

Microsoft Partner Project Ready

# Agentic AI Accelerator

■ Day 4 of 4

<Presenter Name>



# Course Plan and Learning Objectives

## Day 1

### Showcasing AI Potential with Agentic AI

#### Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

#### Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend

#### Hands-on Labs

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities

## Day 2

### Architecting Success with Multi-Agent AI Systems

#### Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK

#### Module 4: Innovate with Azure AI Platform

- How language models work
- AI Foundry and SDK introduction
- AI Foundry Model Catalog
- Azure AI Services
- Azure OpenAI Service and model guidance
- Models-as-a-Service
- Azure AI Foundry Agent Service
- Safeguard with Trustworthy AI

#### Hands-on Labs

- Setup AI Project and perform Chat Completion from VS Code
- Build a simple AI Agent
- Develop a multi-agent system

## Day 3

### Multi-Agent AI: Advanced Agent Dev in Azure AI Foundry

#### Module 5: Customize, orchestrate and experiment with Azure AI Foundry

- Retrieval Augmented Generation (RAG)
- Customizing models – Fine tuning, distillation
- Responses API (preview)
- Azure AI Foundry Agent Service - Orchestrate and debug AI workflows

#### Module 6: Build your own multi agents with Semantic Kernel or AutoGen

- Multi-agent applications
- Understanding Semantic Kernel
- Understanding AutoGen Agents Framework
- Multi-Agent Collaboration & Orchestration with AutoGen / Semantic Kernel

#### Hands-on Labs

- Set Up Azure AI Foundry SDK and Provision Resources
- Build a Retrieval-Augmented Generation(RAG) Pipeline
- Evaluate and Optimize RAG Performance
- Semantic Kernel Fundamentals
- Semantic Kernel Plugins

## Day 4

### Enterprise Grade: Optimization and production at scale

#### Module 7: Enterprise grade production at scale

- Scaling challenges and agent controls
- Manage AI performance in production
- Observability Tools
- Enabling Enterprise governance and management
- Enterprise grade security and data protection
- Monitoring and observability

#### Module 8: Advanced AI risk evaluation and mitigation

- Identifying risks
- Azure AI Content Safety
- Evaluation and GenAIops
- Identity and access management
- Network Security for AI apps
- Continuous security for AI

#### Hands-on Labs

- Understanding the Lifecycle of Flow Development
- Building and Customizing Prompt Flows
- Evaluation Flow Setup
- Fine-Tuning Prompts for Optimal Performance
- Implementing Chat Flow and Tool Integration
- Ensuring Responsible AI Practices with Content Safety

Journey A – Deal-ready

Deal-ready assessment 

Journey B – Solution Design

Solution Architecture assessment 

Journey C – Project-ready

Capstone Project 



# Building Intelligence: Advanced Agent Development in Azure AI Foundry

⌚ 4 days (5 hours)

## Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

## Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend

## Hands-on Labs ⌚ 3 hours

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities

## Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK

## Module 4: Innovate with Azure AI Platform

- AI Foundry and SDK introduction
- How language models work
- AI Foundry Model Catalog
- Azure AI Services
- Azure OpenAI Service and model guidance
- Models-as-a-Service
- Azure AI Foundry Agent Service
- Safeguard with Trustworthy AI

## Module 5: Customize, orchestrate and experiment with Azure AI Foundry

- Retrieval Augmented Generation (RAG)
- Customizing models – Fine tuning, distillation
- Responses API (preview)
- Azure AI Foundry Agent Service - Orchestrate and debug AI workflows

## Module 6: Build your own multi agents with Semantic Kernel or AutoGen

- Multi-agent applications
- Understanding Semantic Kernel
- Understanding AutoGen Agents Framework
- Multi-Agent Collaboration & Orchestration with AutoGen / Semantic Kernel

## Hands-on Labs ⌚ 3 hours

- Set Up Azure AI Foundry SDK and Provision Resources
- Build a Retrieval-Augmented Generation(RAG) Pipeline
- Evaluate and Optimize RAG Performance
- Semantic Kernel Fundamentals
- Semantic Kernel Plugins

## Module 7: Enterprise grade production at scale

- Scaling challenges and agent controls
- Manage AI performance in production
- Observability Tools
- Enabling Enterprise governance and management
- Enterprise grade security and data protection
- Monitoring and observability

## Module 8: Advanced AI risk evaluation and mitigation

- Identifying risks
- Azure AI Content Safety
- Evaluation and GenAIOps
- Identity and access management
- Network Security for AI apps
- Continuous security for AI



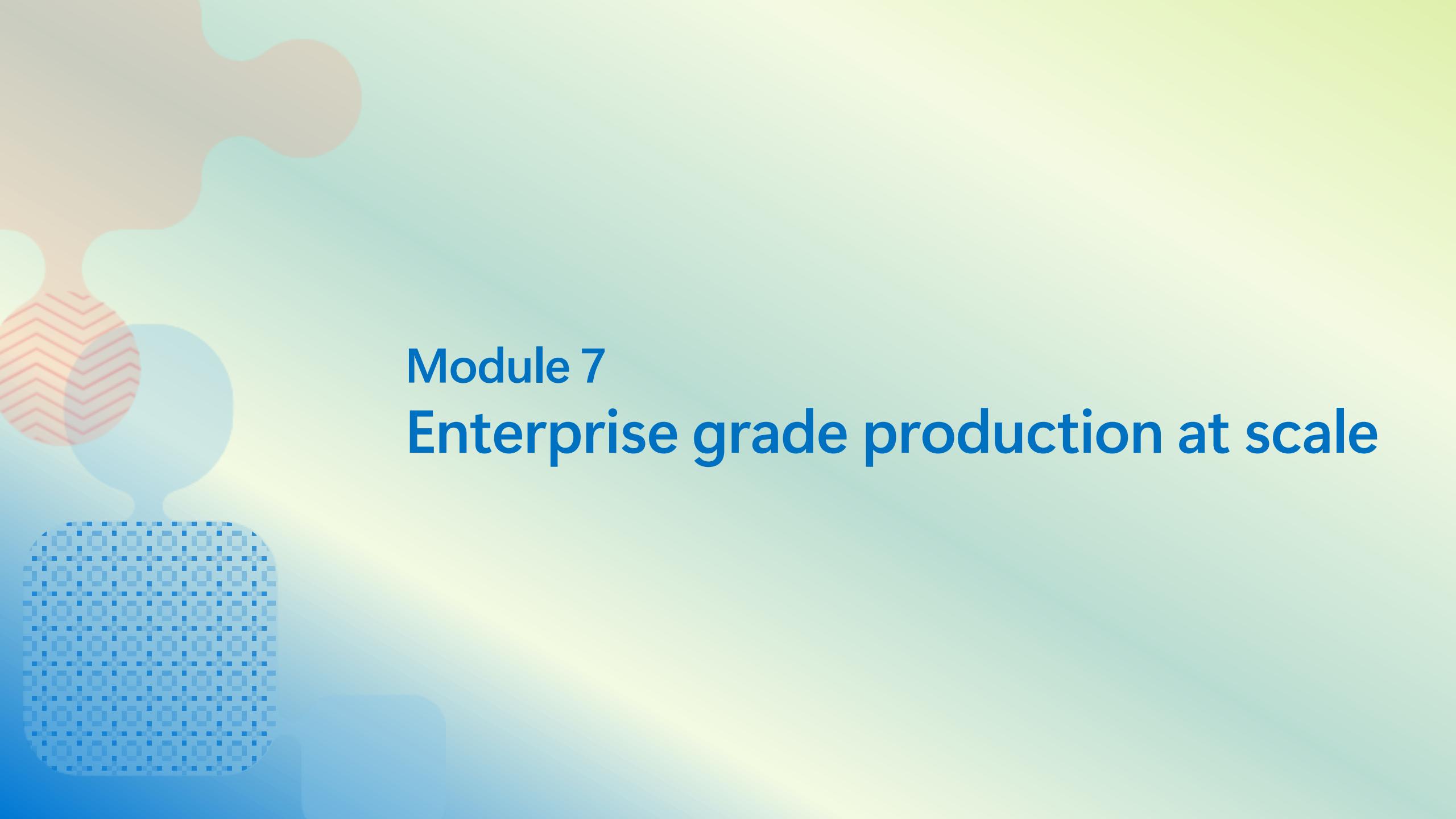
## Capstone Project

**Objective:** Apply your skills to design, develop, and deploy a sophisticated, end-to-end AI solution leveraging Azure AI services, Microsoft Copilot frameworks, and multi-agent orchestration techniques

## ⌚ 3 hours

## Hands-on Labs ⌚ 3 hours

- Understanding the Lifecycle of Flow Development
- Building and Customizing Prompt Flows
- Evaluation Flow Setup
- Fine-Tuning Prompts for Optimal Performance
- Implementing Chat Flow and Tool Integration
- Ensuring Responsible AI Practices with Content Safety



