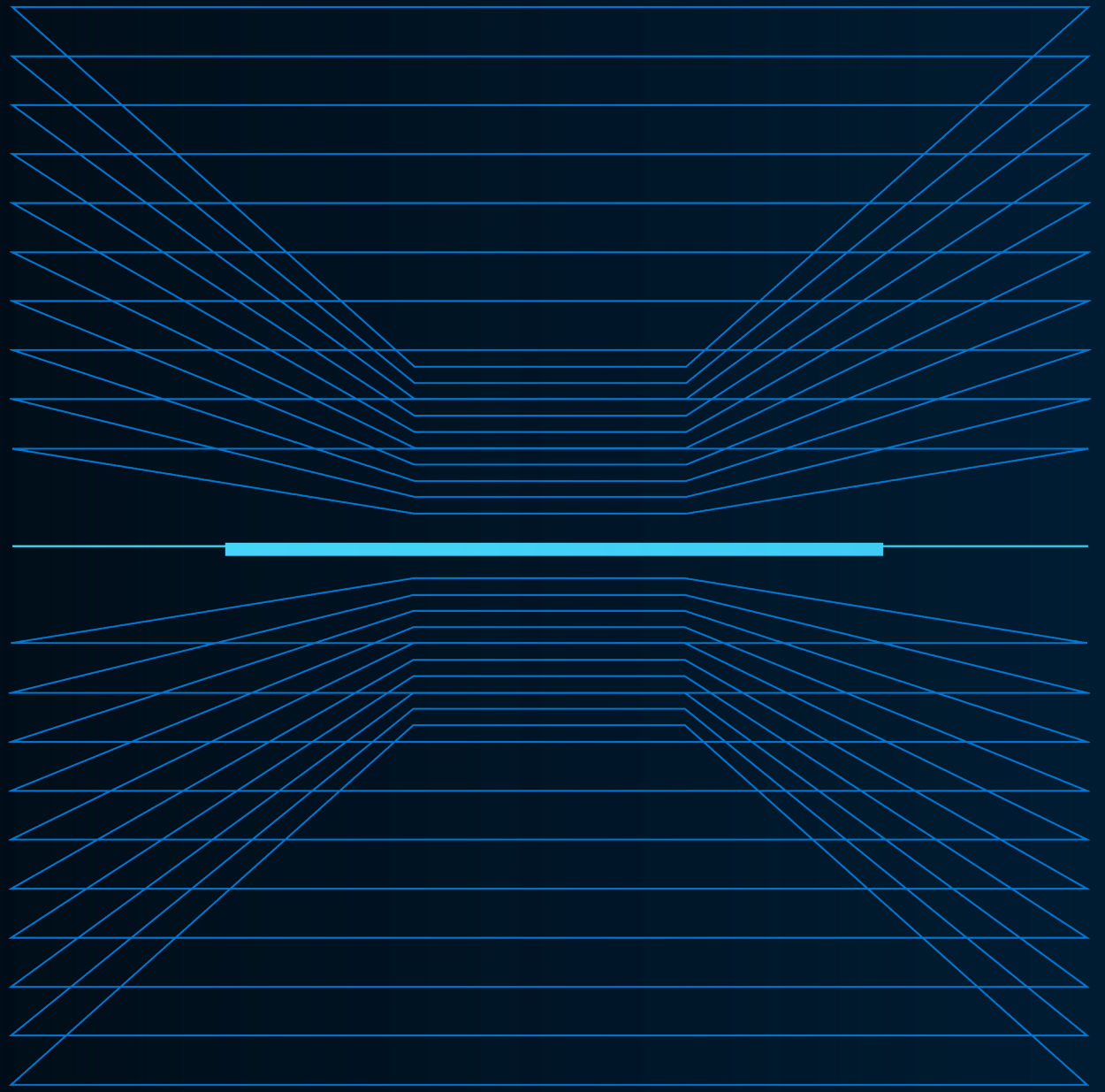
Azure Open AI introduction

(interesting demo included)

