



# Cloud & Edge Computing

## Workshop Series Day 1

Embrace the **future of IT** today

Tom Claes – CSA Azure Core



# Agenda – Day 1

Introduction to Cloud Computing

Introduction to Azure Cloud Platform

Computing capabilities

Product portfolio

Introduction to Azure OpenAI

Break 10'

Introduction to Edge computing

Introduction to Azure IoT & Edge AI

Recap - Introduction Day 2

Quiz with (limited edition) prizes for top 3 😊

Questions & Wrap-Up 15'

# Agenda – Day 2

Introduction to Containerization & k8s

Refresh on Generative AI & LLMs

Break 10'

Hands-on project (Docker, Azure & Web App)

Wrap up 20'

Q&A



# What is the cloud?

An introduction to cloud computing  
with lots of cool stuff

Presenter: Gary Baker

@GARYBAKER



# Introduction to Cloud Computing

*"Simply put, cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the internet to offer faster innovation, flexible resources, and economies of scale."*

- Capabilities are hosted at a remote data center managed by a CSP
  - Compute
  - Storage
  - Networking
  - Development tools
- Well known CSPs (Cloud Solution Providers)
  - Azure
  - GCP
  - AWS



# Advantages of the Cloud

*"Cloud computing is a big shift from the traditional way businesses think about IT resources."*

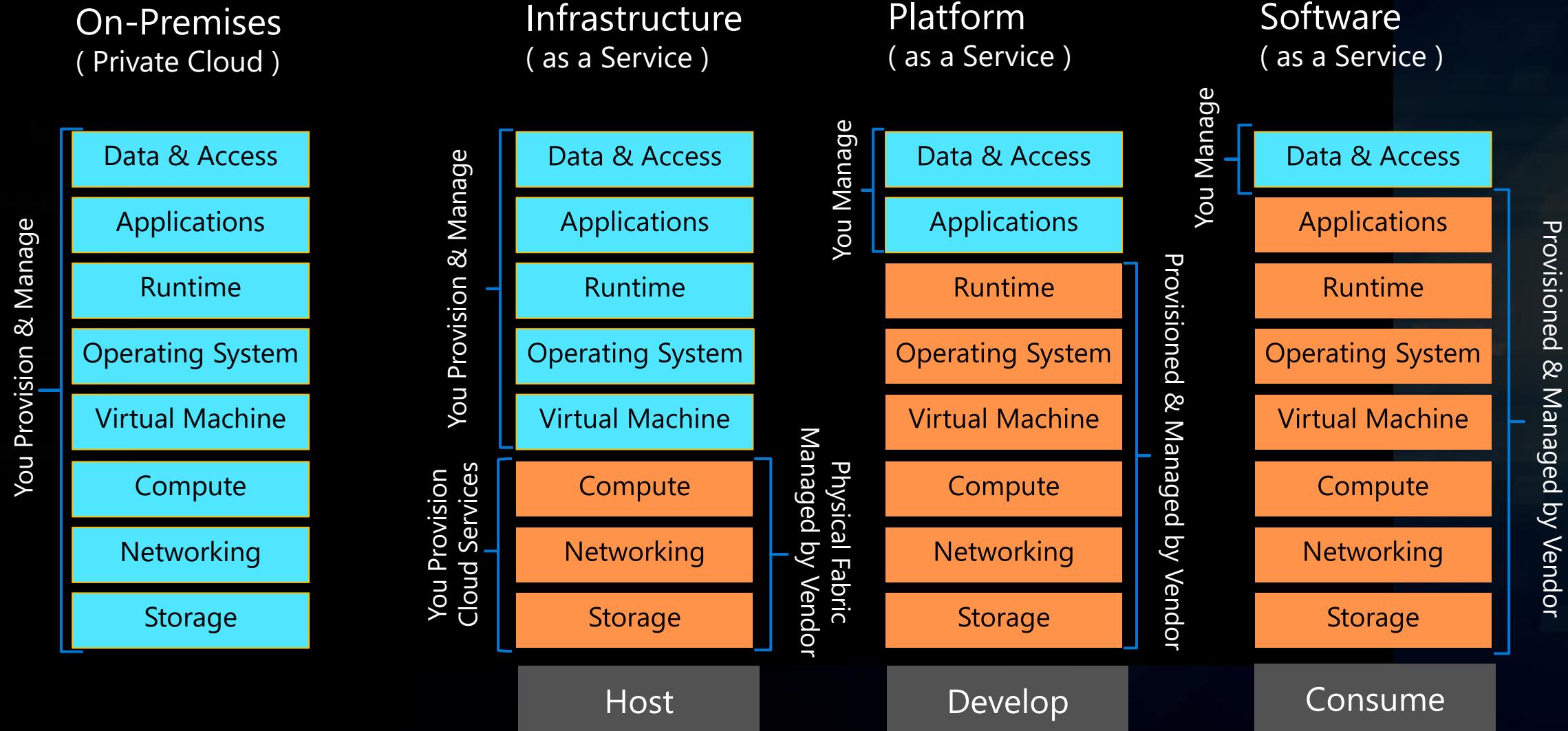
- Optimization of costs
  - CapEx to OpEx
- Speed of services
  - Provisioning of compute in minutes
- Ability to scale elastically
- Focus on core business, less on time-consuming IT management
- Performance of services
- Reliability
- Security



Questions?



# Azure Cloud Computing Capabilities



# Responsibility Zones

Responsibility	SaaS	PaaS	IaaS	On-prem
Data governance & rights management	Customer	Customer	Customer	Customer
Client endpoints	Customer	Customer	Customer	Customer
Account & access management	Customer	Customer	Customer	Customer
Identity & directory infrastructure	Microsoft	Microsoft	Customer	Customer
Application	Microsoft	Microsoft	Customer	Customer
Network controls	Microsoft	Microsoft	Customer	Customer
Operating system	Microsoft	Microsoft	Customer	Customer
Physical hosts	Microsoft	Microsoft	Microsoft	Customer
Physical network	Microsoft	Microsoft	Microsoft	Customer
Physical datacenter	Microsoft	Microsoft	Microsoft	Customer