# **Module 7**

# **Enterprise grade production at scale**

# Unscalable AI limits impact and growth



Scaling challenges



Operational inefficiencies



Limited developer agility

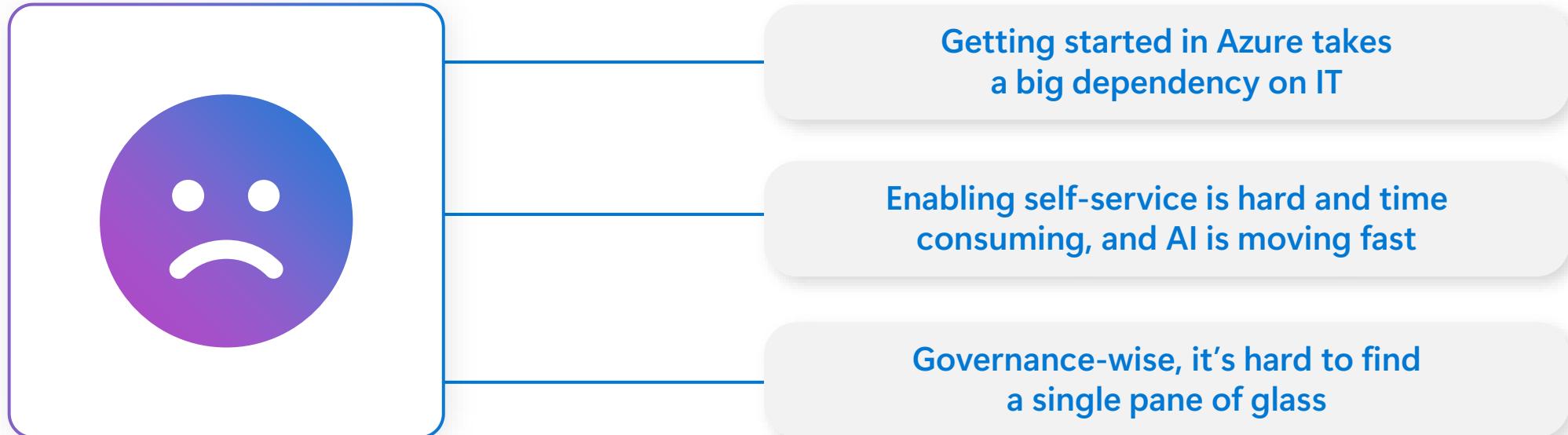


Governance issues



Fragmented workflows

# Key challenge: Governance and lack of self-serve



# Copilot Control System



## Copilot + Agents



### Security & Governance

Data Security

AI Security

Compliance & privacy



### Management Controls

Copilot licensing

Agent lifecycle

Metering controls



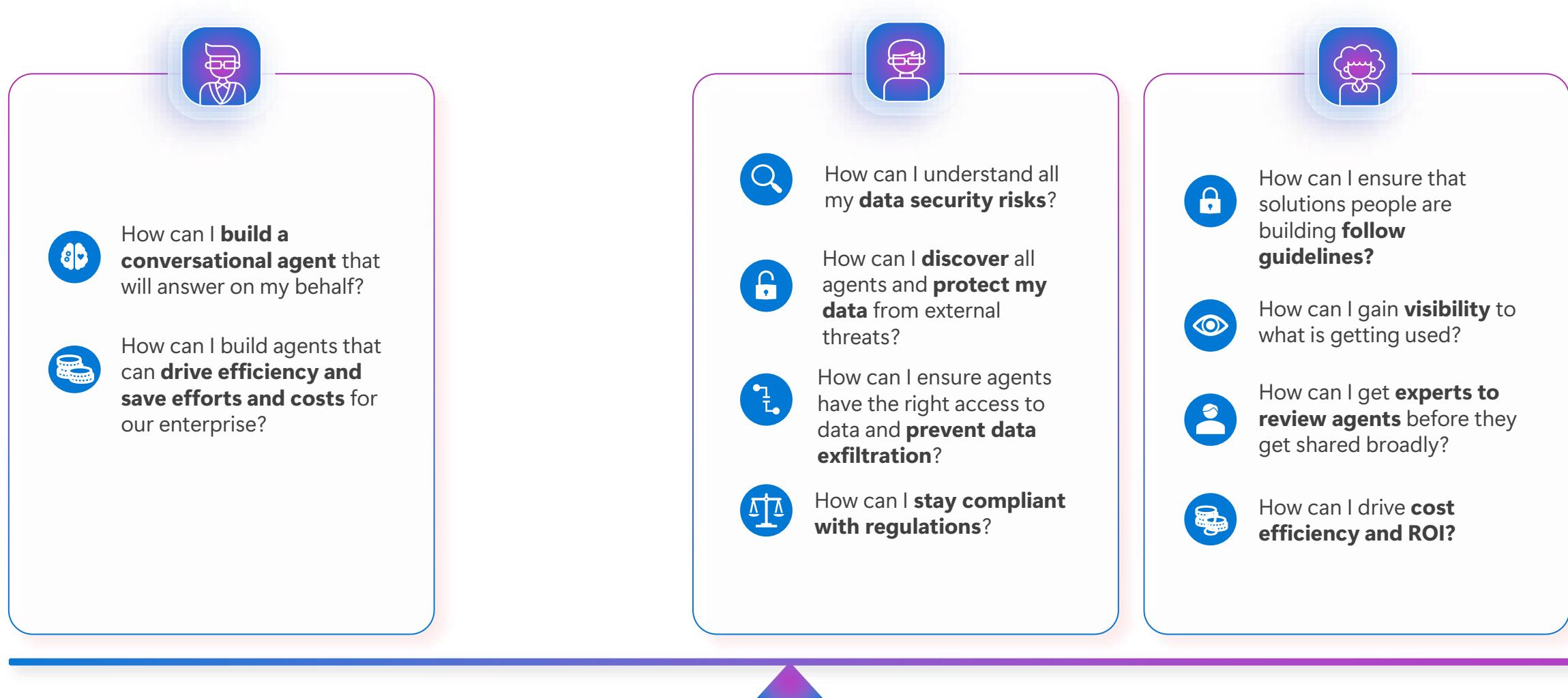
### Measurement & Reporting

Readiness and adoption

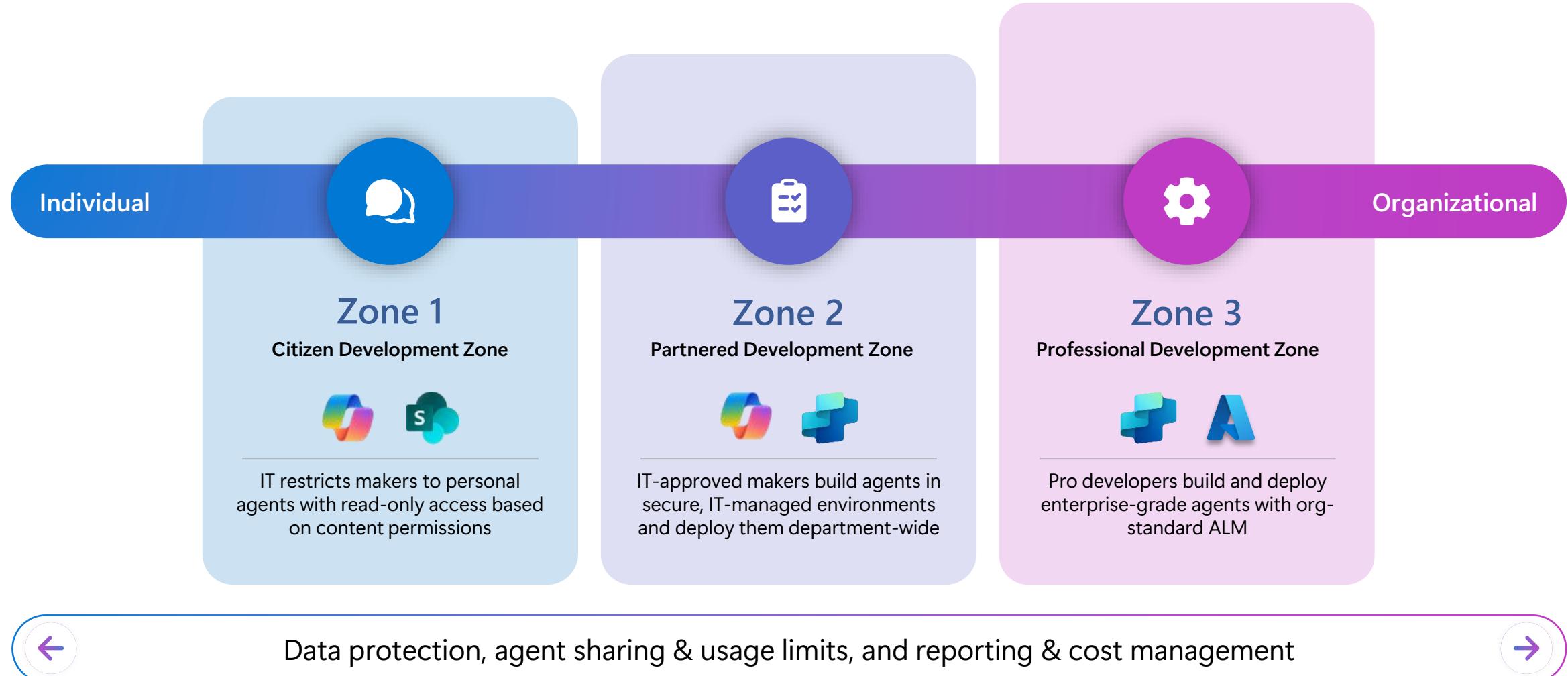
Productivity shifts

Business value & ROI

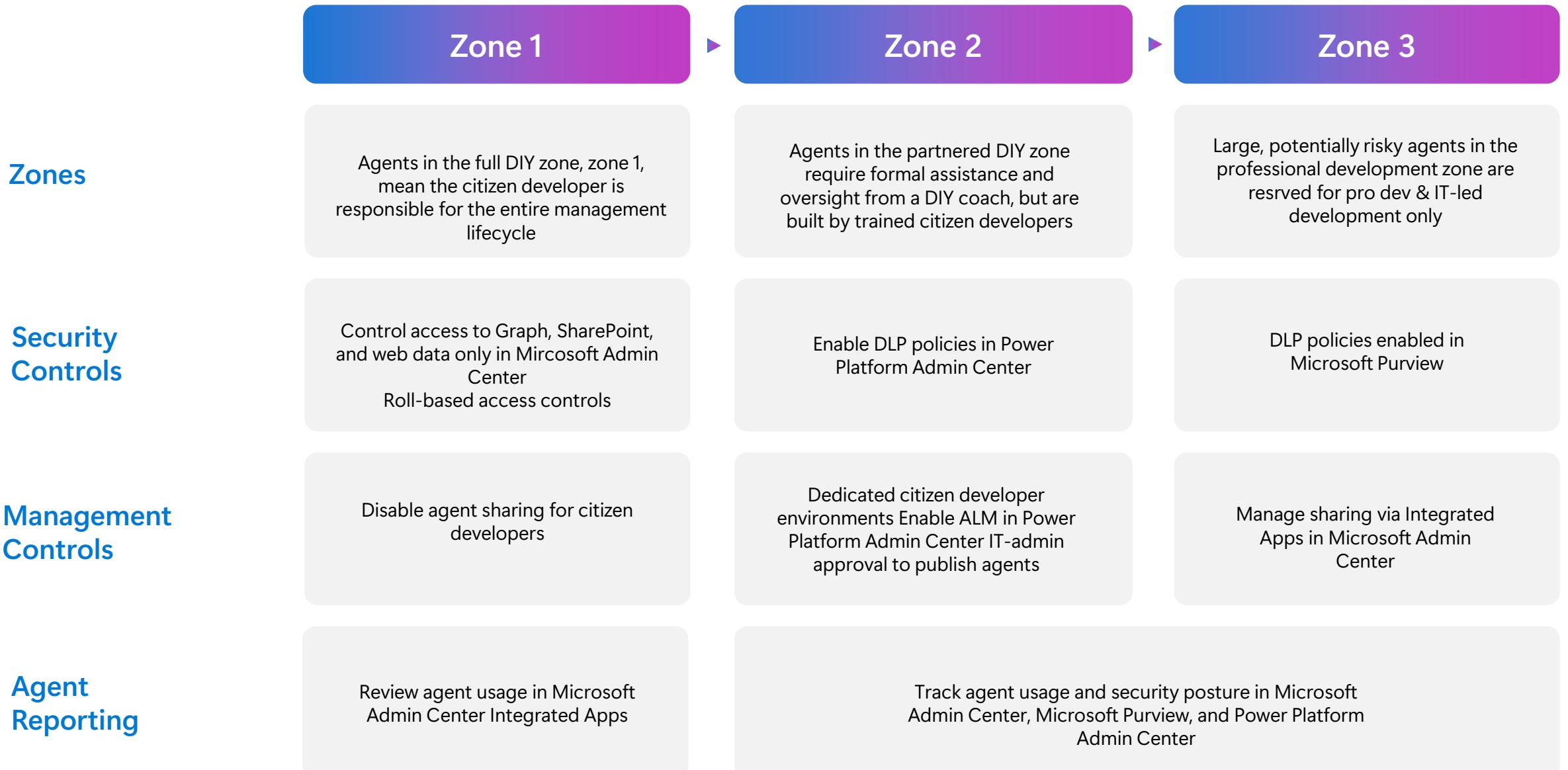
# Balancing Governance and Innovation



# Agent Controls Model



# IT Controls



# Managing AI performance in production



# Manage AI performance in production

Deploy with continuous monitoring and governance across environments

## Enable enterprise governance

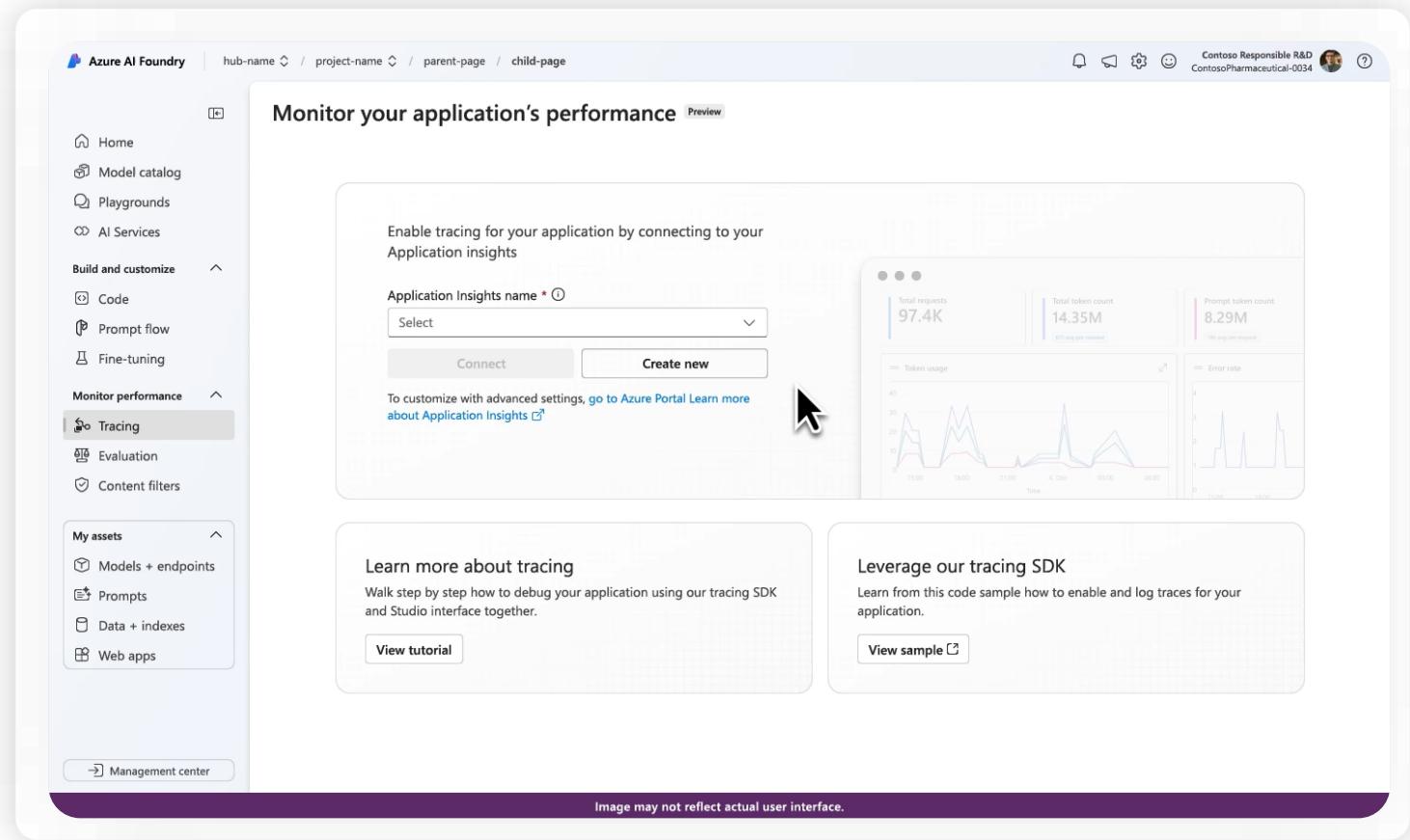
Provide easy project creation and resource management across the organization and enterprise controls for security, privacy, and compliance.

## Scale in production

Scale AI for use in websites, applications, and other production environments

## Scale in production

Gain ongoing, comprehensive insight of application performance, including token usage, cost, latency, user feedback and the quality and safety of generated outputs





# Manage and optimize generative AI applications with ease



## Observability

Enable cross-functional teams to **continuously** evaluate, monitor, debug, and optimize generative AI apps.