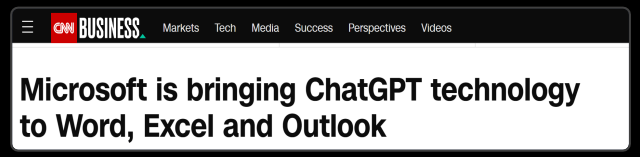
Phi Huynh

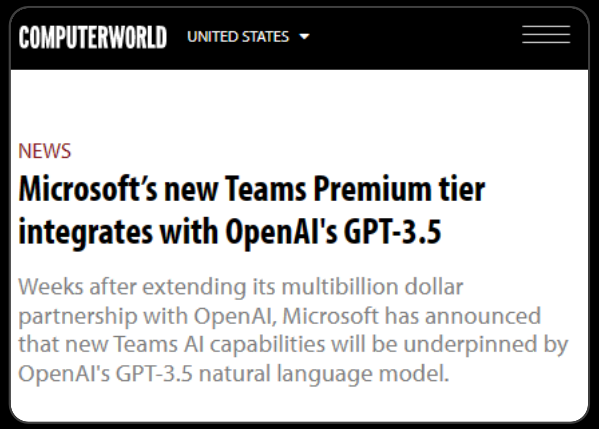
Microsoft MVP

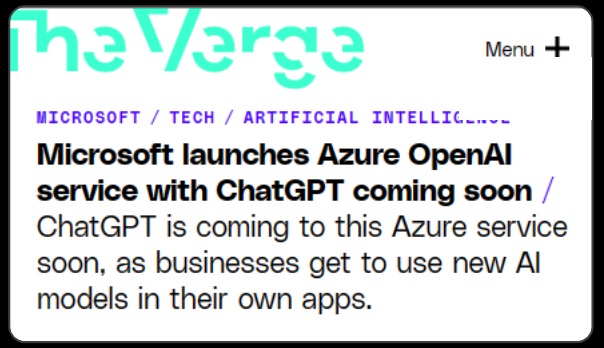
May 2023

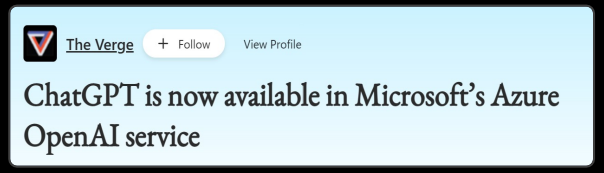


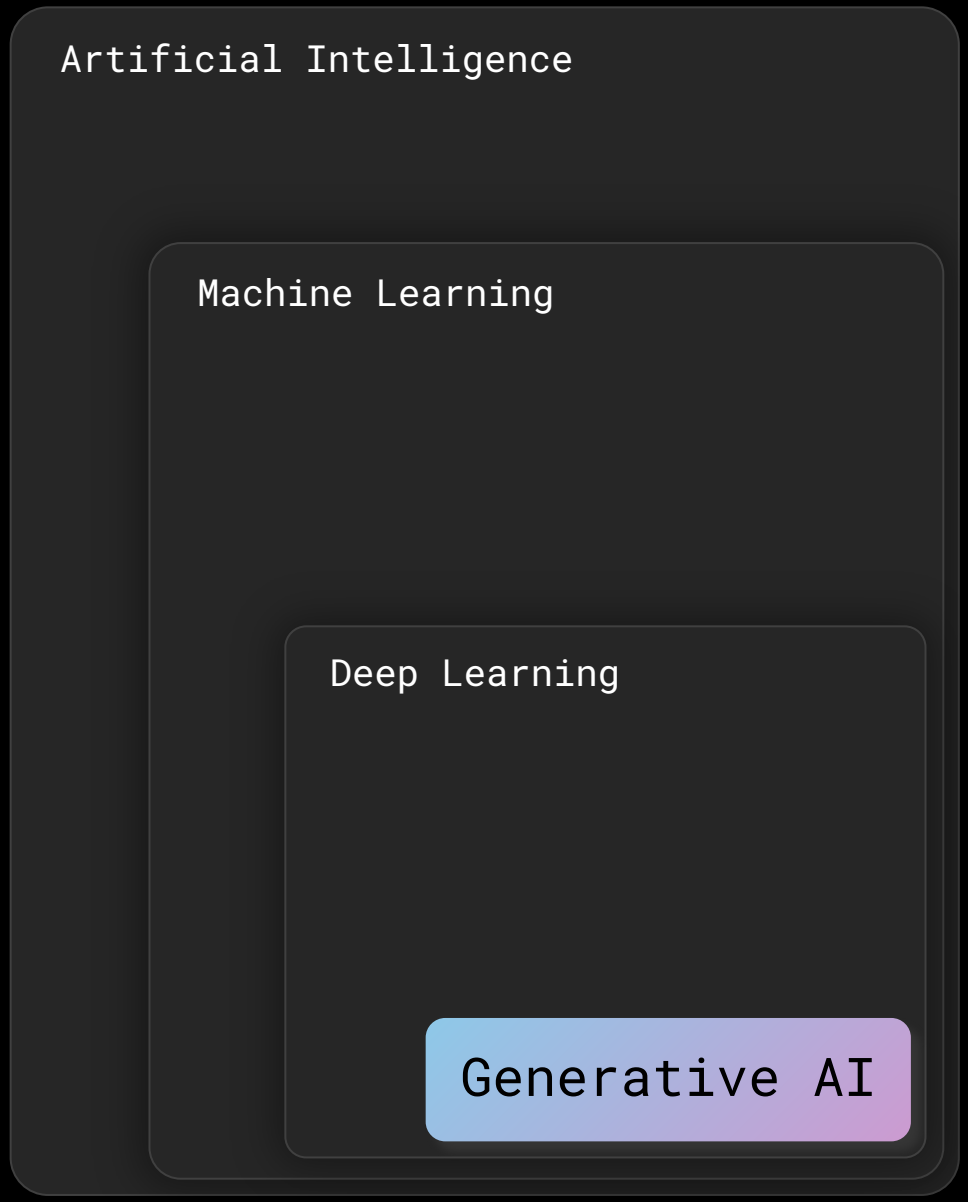
The AI Technology is Here











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

Ensure that artificial general intelligence (AGI) benefits humanity

Empower every person and organization on the planet to achieve more

---

GPT-3.5	ChatGPT and GPT-4	Codex	DALL·E 2
Text	Conversation	Code	Images

GPT-3.5

GPT-4 (preview)