Always retained by customer

Varies by Service Type

Transfers to Cloud Provider

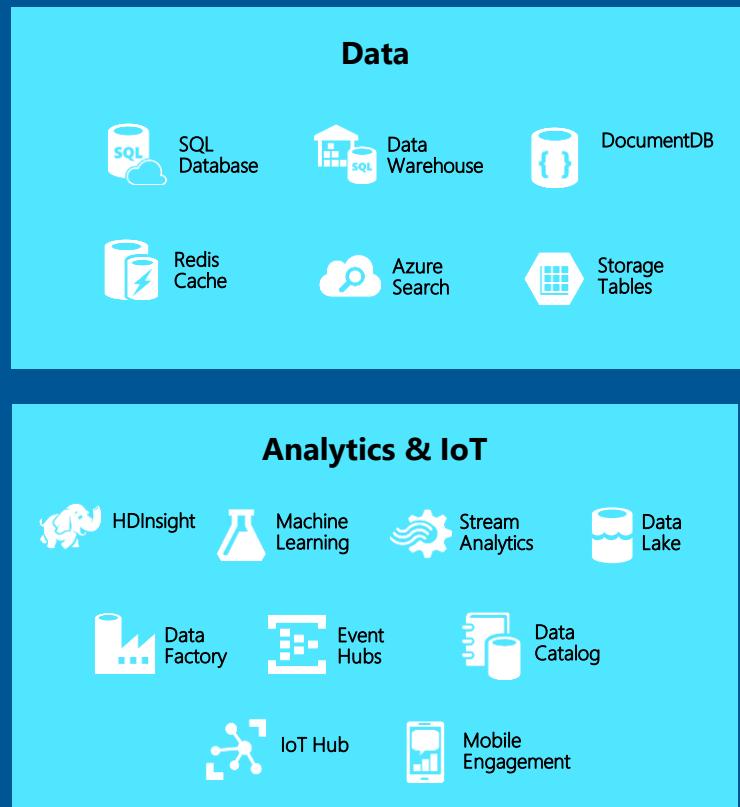
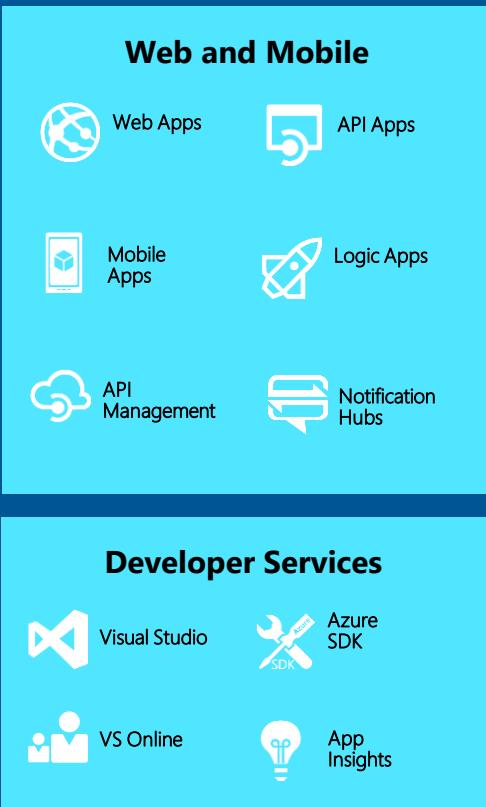
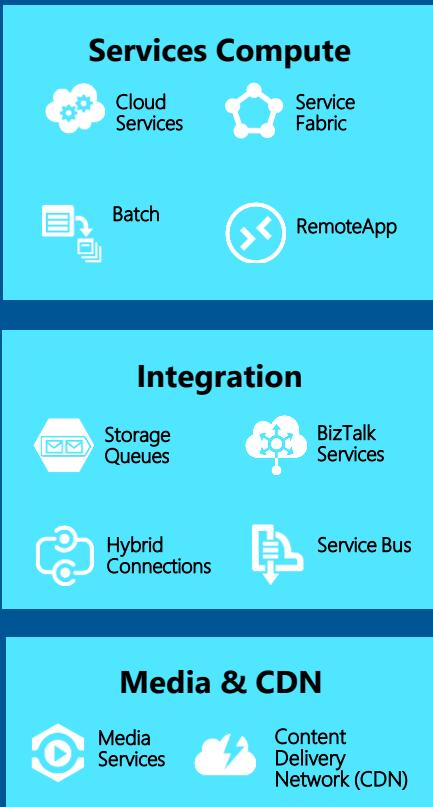
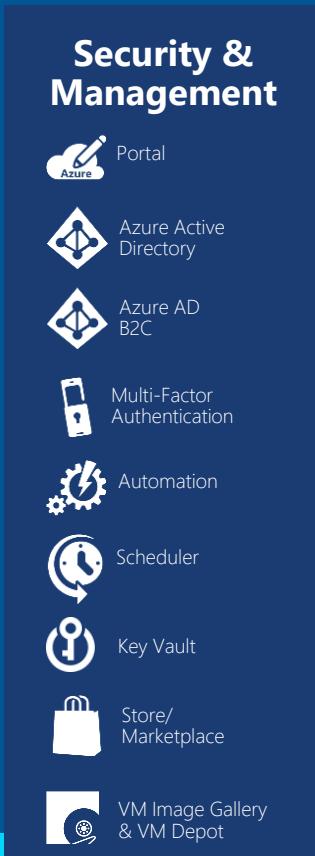


Microsoft



Customer

## Platform Services



## Infrastructure Services

### OS/Server Compute



### Storage



### Networking



## Datacenter Infrastructure (54 Regions, 100+ Datacenters)



**60+**

Azure regions

More than AWS &  
Google combined





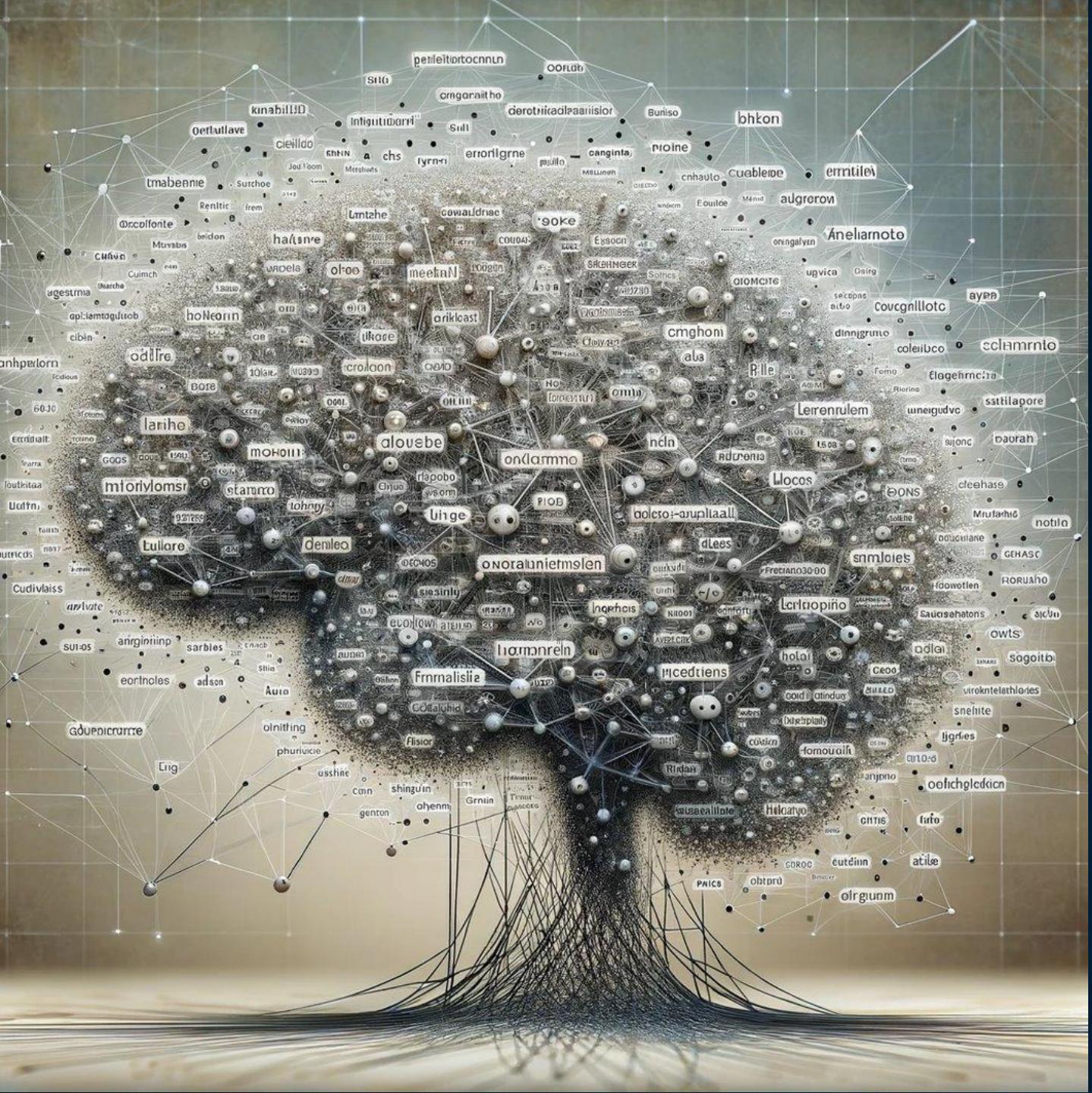
Data Center video

Questions?