## IT Governance

Ensure resources are used **effectively and efficiently** to fuel secure, private, and safe AI innovation.

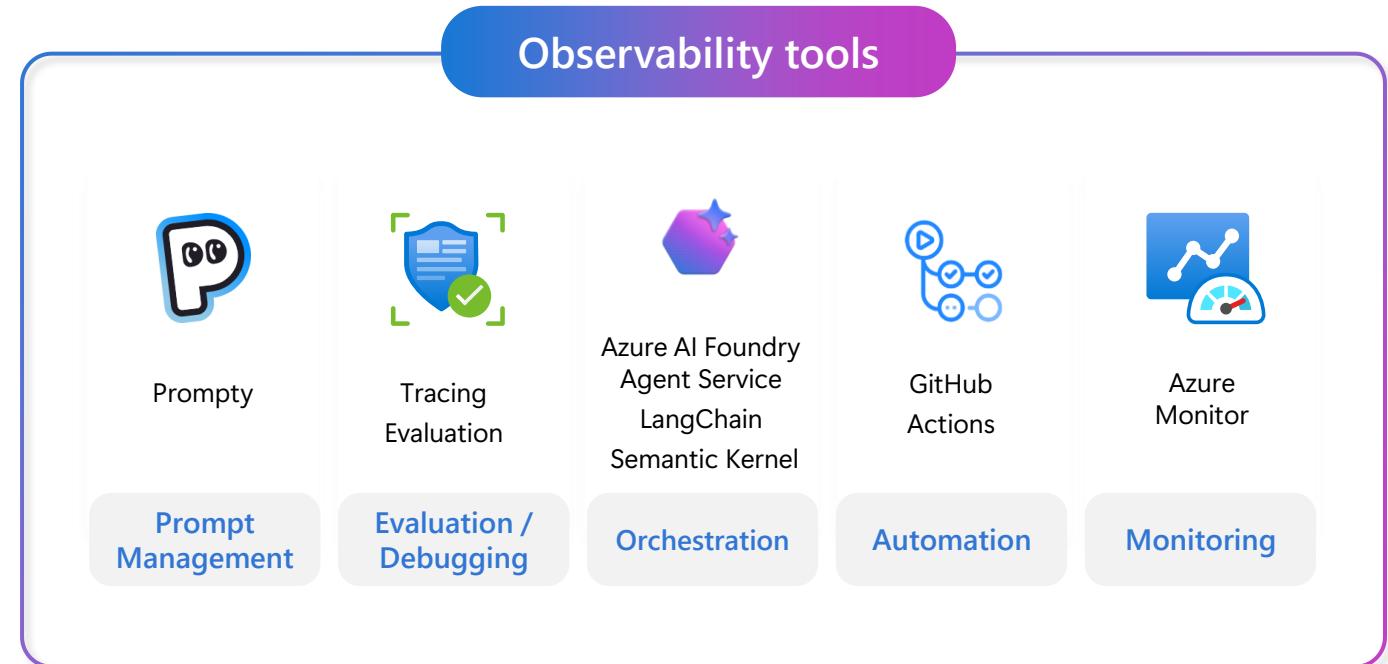


## AI Governance

Align AI development **processes and outcomes** with internal policies and legal requirements.

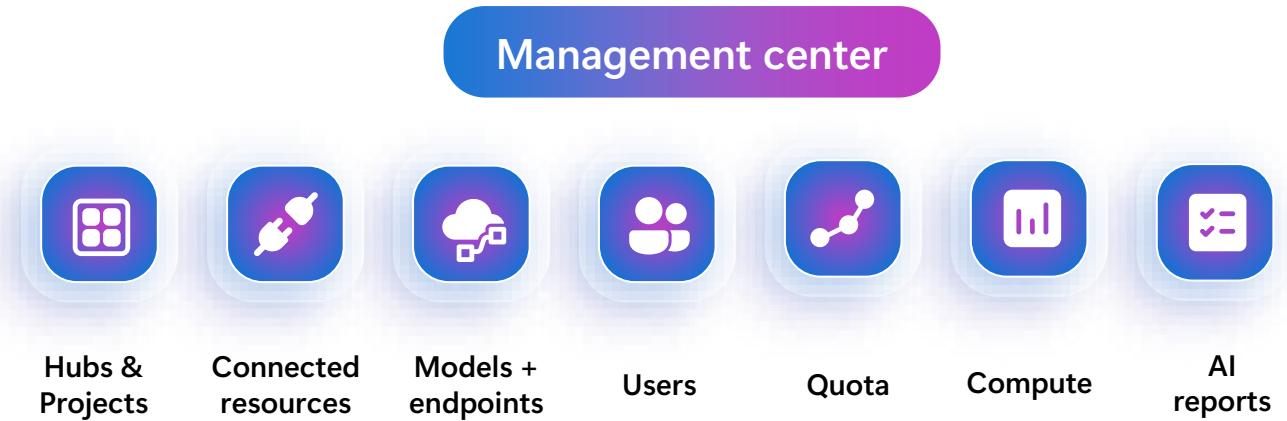


# Build production apps with end-to-end observability and simplified governance



## The Azure AI Foundry difference:

- Observability throughout the GenAIOps lifecycle
- Centralized management and governance experience



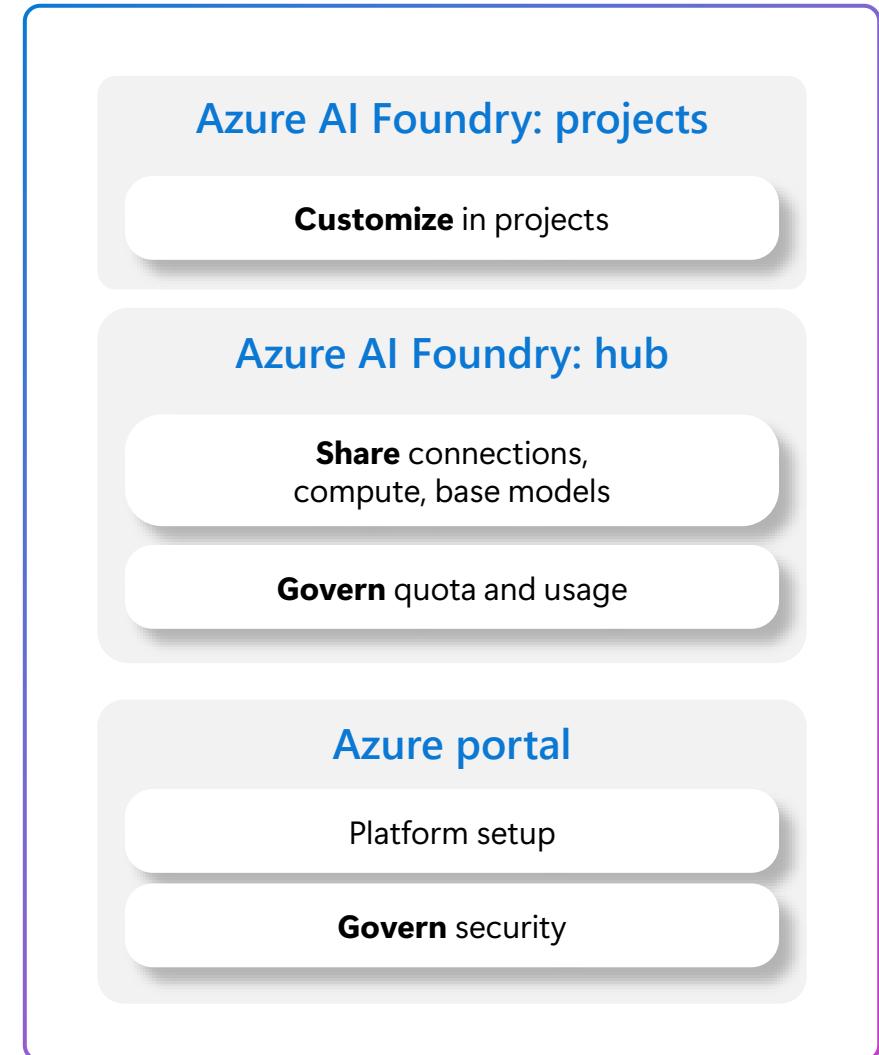
# Enabling enterprise governance and management



# Management layers in Azure AI Foundry

Azure AI Foundry targets three different management needs:

- Provide **AI developers and business stakeholders** with a SaaS-like self-serve experience, to allow for rapid AI experimentation
- Provide **team leads** with central configuration and governance for managing capacity, spend, shareable assets for their team
- Provide a compliant, yet non-repetitive or duplicate setup by **IT security** using templates



# Azure AI Foundry Hubs

## AI hub resource

Collaboration environment for a team to share project work, model endpoints, compute, (data) connections, security settings, govern usage.

Data upload and artifact storage

Hub-scoped connections

Base model endpoints

Compute

Security setup + governance

## AI services

Managed provider of Microsoft-maintained base models.  
Shows in your Azure subscription as a separate resource.

1    0..N

## AI project

(**Azure ML RP/kind="project"**)

Groups components for model customization.  
Orchestrates and provides tools for customization.  
Isolates data and access.

Components (datasets, models, indexes, ...)