ChatGPT (preview)

Codex

DALL·E 2 (preview)

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

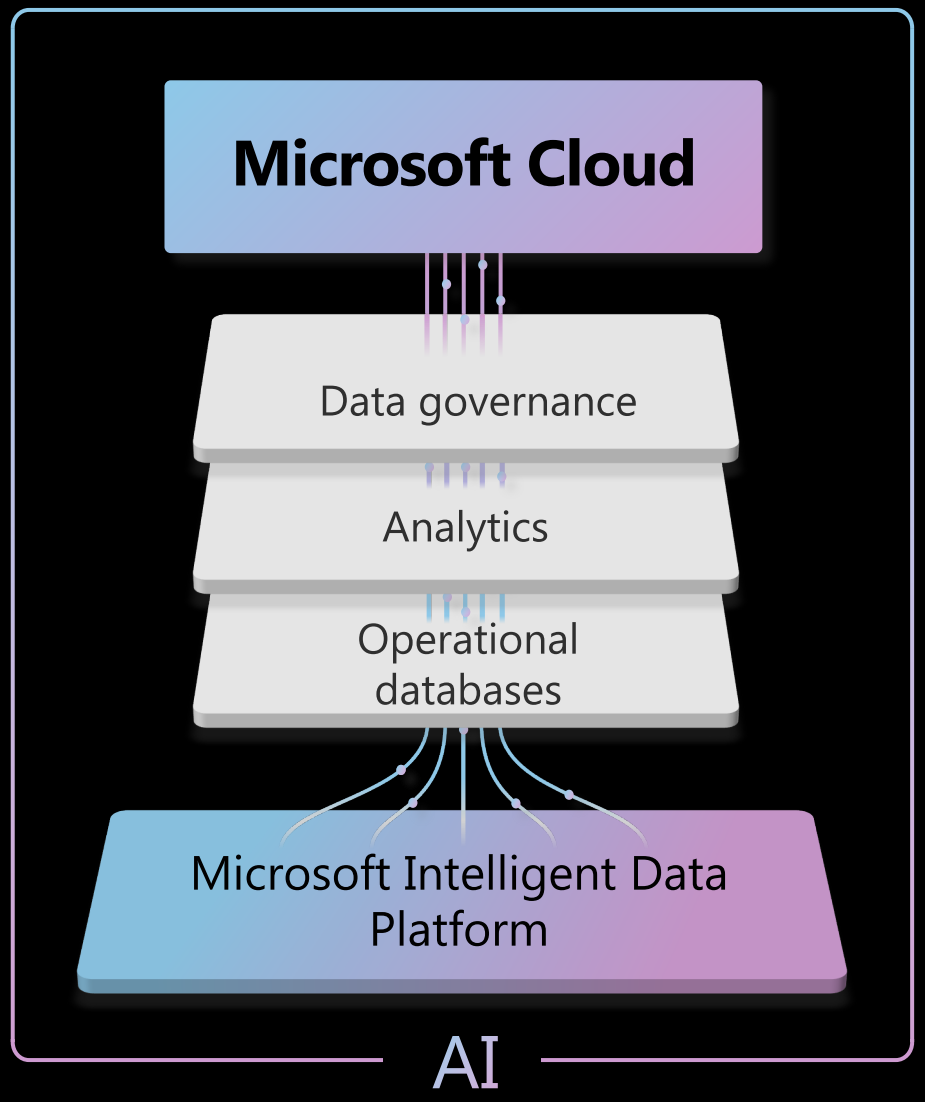
```
SELECT *
FROM customers
WHERE State = 'TX' AND
FirstName = 'Jane'
```