# AI / LLM



Artificial Intelligence

Machine Learning

Deep Learning

Generative AI



## Artificial Intelligence

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



## Machine Learning

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



## Deep Learning

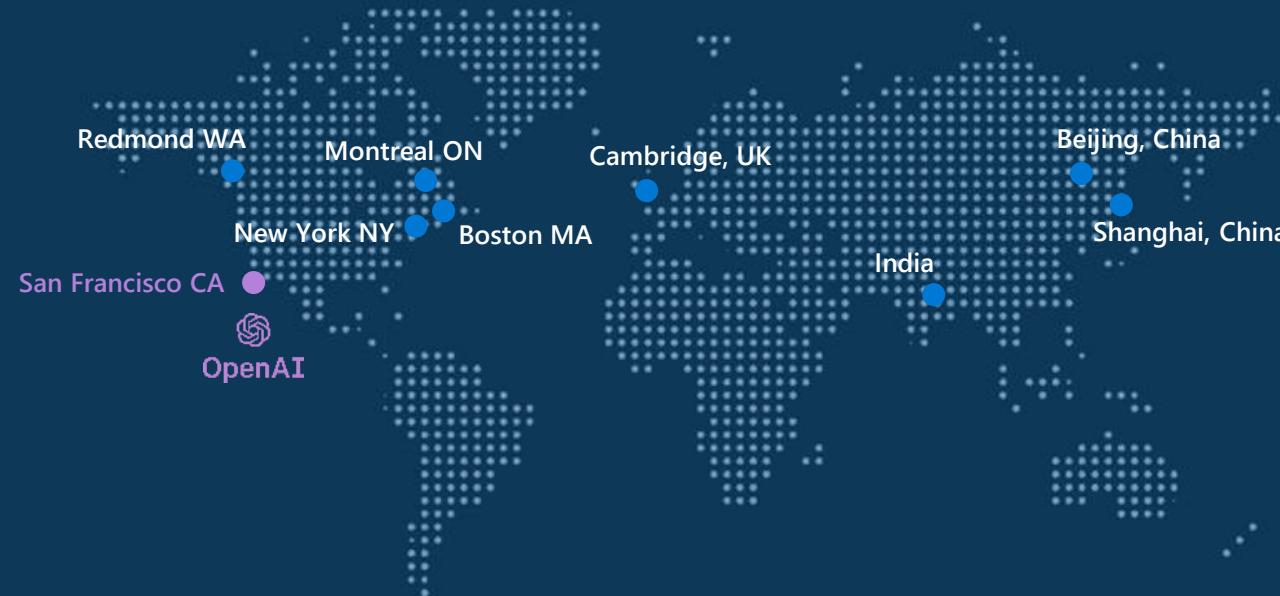
A machine learning technique in which layers of neural networks are used to process data and make decisions



## Generative AI

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

# Leveraging Decades of Experience from Microsoft Research... Recent Microsoft AI human parity milestones



May 2017  
**Human Parity**  
Speech  
Recognition

May 2018  
**Human Parity**  
Machine  
Translation

April 2019  
**Human Parity**  
Conversational  
QnA

July 2020  
**Human Parity**  
Image  
Captioning

January 2021  
**Human Parity**  
Natural Language  
Understanding

November 2021  
**Human Parity**  
CommonsenseQA

BENCHMARKS  
**94.9%**  
Switchboard speech  
recognition test

WMT Chinese-English  
news test

**92.46%\***  
Stanford CoQA  
challenge  
OpenAI & MSFT  
July 2019

**86.54%**  
CIDEr  
NOCAPS Challenge  
Azure Preview  
OpenAI  
June 2020

**90.3%**  
SuperGLUE  
benchmark

**89.4%**  
CommonsenseQA  
benchmark

# Large Language Foundation models

Microsoft AI at Scale

## Project Turing

17B  
Turing-NLG | 2020

340M  
BERT-Large | 2018



Our mission is to expand the boundaries of natural language understanding.

machine reading comprehension, question answering, transfer learning, reinforcement learning, computer vision, and even building interpretable models

Turing-NLG

Megatron NLG Turing

T-ULRv5

DeepSpeed

ZeRO

**175B**

GPT-3 | 2020

**17B**

Turing-NLG | 2020

**340M**

BERT-Large | 2018



**Microsoft**

**OpenAI**



**Microsoft**

**530B**

Megatron-Turing  
NLG | 2021



**175B**

GPT-3 | 2020

**17B**

Turing-NLG | 2020

**340M**

BERT-Large | 2018

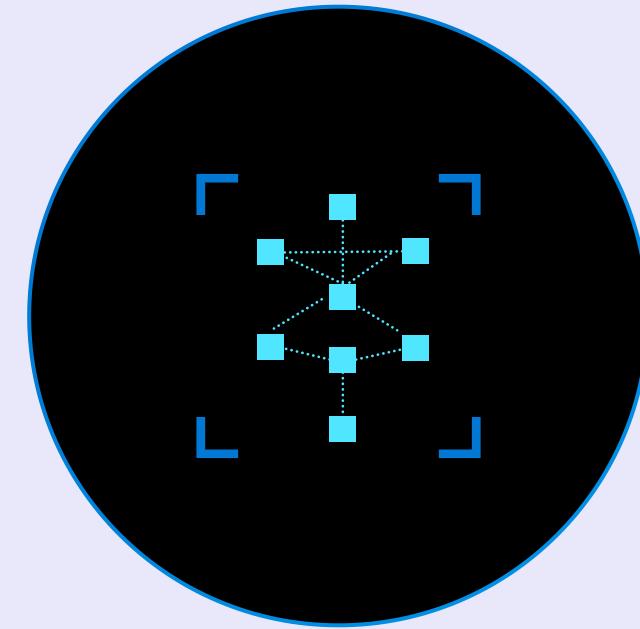
# Embeddings

An embedding is a special format of data representation that can be easily utilized by machine learning models and algorithms.

The embedding is an information dense representation of the semantic meaning of a piece of text.

Each embedding is a vector of floating-point numbers, such that the distance between two embeddings in the vector space is correlated with semantic similarity between two inputs in the original format.

For example, if two texts are similar, then their vector representations should also be similar.