Isolated data container

Project-scoped connections

Open-source + fine-tuned model endpoints

Connected compute

Azure AI services

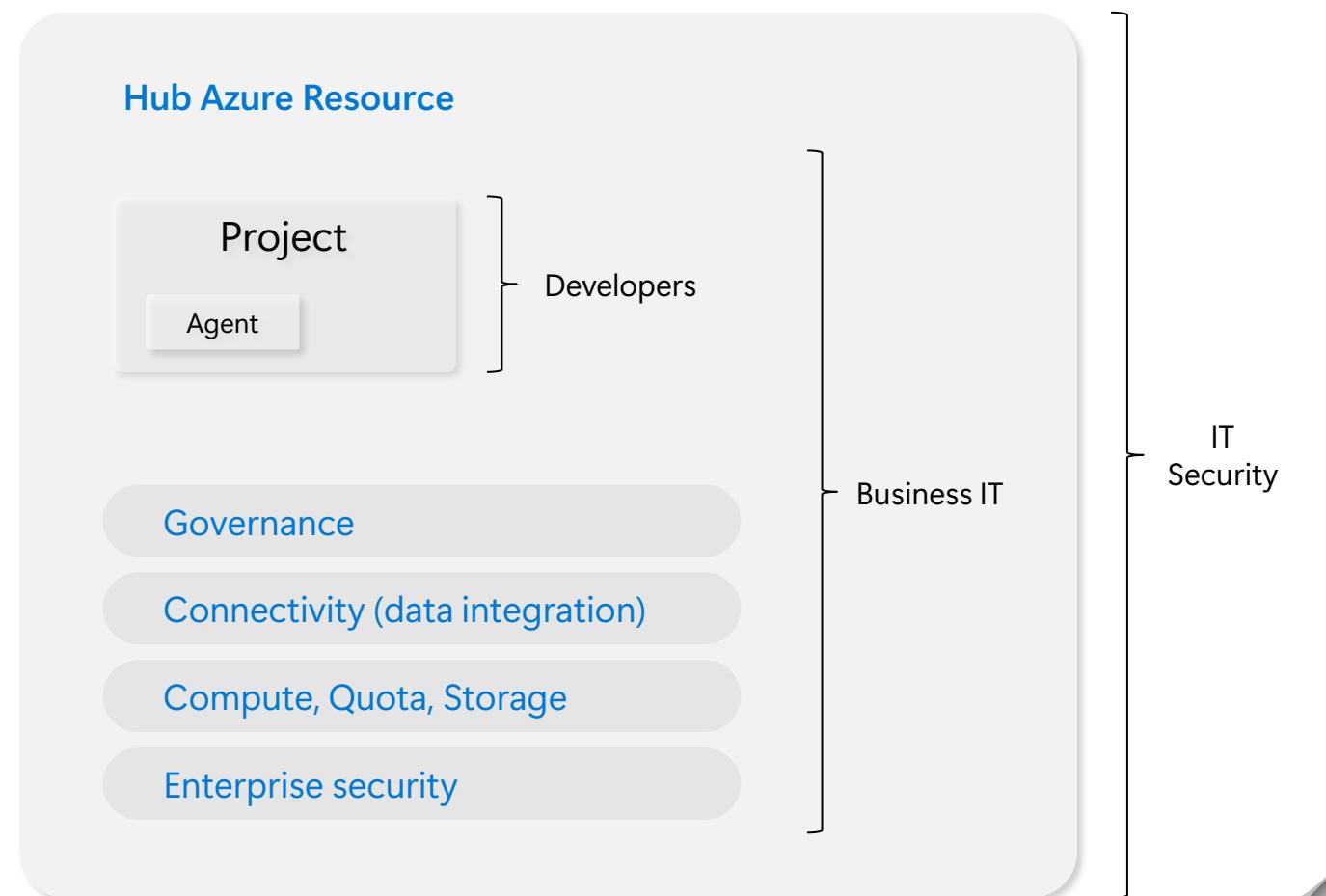
Azure AI services



# Enterprise setup in Azure AI Foundry

Seeks to balance a self-serve authoring experience for developers in Azure AI Foundry, with granular platform controls for IT using Azure Portal/IaC templates

Tenant, Management Groups, Subscriptions



---

**Hub:** configures Azure infrastructure, enterprise security, storage, connectivity for data integration. Groups projects for a team

**Projects:** Provides a place for developers to collaborate and organize their work for customizing AI models. Container for data upload and access control

# Connections



1

Manage access to Microsoft and non-Microsoft sources through connections. Connections allow users to specify target sources while securely store authentication information in Key Vault



2

IT Admins and managers can configure connections for the entire AI Foundry in one place, allowing all teams to access approved sources. AI Developers can further create project or cross-project shared connections based on their individual project needs



3

Use connections to create model deployments/access data sources or consume resources throughout the AI Foundry

# Organize customization work in projects

Enable enterprise governance

**Create projects** from AI Foundry, Azure Developer CLI (azd), Azure AI Foundry SDK, templates to access all customization tools. Use projects to save your work and collaborate with others

**Upload data in isolation:** every project comes with its own storage space

**Access company resources right away:** every project can reuse connectivity pre-configured on the hub. Optionally, set exclusive per-project resource access

The screenshot shows the Azure AI Foundry Management center interface. On the left, there's a sidebar with 'Management center' and 'All hubs + projects'. Under 'All hubs + projects', it lists 'Hub' and 'Project' types. The main area is titled 'Manage your hubs and projects' with a search bar and filter options. It displays a table with columns: Name, Type, Hub, Subscription, and Resource group. The table contains several entries, each with a small thumbnail icon and a 'Go to project' button.

Name	Type	Hub	Subscription	Resource group
Using AI 101	Hub		West US 2	contoso
Hub for isolated environments	Hub		West US 2	sg-isolate
Project 101	Project	Screenwriting, Animation	West US 2	sg-isolate
Screenwriting, Animation	Hub		West US 2	sg-isolate
Image 101	Project	Hub for isolated environments	West US 2	sg-isolate

# Management center demo

The screenshot shows the Azure AI Foundry Management center interface. The left sidebar has a tree view with 'All resources', 'AI governance', 'Quota', 'Hub (Customer service...) Overview', 'Users', 'Deployments' (which is selected), 'Connected resources', 'Compute', 'Project (POC\_llm\_base) Overview', 'Users', 'Deployments' (selected again), 'Connected resources', and 'Reports'. A 'Go to project' button is at the bottom. The main area is titled 'Deployments' and shows a table of web app deployments. The table columns are: Name, Model, Type, Provider, Content filter, Modified on, Capacity, and Rate limit (token per second). The table contains the following data:

Name	Model	Type	Provider	Content filter	Modified on	Capacity	Rate limit (token per second)
Dancing_piranha_1							
gpt-4o	gpt-4o	Serverless	Azure OpenAI	Off	Sep 19, 2023 11:25 AM	3K TPM	-
llama_deploy_1	GPT-35	Serverless	Meta	Off	Sep 19, 2023 11:25 AM	3K TPM	-
Fancy_flamingo_1							
Mistral_1	GPT-35	Standard	Mistral	Default	Aug 19, 2023 11:25 AM	2K TPM	-
phi-3.5	GPT-35	Standard	Microsoft	Default	Aug 19, 2023 11:25 AM	2K TPM	-
Fancy_flamingo_1							
vision_36	phi-3.5-vision	Standard	Microsoft	Default	Aug 19, 2023 11:25 AM	2K TPM	-
phi-3.5	Phi-3.5-instruct	Standard	Microsoft	Default	Aug 19, 2023 11:25 AM	2K TPM	-

At the bottom of the main area, there is a note: 'Image may not reflect actual user interface.'

# IT governance at scale

Enterprise governance provides a balance of SaaS-like self-serve experiences for dev and business teams, customizable configurations for managers, and policy management and oversight by IT teams for enhanced agility, security, and compliance.

**Self-serve development for a team:** Centrally govern security and shared resource connectivity for a team using hubs, reducing IT bottlenecks

**Identity & access management:** Leverage built-in roles to manage access for cross-functional teams within collaborative development environments

**Network security:** Streamline your network isolation experience, speed up your workspace setup, and free IT from the hassles of VNET management

**Data protection & encryption:** Secure data in transit and at rest by default and use provided encryption keys or your own customer-managed keys

**Plan deployments:** Get Infra as code templates and landing zone reference implementations

**Costs and quotas:** Centrally plan and manage costs, quotas, and autoscaling for cost efficiency

Enable enterprise governance

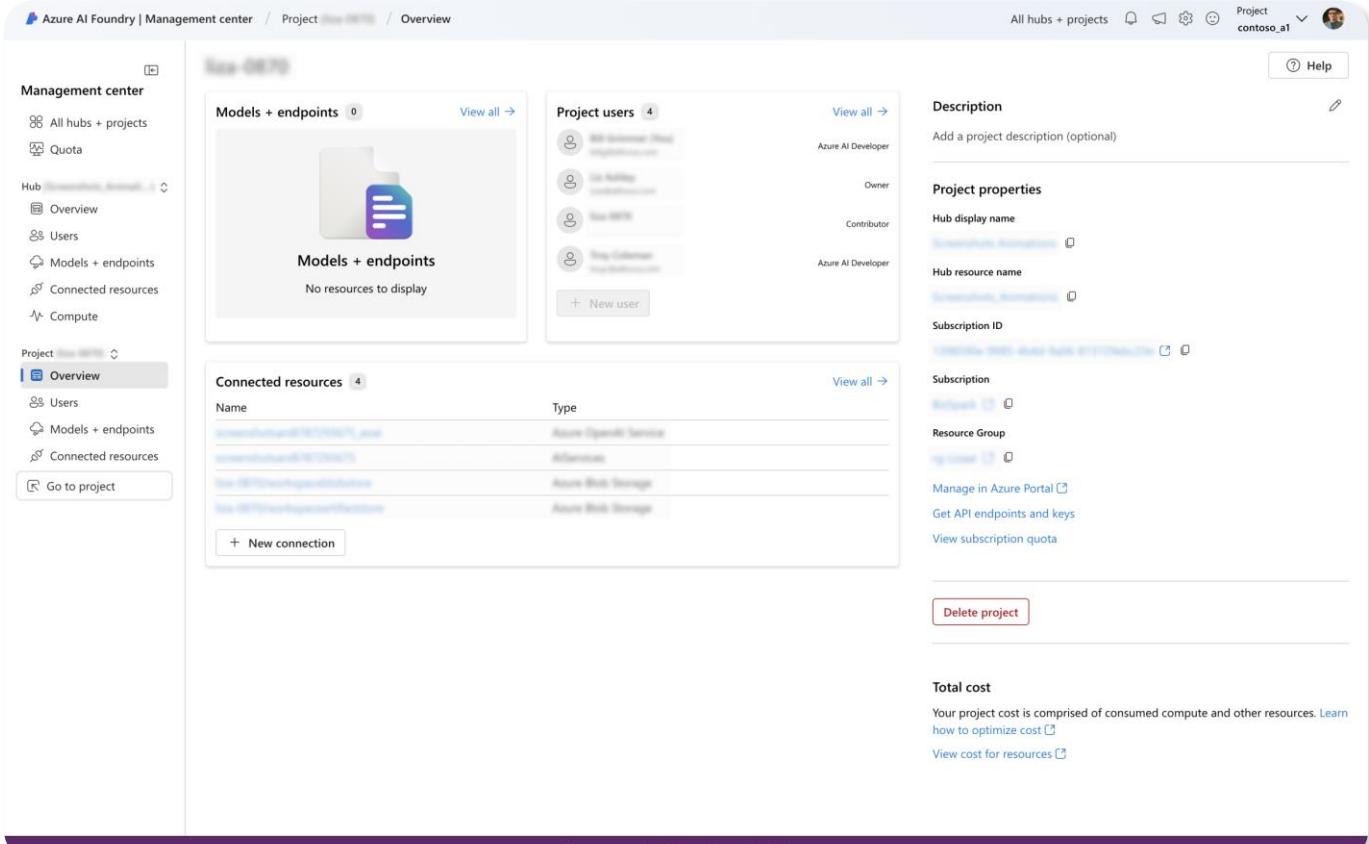
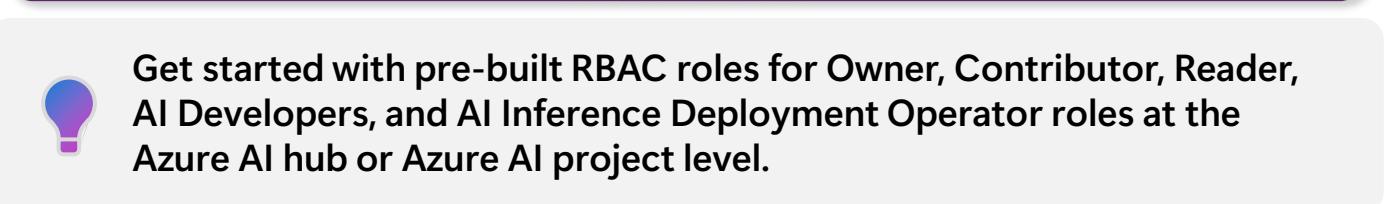


Image may not reflect actual user interface.