Prompt

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

Response

Azure Open AI security & privacy

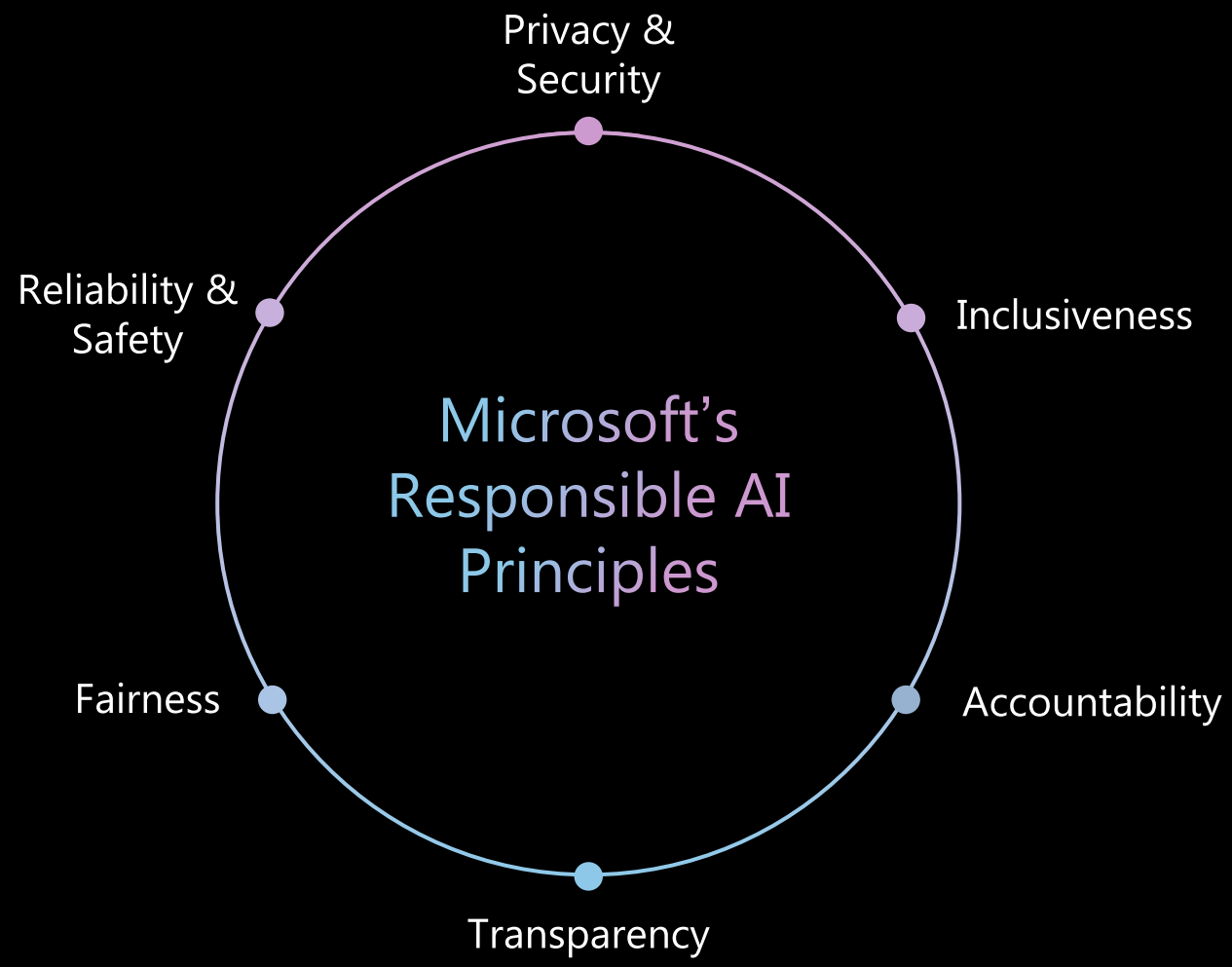


Deployed within your Azure subscription, secured by you, accessed only by you, and tied to your datasets and applications