# Embeddings make it possible to map content to a “semantic space”

A neutron star is the collapsed core of a massive supergiant star

A star shines for most of its active life due to thermonuclear fusion

The presence of a black hole can be inferred through its interaction with other matter

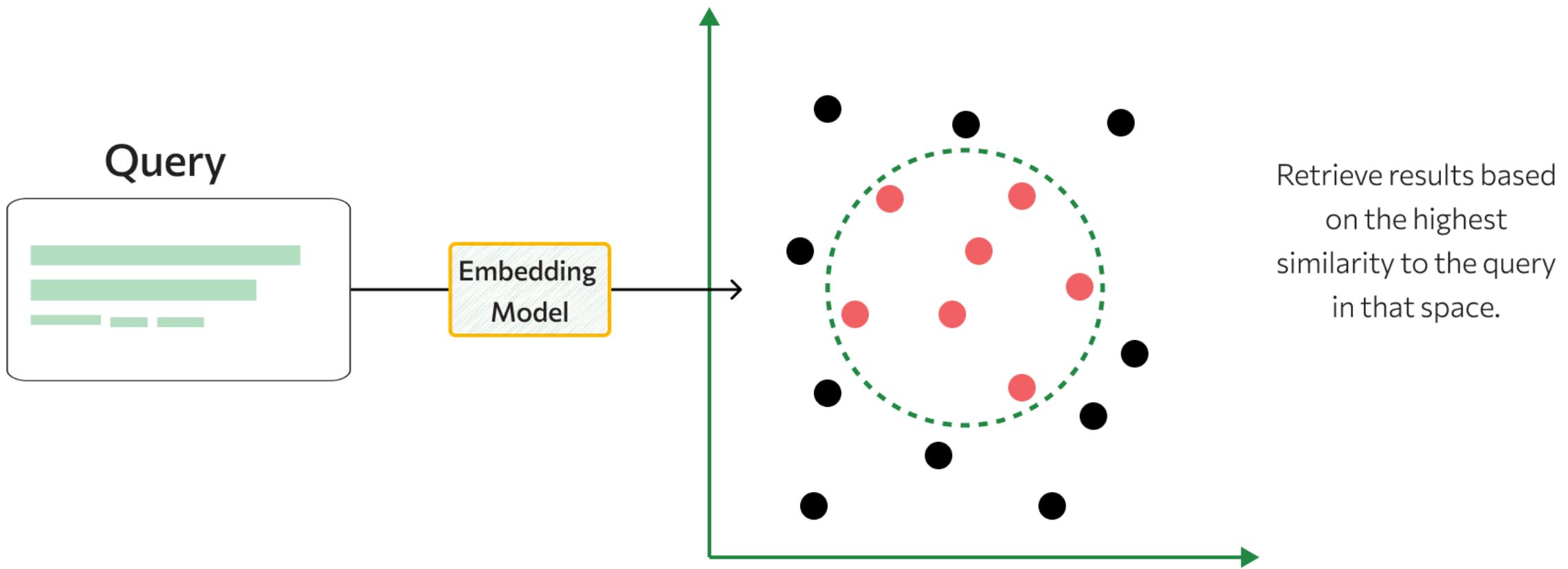


[ 15 34 24 13 ...]

[16 22 89 26 ...]

[ 20 13 31 89 ...]