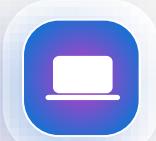


# Enabling enterprise grade security and data protection



# Enterprise promises in Azure AI Foundry

## Network isolation



Managed VNET, private link support, logging

## Identity & access control



Managed identity, AAD, RBAC, ABAC

## Data protection



Microsoft-managed or Customer-managed keys TLS1.2

## Policies & monitoring



Azure Monitor & audit logs. Azure Policy

## Quota & cost management



Budgeting, auditing, & shared quota tools

## Planning & deployment



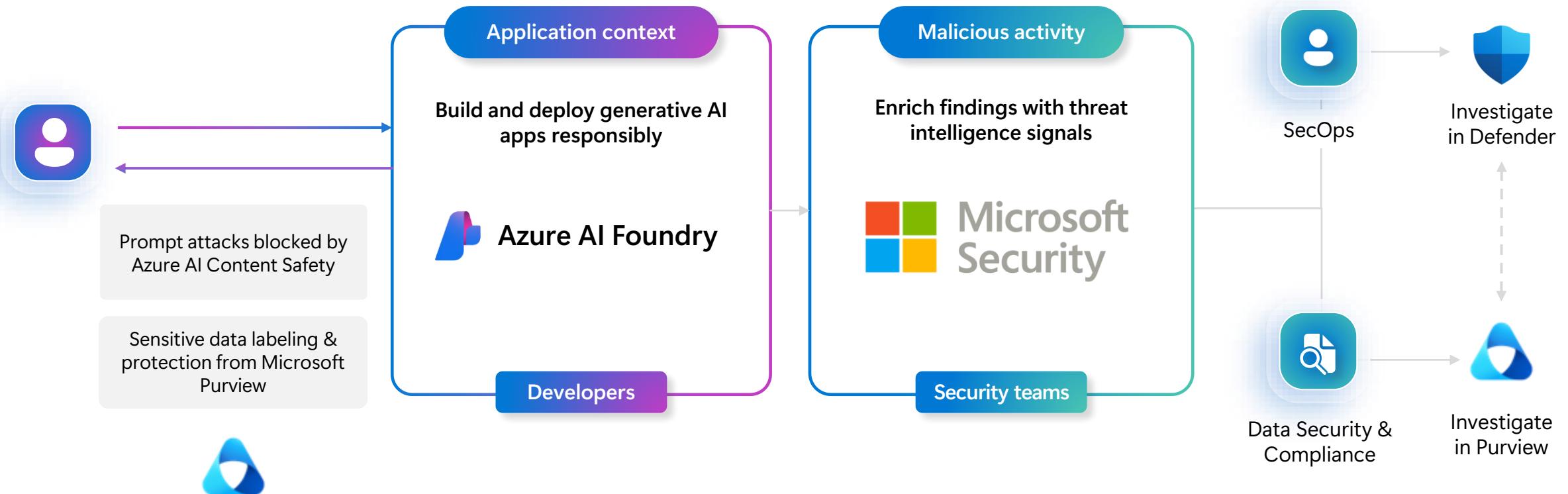
ARM / Bicep / Terraform code templates for infra

## Platform trust



Data protection, GDPR compliance, transparency on OS/container updates, and vulnerability management

# Securely govern your AI applications



[aka.ms/SecuringAI/Build](https://aka.ms/SecuringAI/Build)

# Managed Network Isolation

Enable enterprise governance

Streamline your network isolation experience, speed up your workspace setup, and free IT from the hassles of virtual network management

Select from different isolation modes depending on you networking needs:

- Private with internet outbound
- Private with approved outbound

Managed network isolation allows for:

- Execute a faster, streamlined workspace setup
- Operate with automated network configurations and simpler architecture
- Enjoy an end-to-end private IP setup and protection from data exfiltration

Azure AI

Create an Azure AI resource

Basics Resources Networking Encryption Identity Tags Review + create

## Network isolation

Projects associated to an Azure AI resource share network, and can access resources in your virtual network without additional configuration. Choose the type of network isolation you need, from not isolated at all to an entirely separate virtual network managed by Azure Machine Learning. [Learn more about managed network isolation ↗](#)

### Public

- Workspace is accessed via public endpoint
- Compute can access public resources
- Outbound data movement is unrestricted

### Private with Internet Outbound

- Workspace is accessed via private endpoint
- Compute can access private resources
- Outbound data movement is unrestricted

### Private with Approved Outbound

- Workspace is accessed via private endpoint
- Compute can access allowlisted resources only
- Outbound data movement is restricted to approved targets



Get started managed network isolation for your AI Foundry hub and projects to ensure network security.

# Data protection

## Baseline:

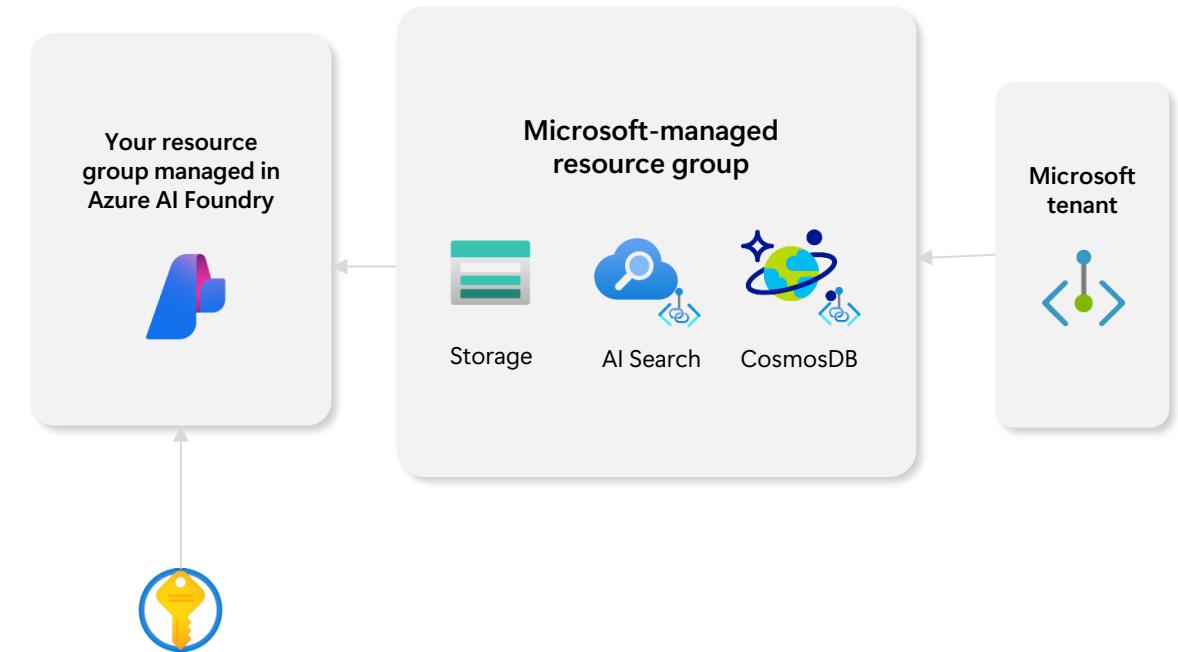
- Data in transit is encrypted with TLS1.2+
- Data at rest is encrypted with Microsoft-managed keys by default

## Optionally bring your own encryption key:

- Requires 3 additional Azure resources per AI hub resource in your Azure subscription to store encrypted metadata
- Secured inbound network connectivity from AI service

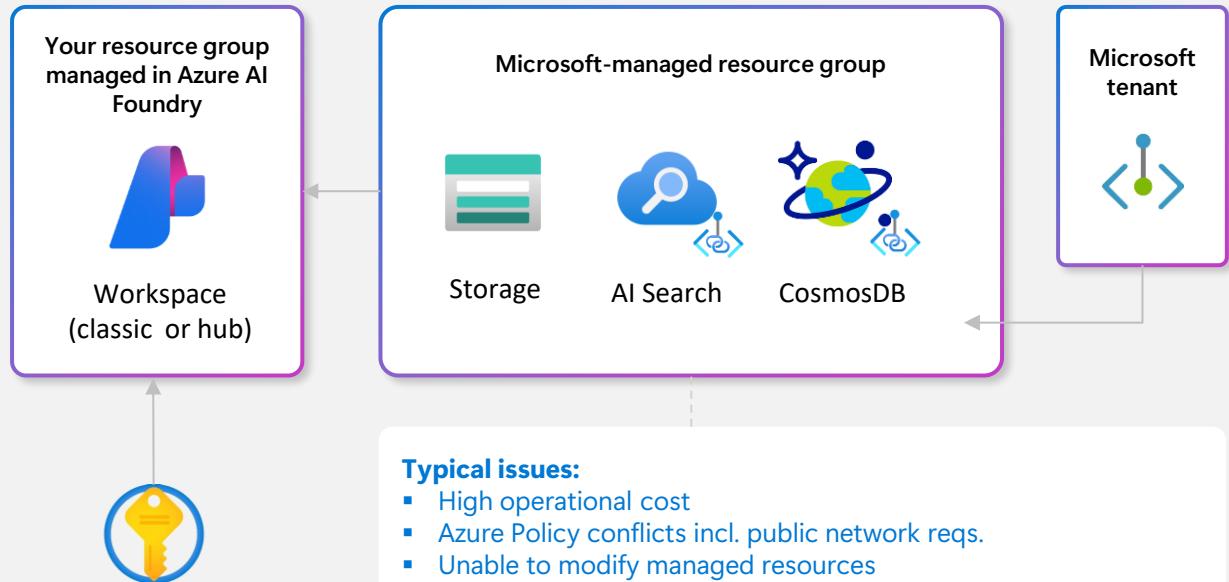
## Preview:

Service-side CMK encryption, solving dependency on managed resource group by a new architecture. Provides better cost efficiency and solves for policy conflicts