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

Secure networking through private endpoints and VNETs

https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy

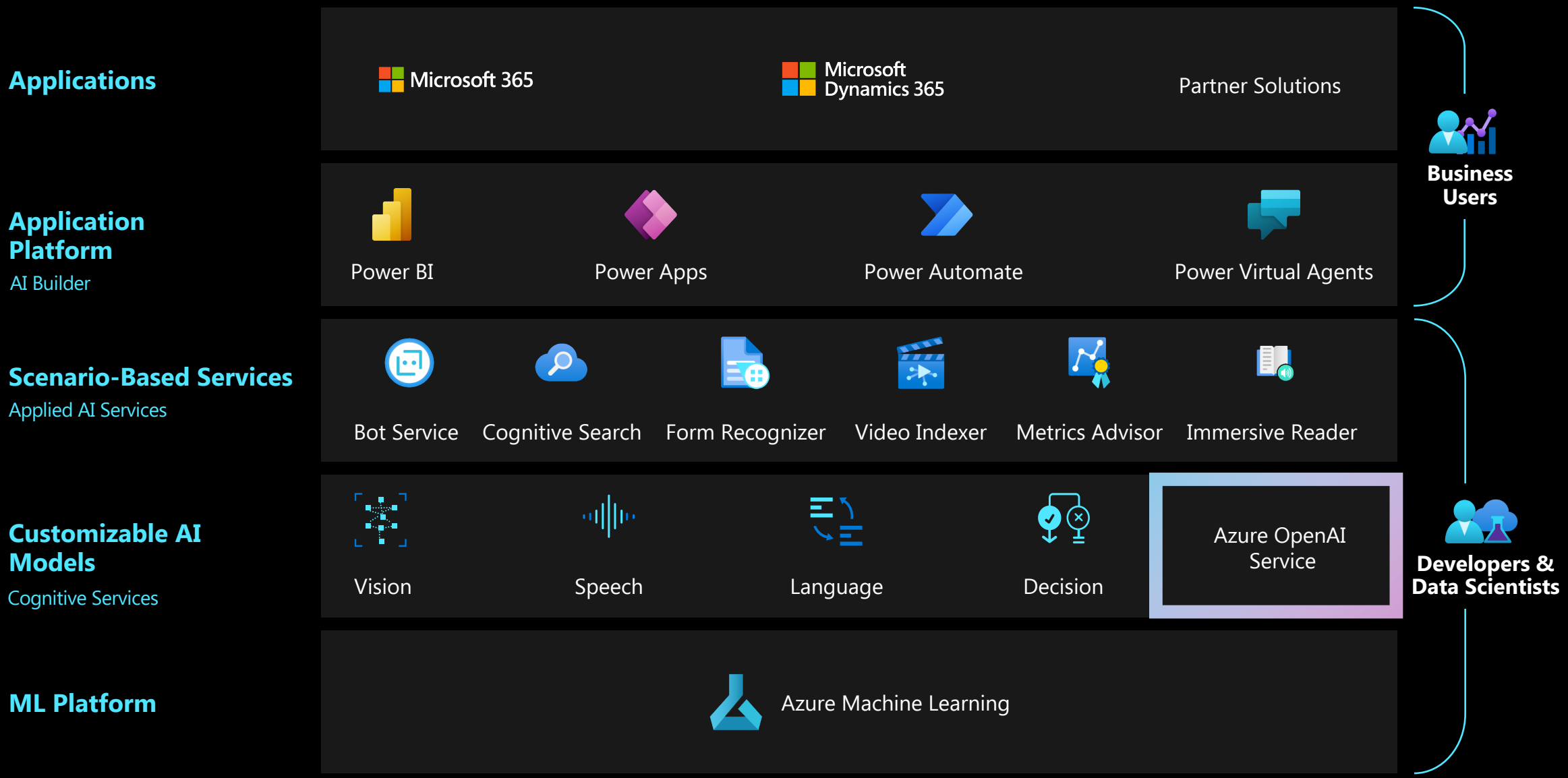


Building blocks to enact principles

- Tools and processes
- Training and practices
- Rules
- Governance

https://azureaidevs.github.io/hub/2023-aia/day6

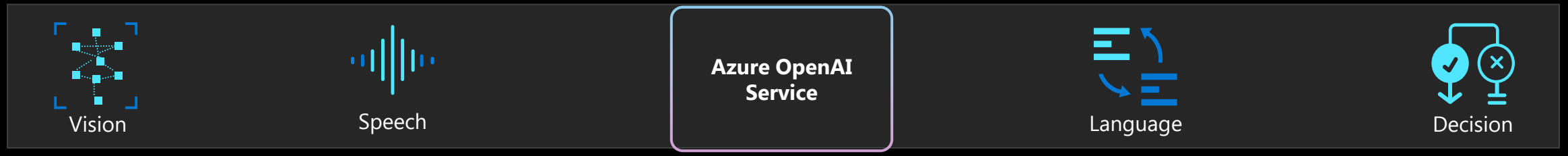
Azure AI


Applications
Microsoft 365
Microsoft Dynamics 365
Partner Solutions
Business Users
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
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