## Embedding Vector Space





*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

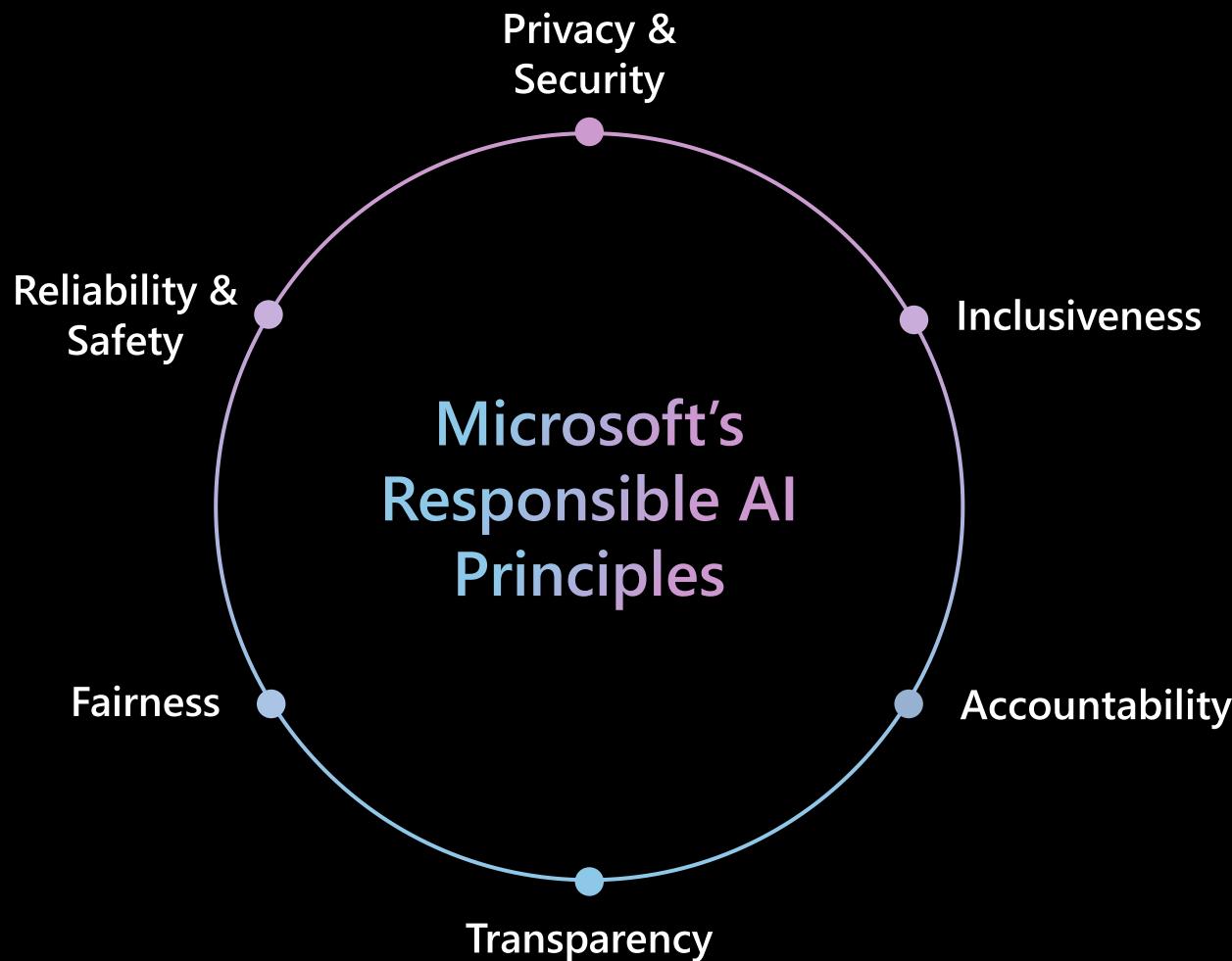
Conversation

Codex

Code

DALL·E 3

Images



## Building blocks to enact principles



Tools and processes



Training and practices

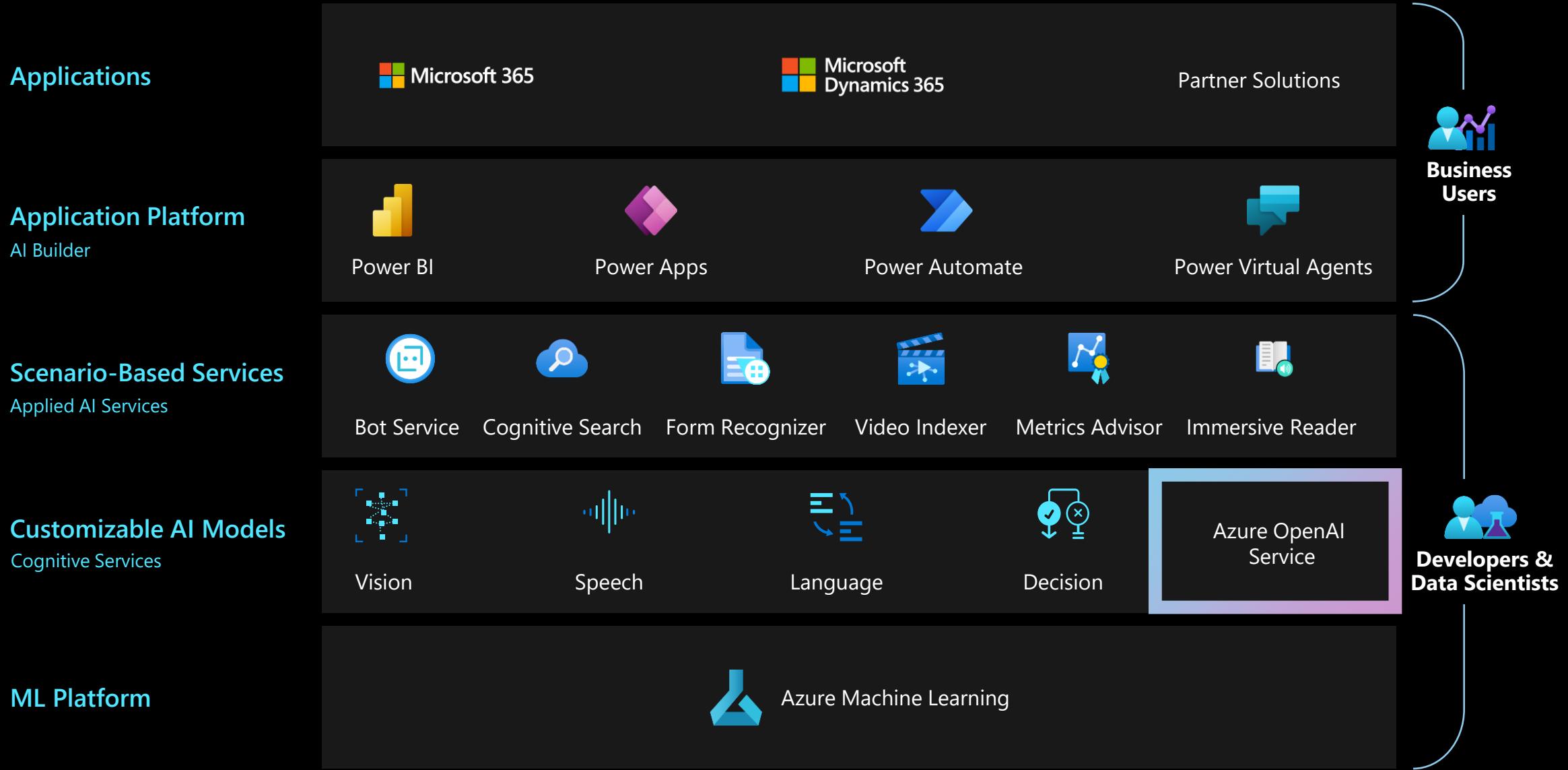


Rules

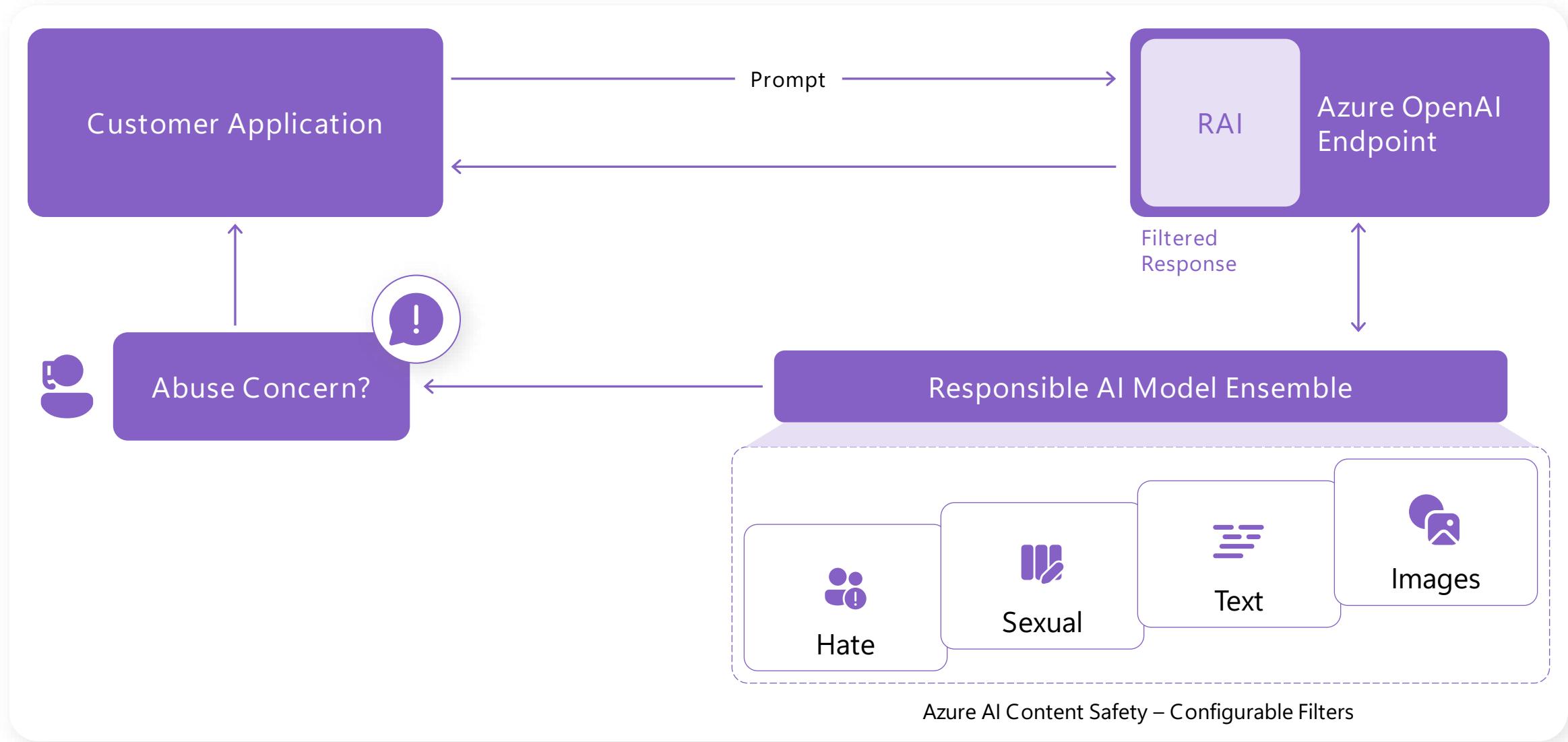


Governance

# Azure AI



# Responsible AI in Azure OpenAI Service



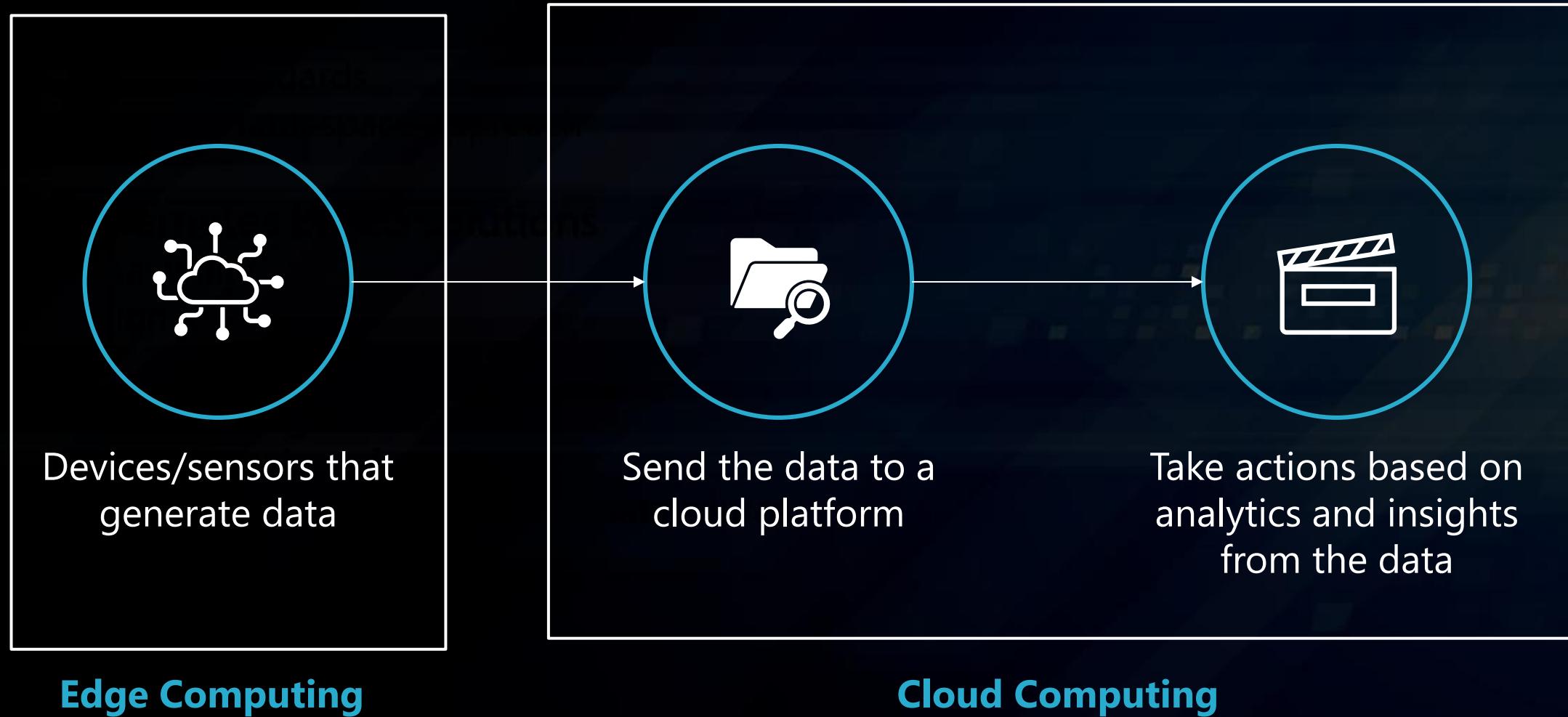
//

GitHub Copilot

# Edge Computing



# Internet of Things - overview





# Introduction to Edge Computing



# Introduction to Edge Computing

*"Edge computing is a distributed computing framework that allows IoT devices to quickly process and act on data at the edge of the network, in a decentralized manner."*

- Solves three interrelated challenges:
  - Connecting a device to a network from a remote location