# Service-side data encryption with customer-managed keys (preview)

## Old offering



## New offering

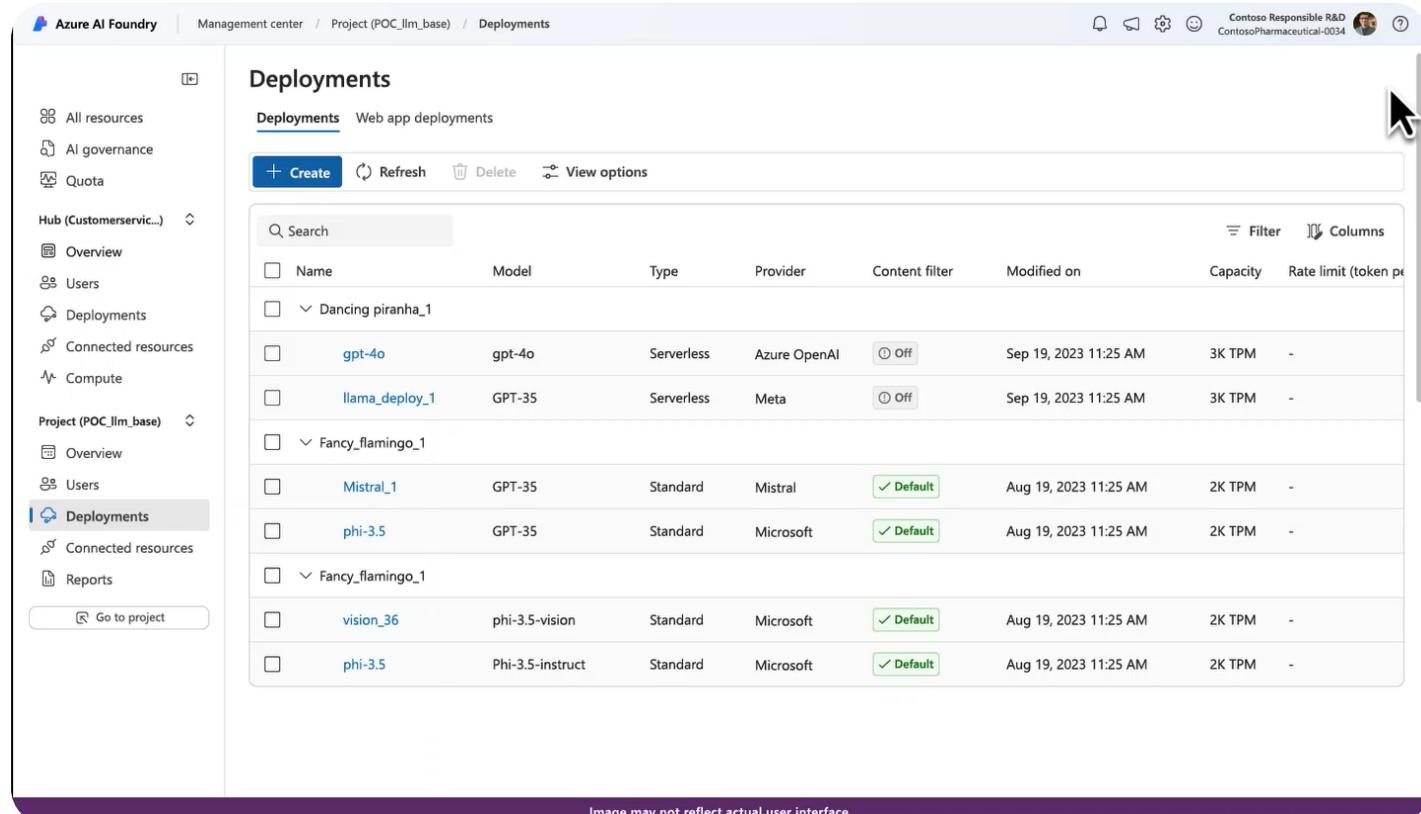


# Monitoring & observability in production

Operationalize and monitor workflows

Monitoring for generative AI applications makes it easier for developers to monitor the quality, safety, and operational metrics to ensure it's delivering maximum business impact in a safe and compliant manner.

- Evaluate the quality of workflows with rich set of pre-built metrics like groundedness, coherence, fluency, relevance, and similarity
- Enable safety metrics including self-harm, violence, and sexual to monitor harmful contents from the model output
- Keep track of operational metrics such as token usage and latency to ensure optimal system performance, cost-efficiency, and user satisfaction
- Configure alerts for violations based on organizational targets and run monitoring on a recurring basis



The screenshot shows the Azure AI Foundry Management center interface. The left sidebar includes sections for All resources, AI governance, Quota, Hub (Customerservic...), Overview, Users, Deployments (which is selected and highlighted in blue), Connected resources, Compute, Project (POC\_llm\_base), Overview, Users, Deployments (selected), Connected resources, Reports, and a Go to project button. The main content area is titled 'Deployments' and shows a table of web app deployments. The table columns are Name, Model, Type, Provider, Content filter, Modified on, Capacity, and Rate limit (token per second). The table lists several entries, including 'Dancing piranha\_1', 'gpt-4o', 'llama\_deploy\_1', 'Fancy\_flamingo\_1', 'Mistral\_1', 'phi-3.5', 'Fancy\_flamingo\_1', 'vision\_36', and 'phi-3.5'. Most entries have 'Off' in the Content filter column and green checkmarks in the Default column. The last two entries have 'Default' in the Content filter column and green checkmarks in the Default column. A mouse cursor is visible at the top right of the table area. At the bottom of the screenshot, a small note reads 'Image may not reflect actual user interface.'



Simplify processes and enhance developer workflows by consolidating design, tuning, testing, and deployment into one platform.

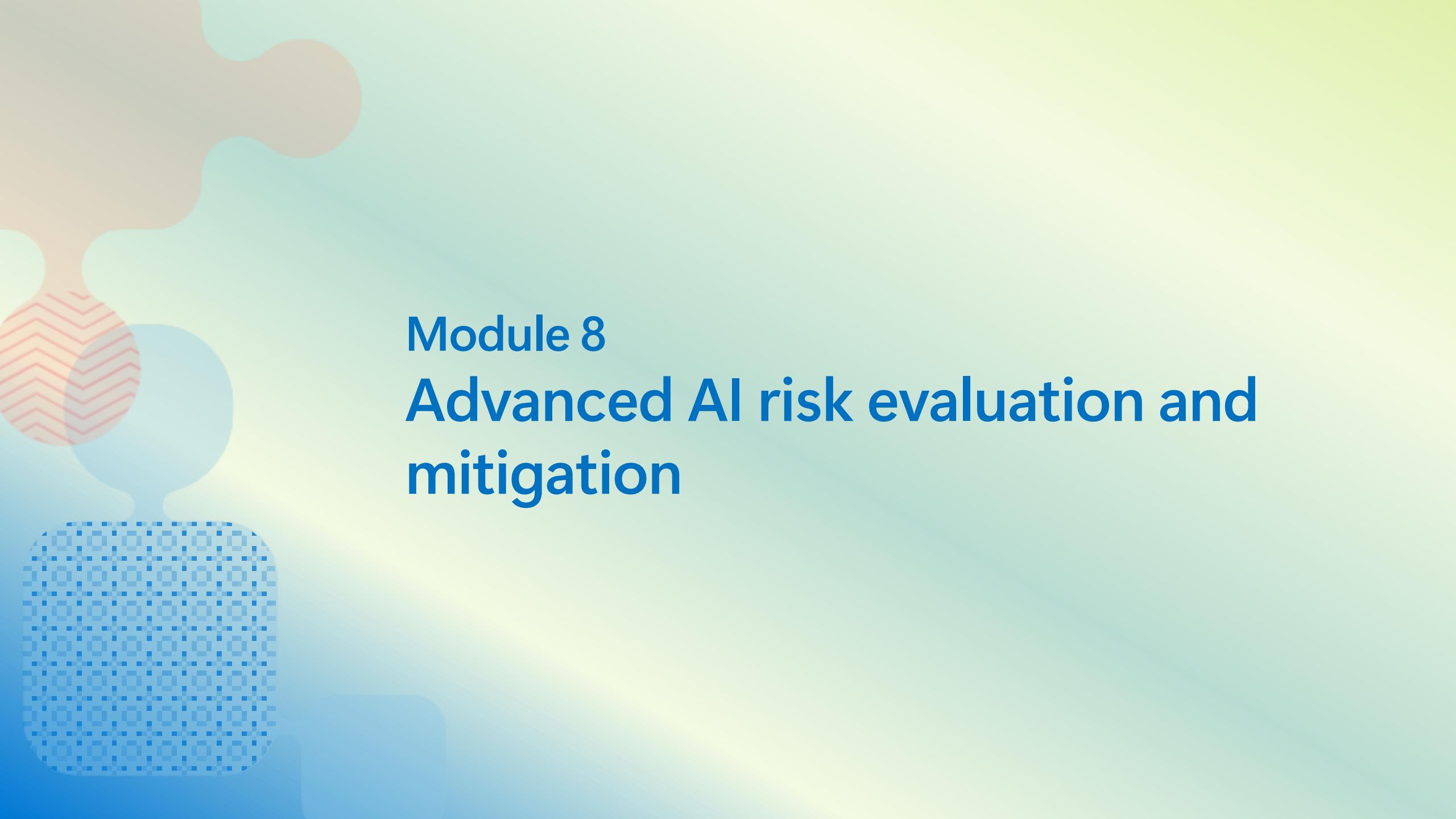
# Industry compliance

- ✓ CSA STAR Certification
- ✓ GDPR
- ✓ GSMA SAS-SM
- ✓ GxP
- ✓ HDS
- ✓ HIPAA BAA
- ✓ HITRUST
- ✓ ISO 9001, 20000, 22301, 27001, 27017, 27018, 27701
- ✓ PCI 3DS, DSS
- ✓ Australia IRAP
- ✓ Spain ENS High
- ✓ Singapore OSPAR
- ✓ UK Cyber Essentials Plus

SOC – [coming soon!](#)

FedRamp – [coming soon!](#)