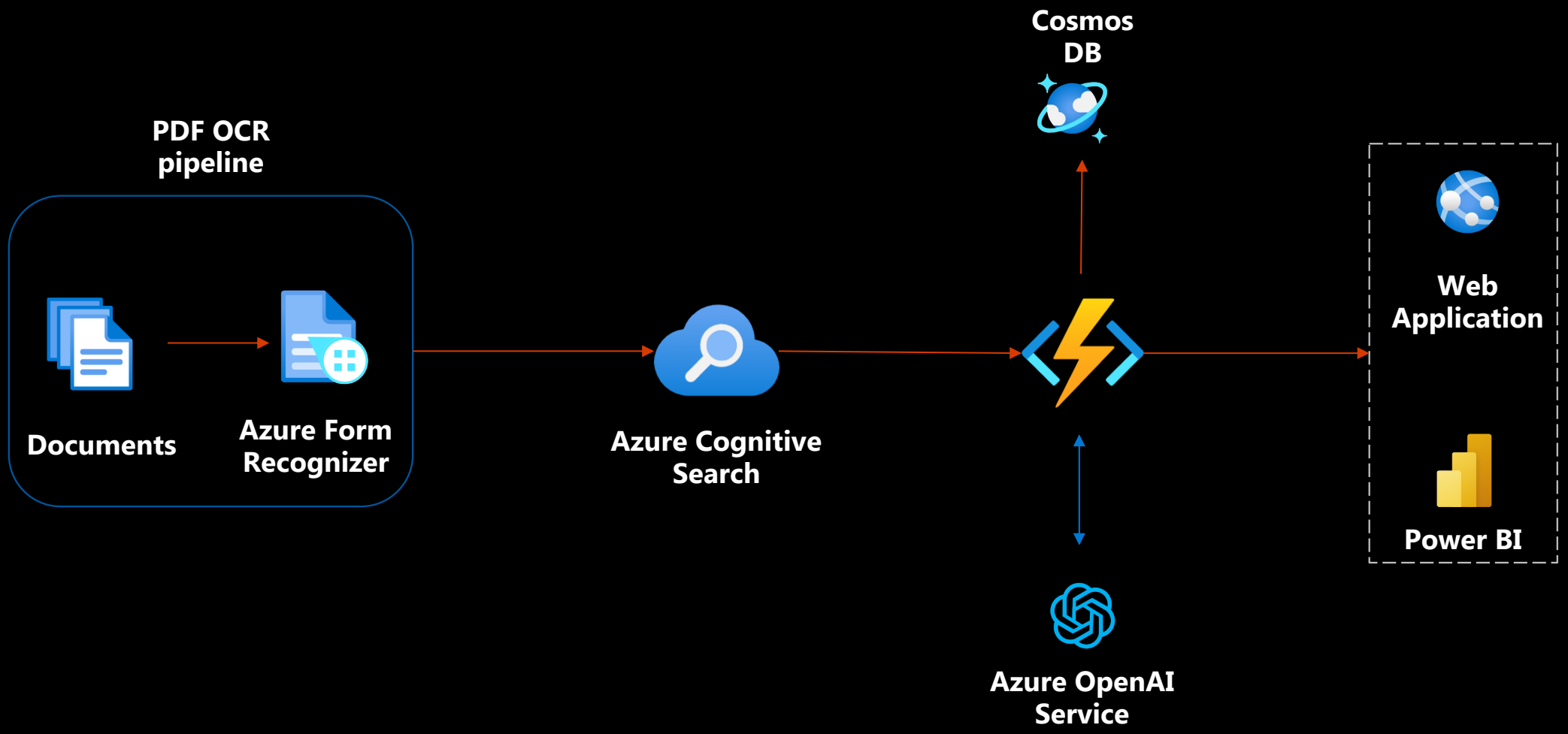


Azure AI Cognitive Services

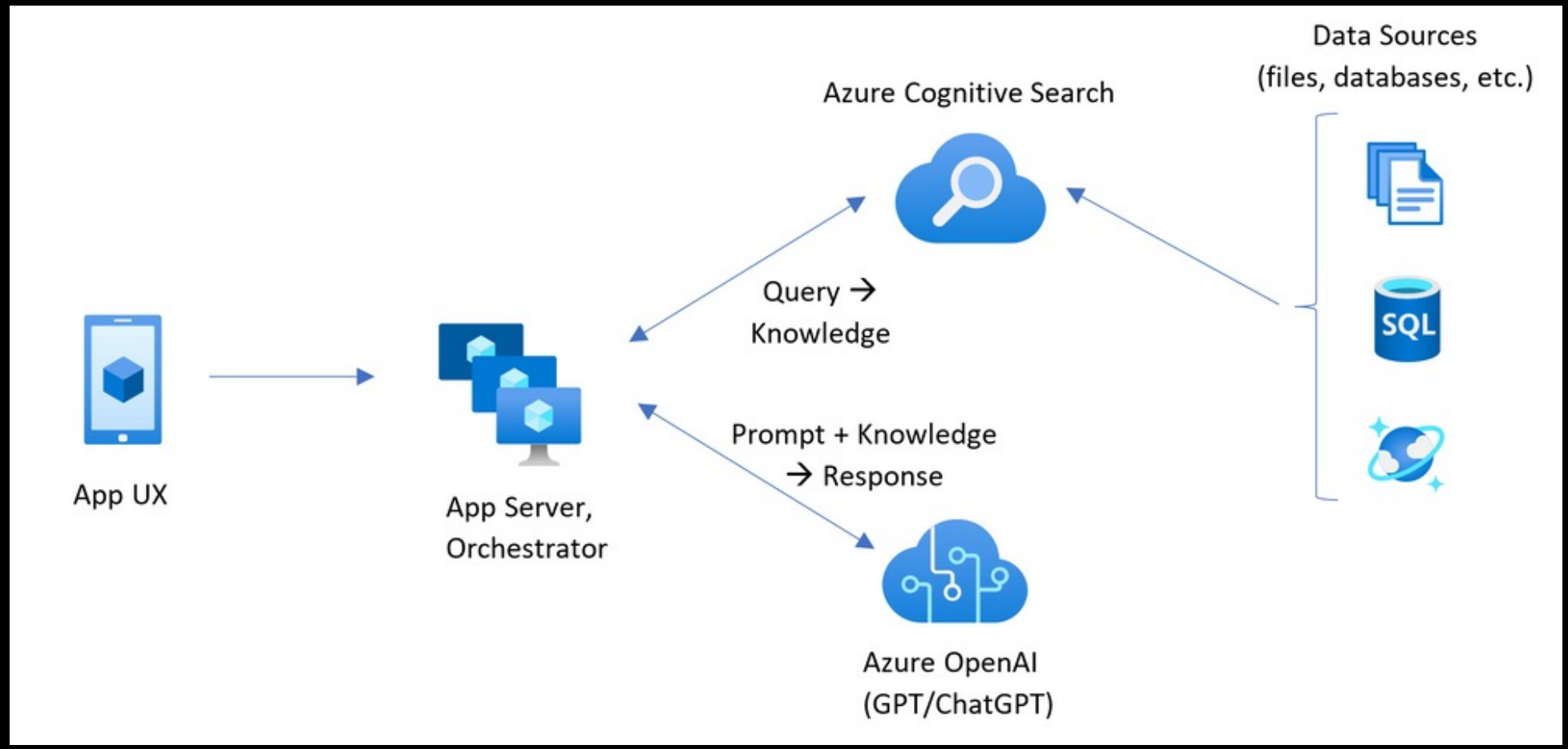
Azure Open AI use cases

Document Process Automation

Extract rich insights from documents and summarize
https://azureaidevs.github.io/hub/2023-aia/day10