    - Desert, ocean...
  - Slow data processing due to network or computing limitations
  - Edge devices causing network bandwidth issue
    - E.g., security cameras only sending relevant data to the cloud
- Use cases:
  - Autonomous driving
  - Vessels (remote areas)
  - Security cameras
  - Agriculture



# Benefits of Edge Computing

*"From workplace safety to security and productivity, the benefits of edge computing are vast."*

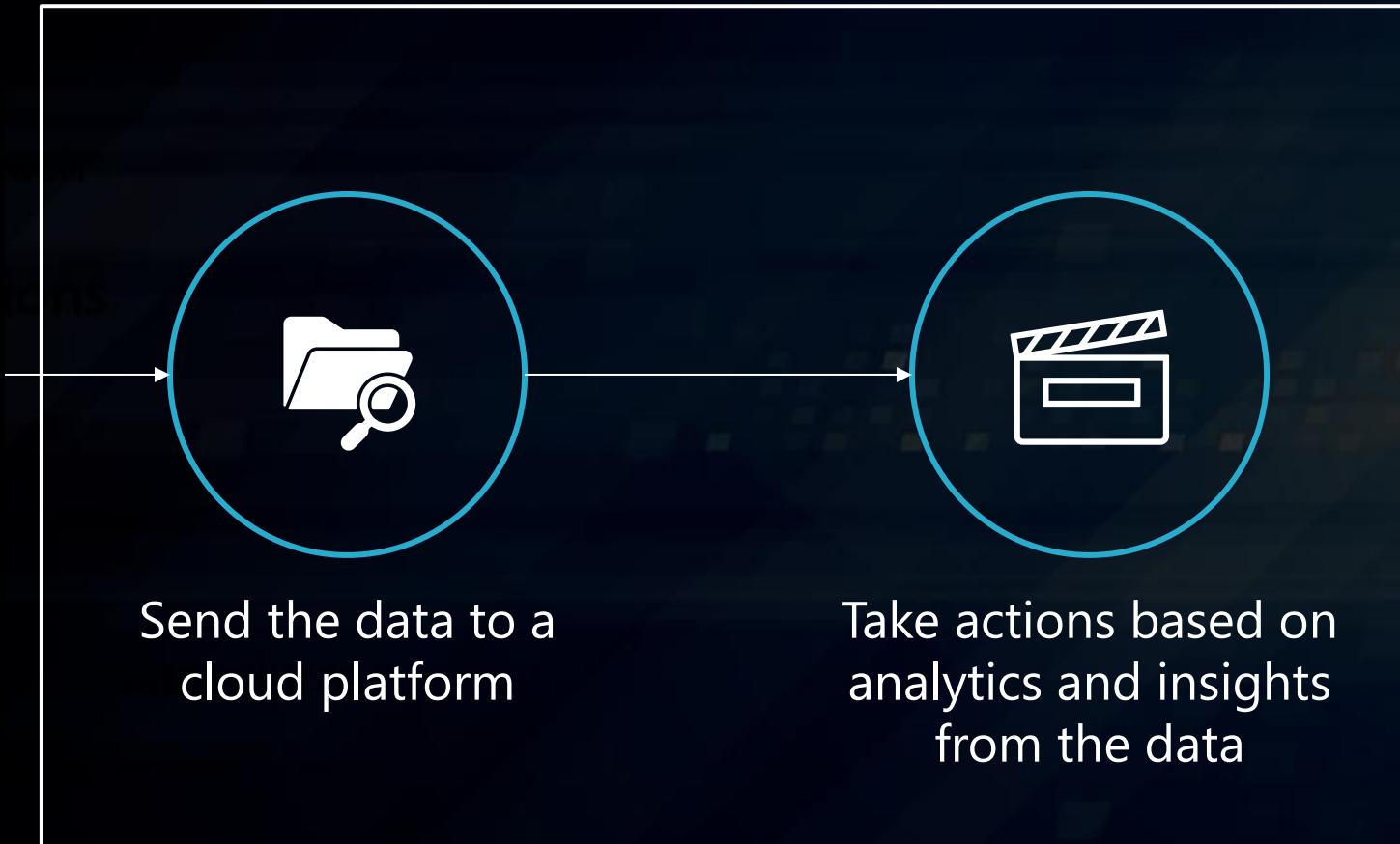
- More efficient operations
  - Day-to-day operations
- Faster response times, reduced latency
- Improved workplace safety
- Reduced IT costs
  - Less data sent to the cloud
  - Reduced bandwidth
- Increased reliability
  - Reducing reliance on a central cloud infrastructure
- Ability to operate in remote areas
  - Deserts
  - Oceans
  - Farming in rural areas