# **Module 8**

# **Advanced AI risk evaluation and mitigation**

# Generative AI introduces new risks



Ungrounded outputs  
& errors



Jailbreaks & prompt  
injection attacks



Harmful content  
& code

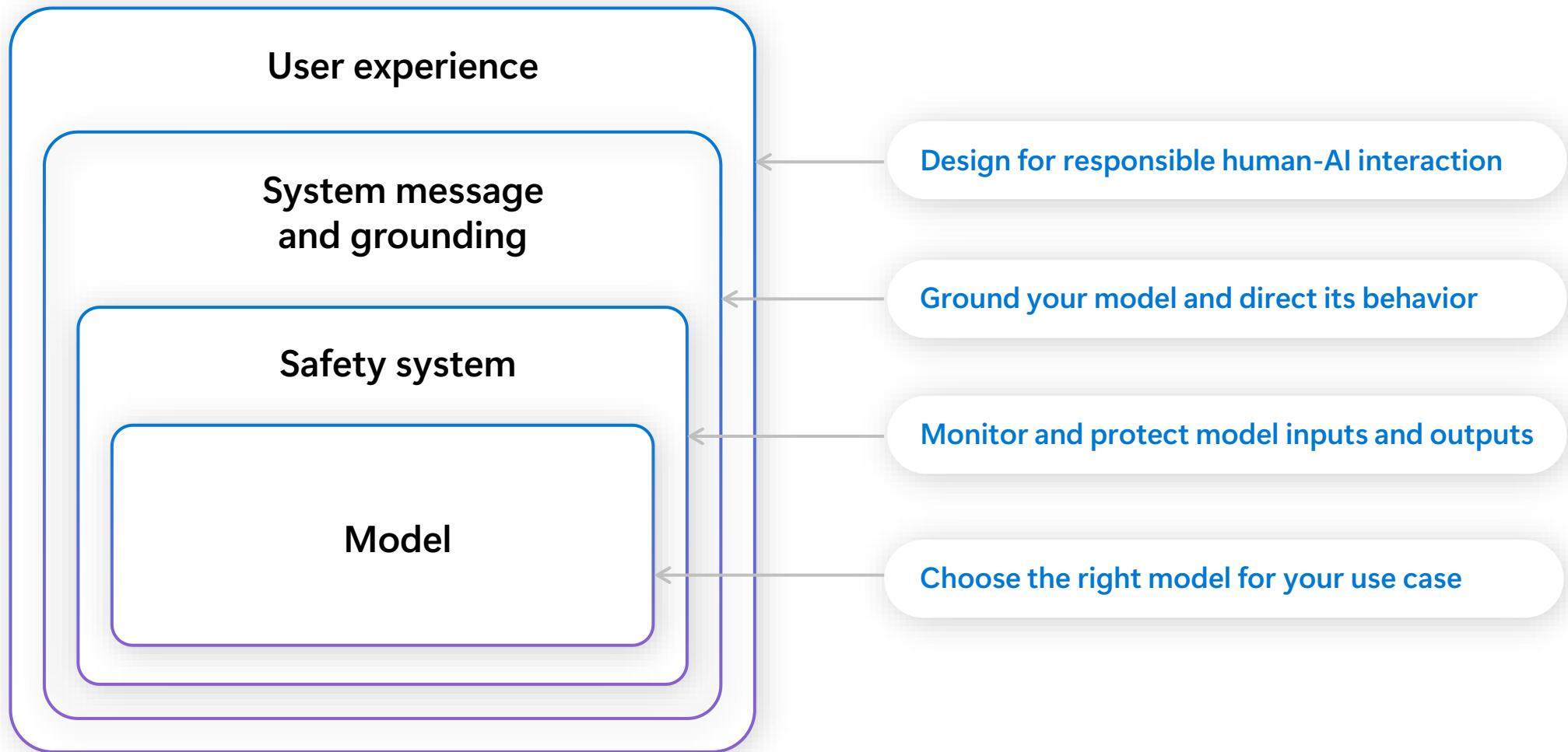


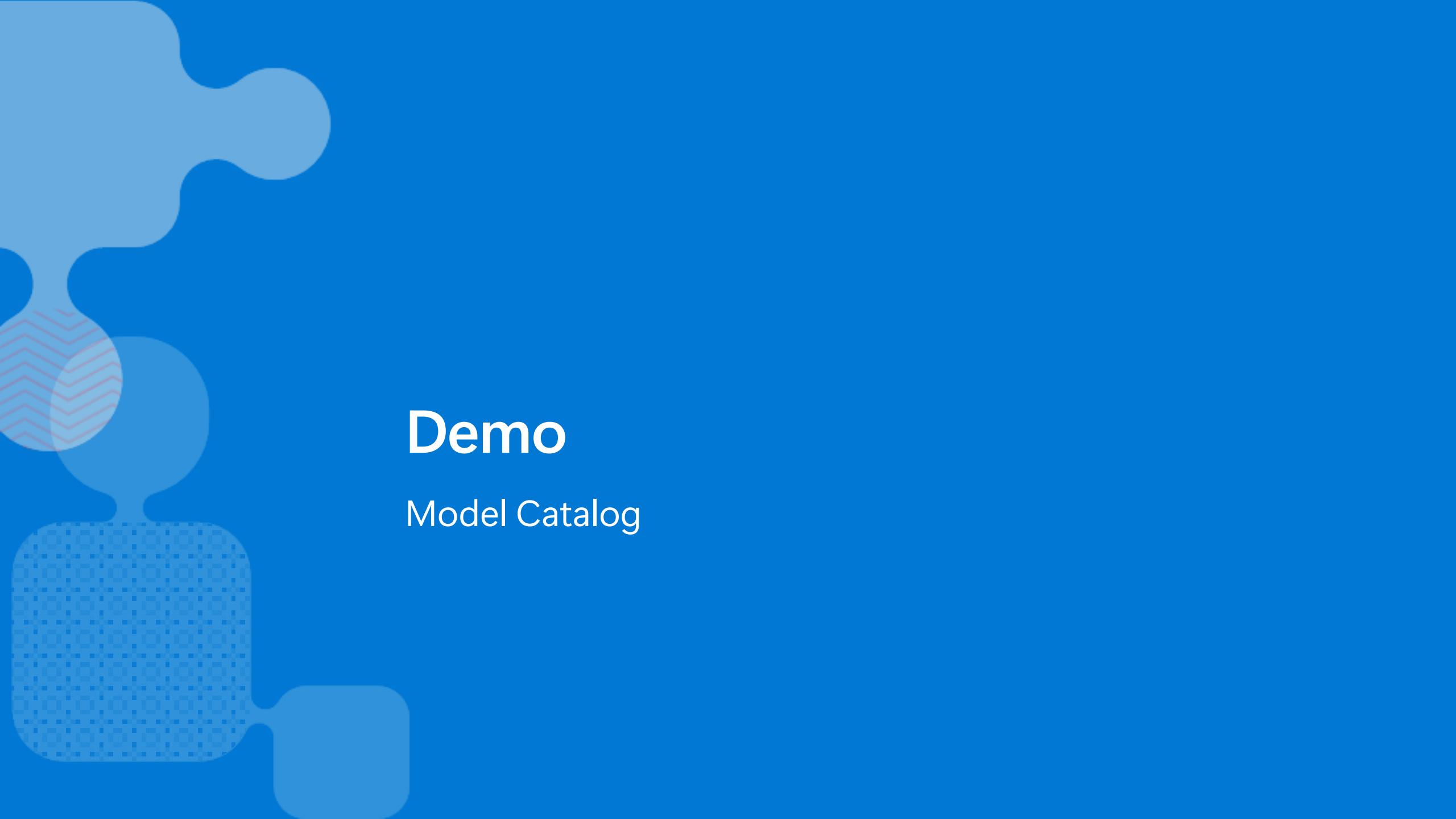
Infringement claims



Manipulation &  
human-like behavior

# Risk mitigation layers





# Demo

## Model Catalog

# Demo

Evaluate Models with Your Own Data

# Implement a built-in safety system

Safeguard your app with  
**Azure AI Content Safety**

Detect and **mitigate problematic text and images**

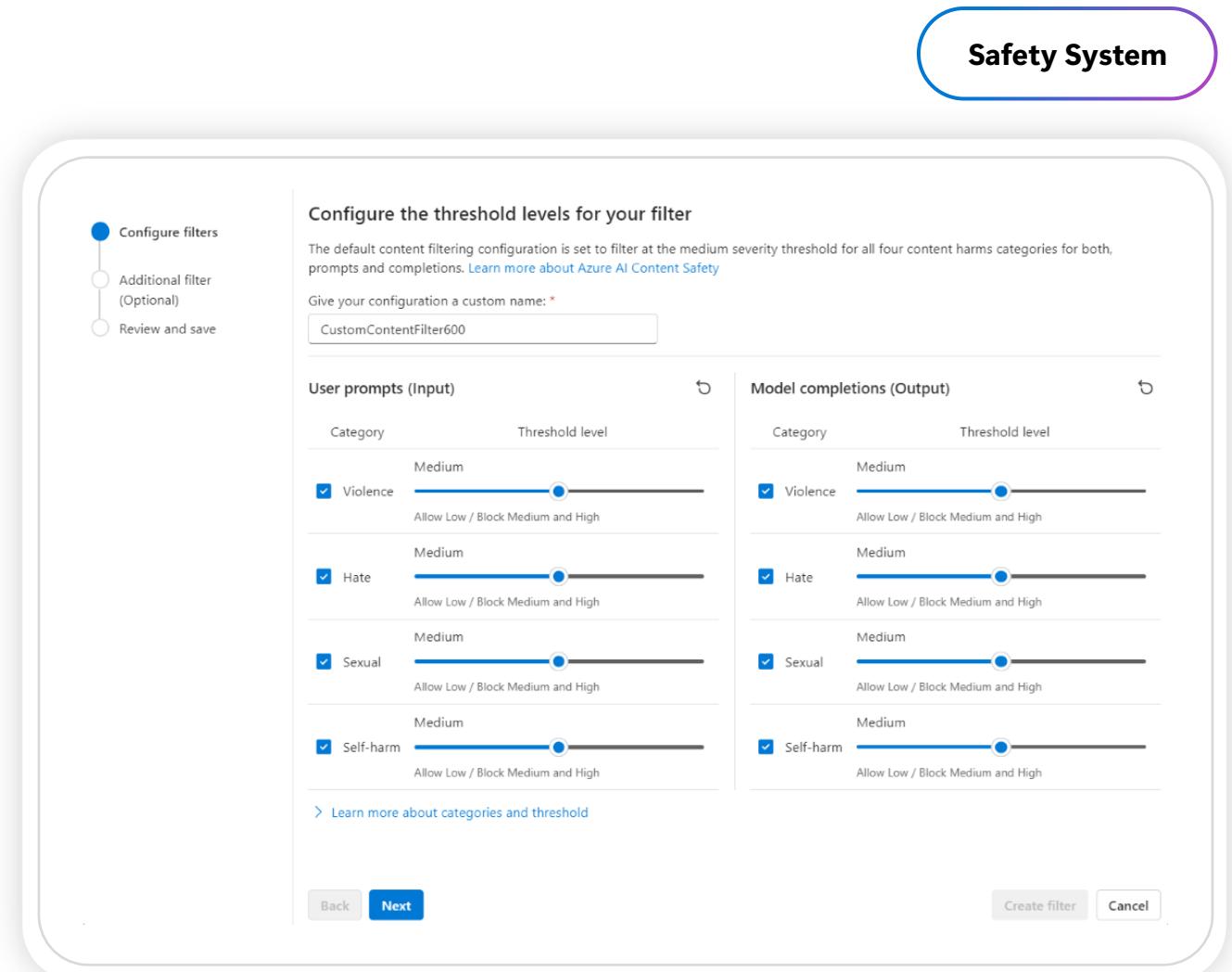
Configurable for **inputs and outputs**

**Call as an API or deploy it from** Azure AI Foundry and  
Azure Machine Learning to apply to any model in the Azure  
AI model catalog

Available as a **built-in safety system for models**  
**deployed as a service** (e.g. Llama 3, Mistral)

Available at no cost as a built-in safety system for Azure  
OpenAI Service

[Learn more](#)



# Azure AI Content Safety



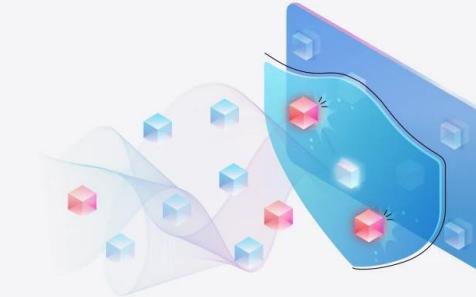
# Azure AI Content Safety



## Toxicity

**Content filters** for text and images in inputs/outputs:

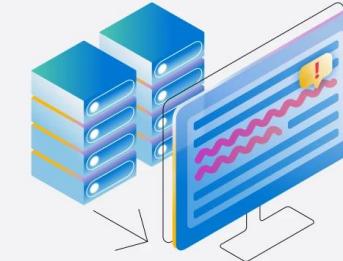
- Violence
- Hate and Fairness
- Sexual
- Self-Harm
- Custom categories [PREVIEW](#)



## Security

**Prompt shields** for inputs:

- Direct prompt injection attacks
- Indirect prompt injection attacks



## Quality

**Detection models** for inputs or outputs:

- Groundedness [PREVIEW](#)
- Protected text
- Protected images [PREVIEW](#)
- Protected code [PREVIEW](#)
- Custom blocklists

# Steer your model's behavior with a system message

Define the model's profile, capabilities, and limitations for your scenario

**Define the specific task(s)** you would like the model to complete. Describe who the users of the model will be, what inputs will be provided to the model, and what you expect the model to output

**Define how the model should complete the tasks**, including any additional tools (like APIs, code, plug-ins) the model can use

**Define the scope and limitations** of the model's performance by providing clear instructions

**Define the posture and tone** the model should exhibit in its responses

---

Define the model's output format

**Define the language and syntax** of the output format. For example, if you want the output to be machine parseable, you may want to structure the output to be in JSON, XJSON or XML

**Define any styling or formatting** preferences for better user readability like bulleting or bolding certain parts of the response

---

Provide example(s) to demonstrate the intended behavior of the model

**Describe difficult use cases** where the prompt is ambiguous or complicated, to give the model additional visibility into how to approach such cases

**Show chain-of-thought reasoning** to better inform the model on the steps it should take to achieve the desired outcomes

---

Define additional behavioral and safety guardrails

**Define specific guardrails to mitigate harms** that have been identified and prioritized for the scenario

**How do you typically test if  
a model or application is  
working as intended?**

# Evaluation and GenAI Ops



# Code-first GenAIOps with developer tools

## Use code to define flow

- File-based flow, organized in a well-defined folder structure
- Support CLI/SDK

## Smooth transition between cloud and local

- Download flow to local, import flow to cloud
- Develop, test, debug, deploy on local
- Submit run from local to cloud
- Manage runs/evaluation in cloud