Enterprise chat bot model with Azure Open AI



https://techcommunity.microsoft.com/t5/ai-applied-ai-blog/revolutionize-your-enterprise-data-with-chatgpt-next-gen-apps-w/ba-p/3762087

Toolset & frameworks to build apps on LLM

Frameworks for LLM

Langchain

Semantic Kernel

Ref:

https://github.com/microsoft/semantic-kernel

https://python.langchain.com/en/latest/index.html

https://github.com/facebookresearch/faiss

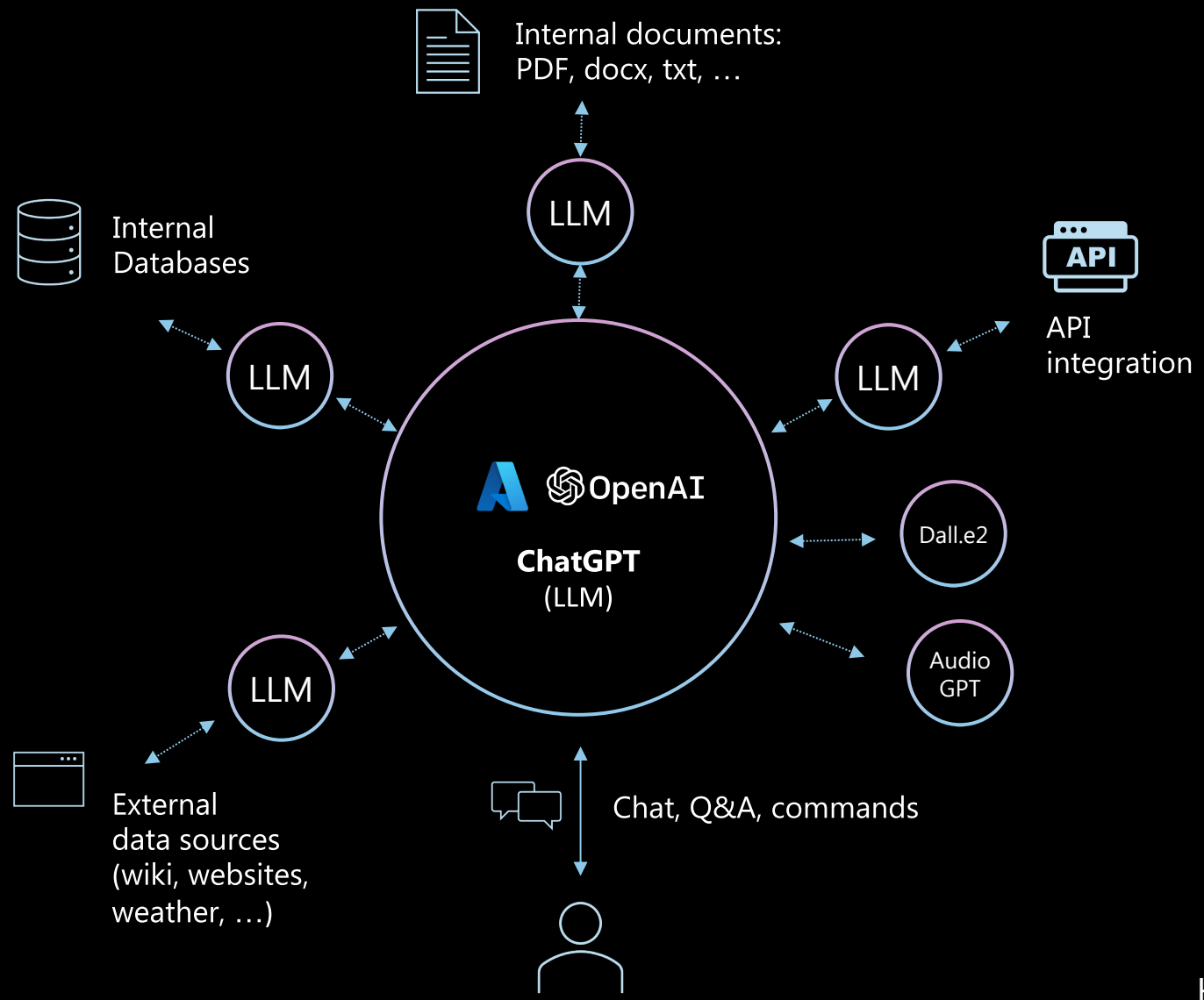
Vector databases

Chroma

FAISS

Pinecone

New app model with LLM

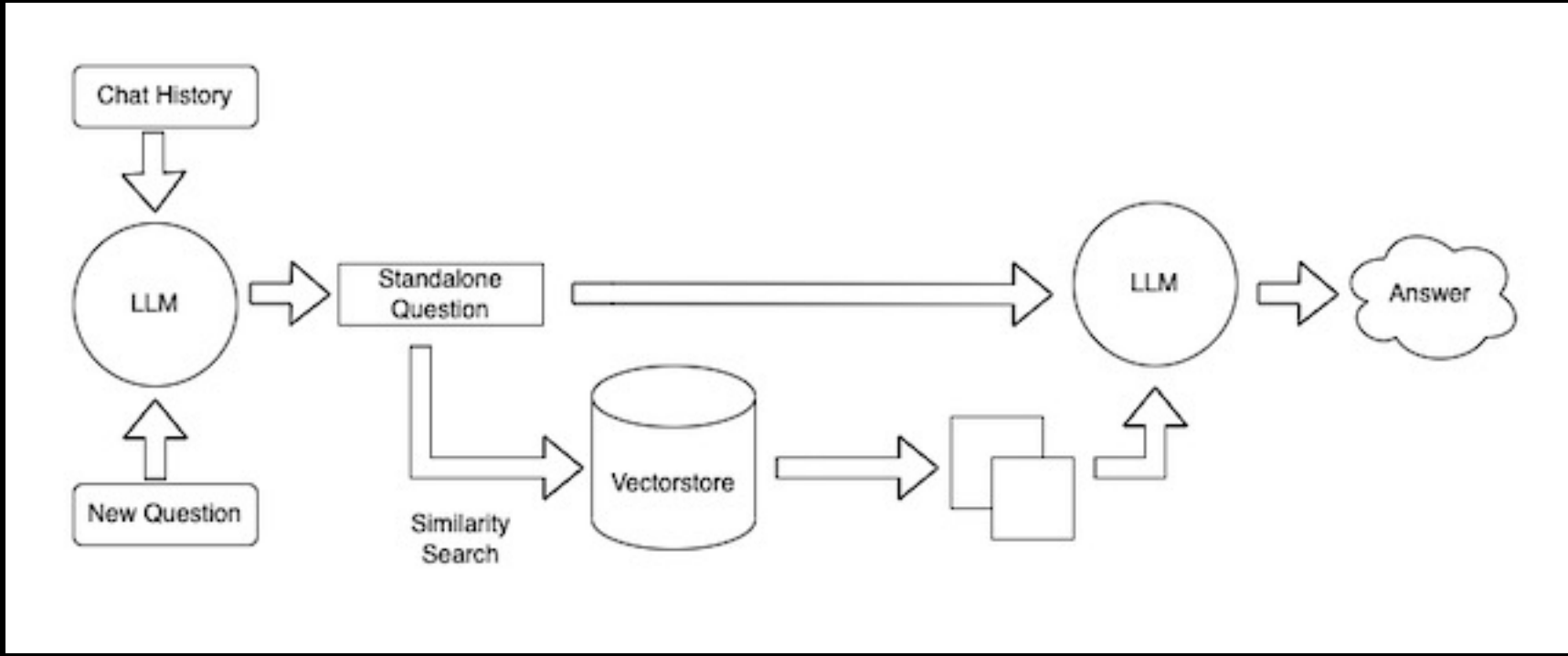


What can we do with this app model?

- Customer support
- Chatbot
- Automation process
- CoPilot

How it works: https://arxiv.org/pdf/2205.00445.pdf

How a vector database work in LLM app? (RAG technique)



https://betterprogramming.pub/build-a-chatbot-on-your-csv-data-with-langchain-and-openai-ed121f85f0cd

Prompt engineering

https://www.promptingguide.ai/

https://learn.microsoft.com/en-us/azure/cognitive-services/openai/concepts/prompt-engineering

https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api

Pricing

Price per tokens

Models	Per 1,000 tokens Standard
Text-Ada	$0.0004
Text-Babbage	$0.0005
Text-Curie	$0.002
Text-Davinci	$0.02
Code-Cushman	$0.024
Code-Davinci	$0.10
ChatGPT (gpt-3.5-turbo)	$0.002

GPT-4	Prompt (Per 1,000 tokens)	Completion (Per 1,000 tokens)
8K context	$0.03	$0.06
32K context	$0.06	$0.12

Fine tuned models (training & hosting) pricing

Models	Training per compute hour Standard
Text-Ada	$20
Text-Babbage	$22
Text-Curie	$24
Text-Davinci	$84
Code-Cushman	$26

Models	Hosting per hour Standard
Text-Ada	$0.05
Text-Babbage	$0.08
Text-Curie	$0.24
Text-Davinci	$3
Code-Cushman	$0.54

See more here: https://azure.microsoft.com/en-us/pricing/details/cognitive-services/openai-service/

Thank you!