Questions?

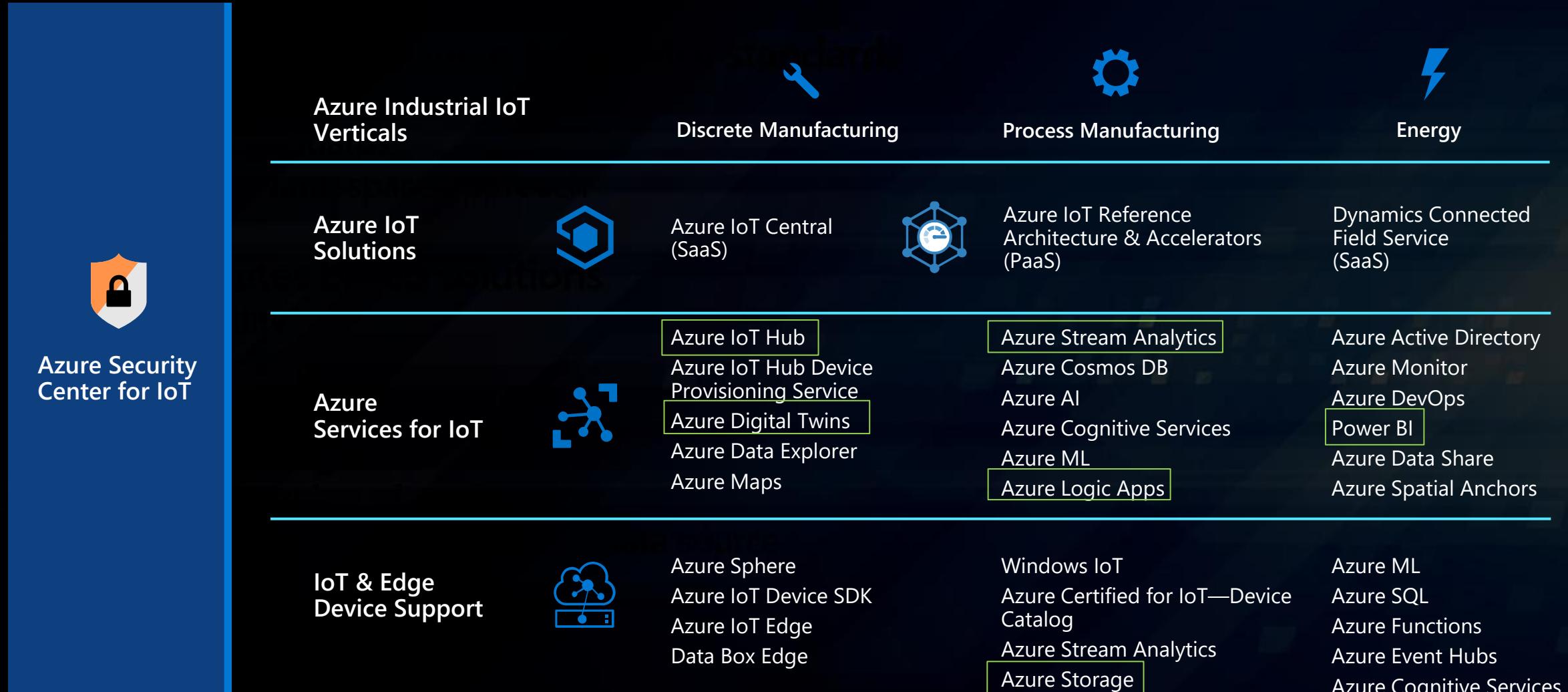


# Internet of Things - overview

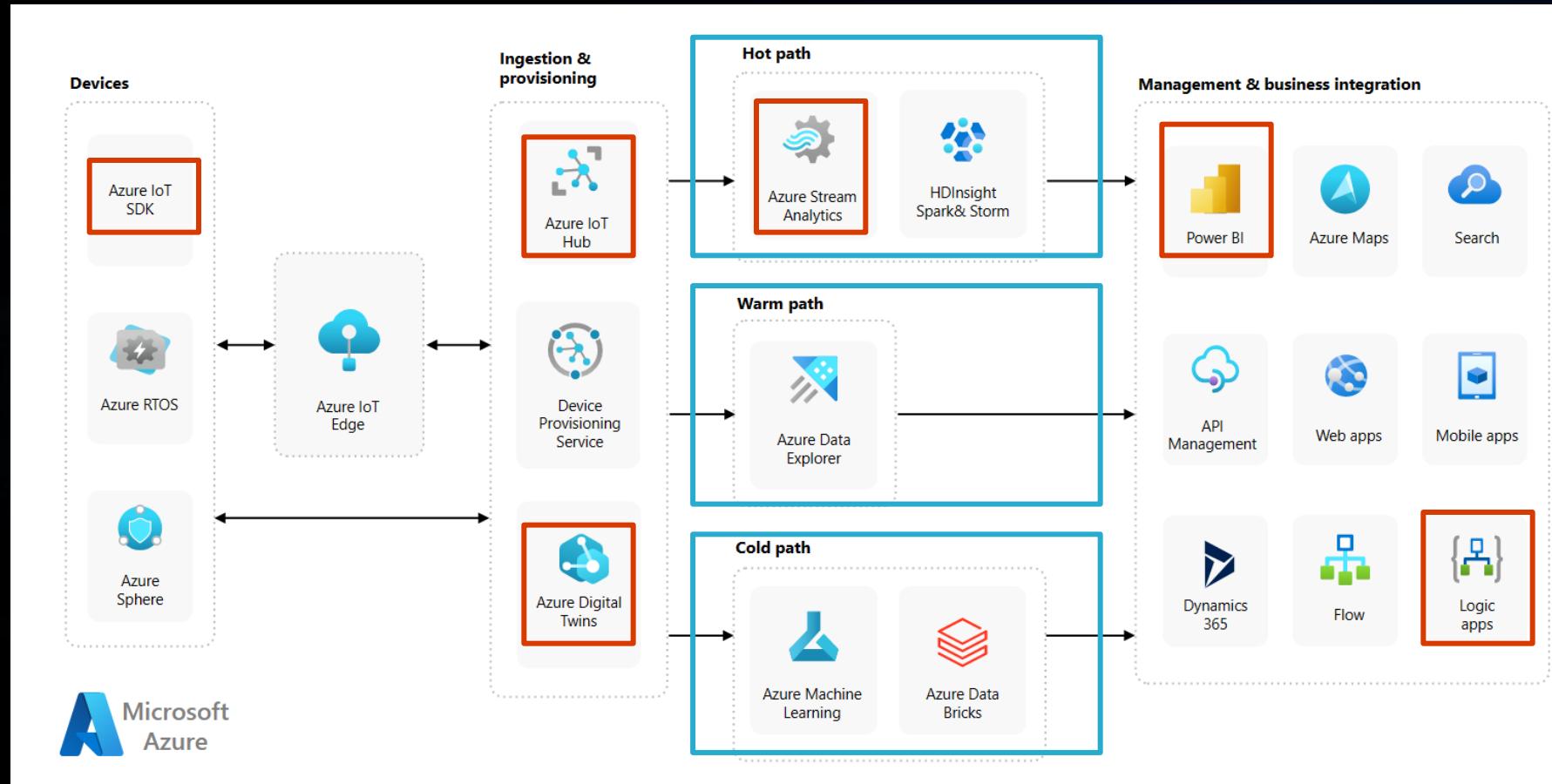


**Cloud Computing**

# Microsoft's comprehensive IoT product portfolio



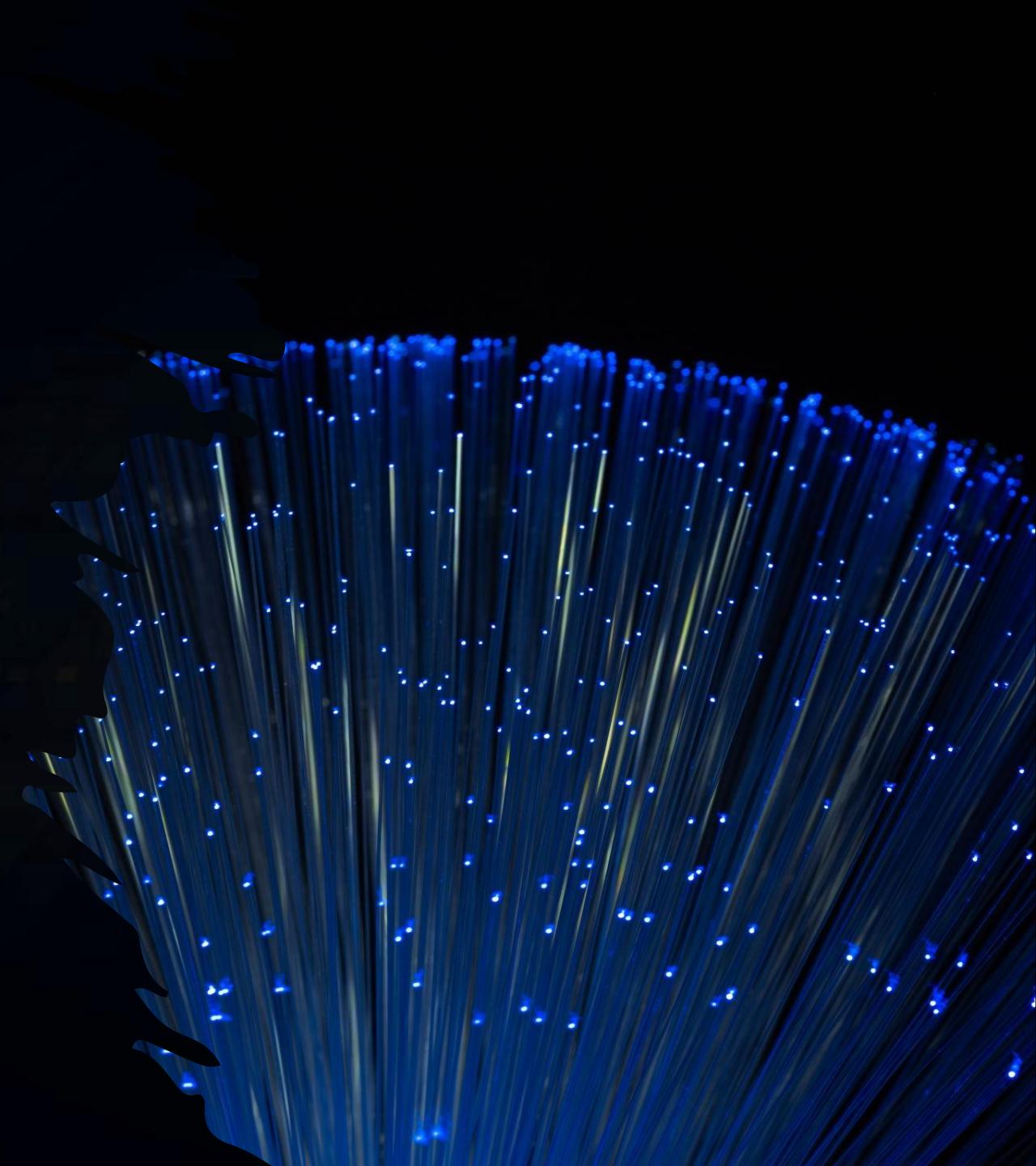
# Reference architecture



*Anyone that is familiar with one of the services mentioned on the slide?*

# Hot path Data flows

- Real-time data processing
- Typically uses a stream processing engine
  - E.g., Azure Stream Analytics
- Outputs can be
  - Alerting/notifications
  - Visualizations
  - Unstructured database for data analytics on historical data



# Warm path Data flows

- Near real-time data processing
  - More detailed processing
- Azure Data Explorer
  - Time-series data
  - Store large amounts of data
- Outputs can be
  - Time-series analytics
  - Visualization



# Cold path Data flows

- Batch processing at longer intervals, like hourly or daily
- Operates over large volumes of data, which can be stored in Azure Data Lake Storage
- Results don't need to be as timely as in the hot or warm paths
- Use ML to analyze this stored data

# Future of IoT at Microsoft

- Focus on Open Source & Industry standards
  - MQTT protocol
  - ISA 95 standards
  - Unified Namespace approach
- Kubernetes based solutions
  - Scalability
  - High availability
- AI at the Edge
  - Cloud: training of models
  - Edge: inferencing close to the data source

# Edge AI

## Computer vision at the Edge

- Safety/security cameras
- Autonomous driving

Agriculture: anomaly detection and monitor livestock in rural areas

Smart apps:

- Chatbot at the edge with out internet connection
- Predictive maintenance in the factories

Questions?



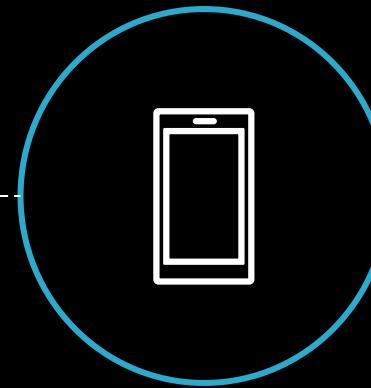
# Recap of the day



Cloud Computing



Azure (OpenAI)



Edge Computing



Azure (IoT) Platform

Questions?



## Agenda – Day 2

Introduction to Containerization  
Refresh on Generative AI & LLMs  
Break 10'  
Hands-on project (Docker, Azure & Web App)  
Wrap up & Q&A

# Introduction to Day 2 – Building E2E Solution