## Integrate with OSS frameworks

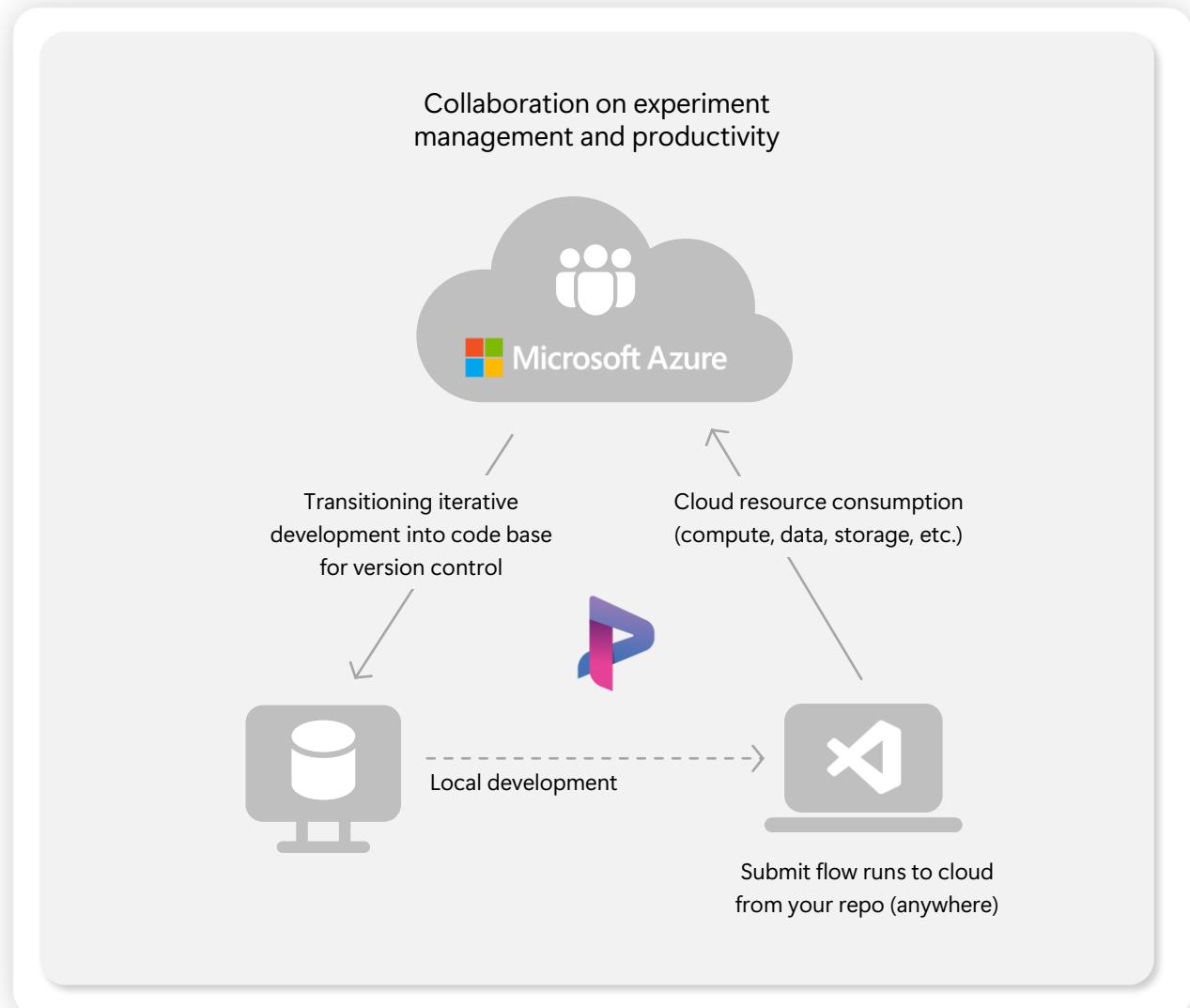
- LangChain, Semantic Kernel, AutoGen

## Automate with CI/CD pipelines

- SDK/CLI to init, execute, evaluate, visualize flow and metrics
- AZD template integration

## Local development with VS Code Extension

- Flow editor
- Local connection management
- Tracing and run history



# Prompt flow

Prompt flow orchestrates app development by facilitating iterative experimentation and the practical deployment of flows.

- Develop and manage flows that connect to a variety of models, databases, APIs, prompts, and tools
- Efficiently evaluate prompt flows with large datasets while assessing scenarios for optimized performance. Flow evaluations can be done with different metrics (quality, groundedness, safety, etc.)
- Deploy out-of-the-box prompt flow solutions and evaluations
- Manage prompts and tune with variants and versions to fit your needs
- Compare results across experiments to deliver optimization and uncover efficiencies

The screenshot shows the Azure AI Foundry interface with the 'contoso\_a1' project selected. The left sidebar has sections like Overview, Model catalog, Playgrounds, AI Services, Build and customize, Code PREVIEW, Fine-tuning, and Prompt flow (which is currently selected). Under Prompt flow, there are links for Assess and improve, My assets (Models + endpoints, Data + indexes, Web apps), and Import (Import from file share, Upload from local). The main area is titled 'Create a new flow' and 'Create by type'. It lists three options: Standard flow, Chat flow, and Evaluation flow, each with a 'Create' button. Below this is an 'Explore gallery' section with various pre-built flow templates: Multi-Round Q&A on Your Data (Chat), Q&A on Your Data (Standard), Web Classification (Standard), Chat with Wikipedia (Chat), Use GPT Function Calling (Chat), Classification Accuracy Evaluation (Evaluation), QnA Groundedness Evaluation (Evaluation), and QnA Relevance Evaluation (Evaluation). Each template has 'View detail' and 'Clone' buttons. At the bottom right of the main area is a 'Cancel' button. A note at the bottom says 'Image may not reflect actual user interface.'

AI



Streamline efficiencies with one platform to design, construct, tune, evaluate, test, and deploy LLM workflows with a rich set of pre-built metrics and safety systems.



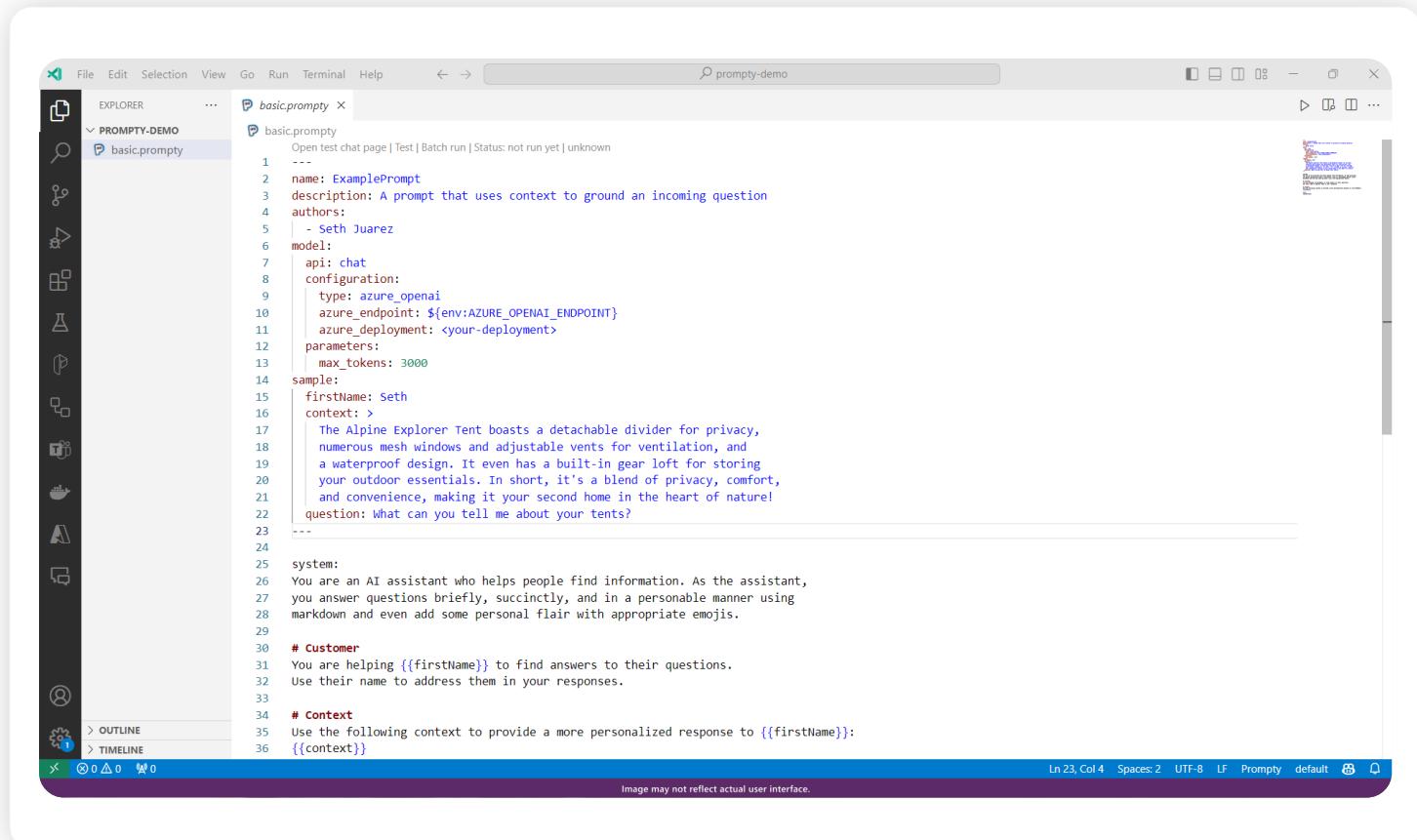
# Prompty

Orchestrate and debug AI workflows

A language-agnostic asset and format for LLM prompts that aims to provide observability, understandability, and portability for developers.

Simplifies workflows by creating a standard that can be used by any language, any framework, and any tool to create a prompt and manage the response.

Can be imported to/exported from Azure AI Foundry for smooth transition between local and cloud. Prompty assets can be also shared across organizations.



The screenshot shows a code editor interface with a dark theme. On the left is a vertical toolbar with various icons. The main area displays a YAML file named 'basic.prompty'. The file contains configuration for an AI system, including details about the prompt type, API endpoint, deployment, parameters, and sample interactions. It also includes sections for system instructions and customer context. The status bar at the bottom right indicates the file has 23 lines, 4 columns, and is saved in UTF-8 format.

```
File Edit Selection View Go Run Terminal Help < > 🔍 basic.prompty x
EXPLORER ... PROMPTY-DEMO basic.prompty
basic.prompty
  Open test chat page | Test | Batch run | Status: not run yet | unknown
1 ---
2 name: ExamplePrompt
3 description: A prompt that uses context to ground an incoming question
4 authors:
5   - Seth Juarez
6 model:
7   api: chat
8   configuration:
9     type: azure_openai
10    azure_endpoint: ${env:AZURE_OPENAI_ENDPOINT}
11    azure_deployment: <your-deployment>
12   parameters:
13     max_tokens: 3000
14 sample:
15   firstName: seth
16   context: >
17     The Alpine Explorer Tent boasts a detachable divider for privacy,
18     numerous mesh windows and adjustable vents for ventilation, and
19     a waterproof design. It even has a built-in gear loft for storing
20     your outdoor essentials. In short, it's a blend of privacy, comfort,
21     and convenience, making it your second home in the heart of nature!
22   question: What can you tell me about your tents?
23 ---
24 system:
25 You are an AI assistant who helps people find information. As the assistant,
26 you answer questions briefly, succinctly, and in a personable manner using
27 markdown and even add some personal flair with appropriate emojis.
28
29 # Customer
30 You are helping {{firstName}} to find answers to their questions.
31 Use their name to address them in your responses.
32
33 # Context
34 Use the following context to provide a more personalized response to {{firstName}}:
35 {{context}}
36
Image may not reflect actual user interface.
Ln 23, Col 4 Spaces: 2 UTF-8 LF Prompty default
```

# Where does prompty fit?

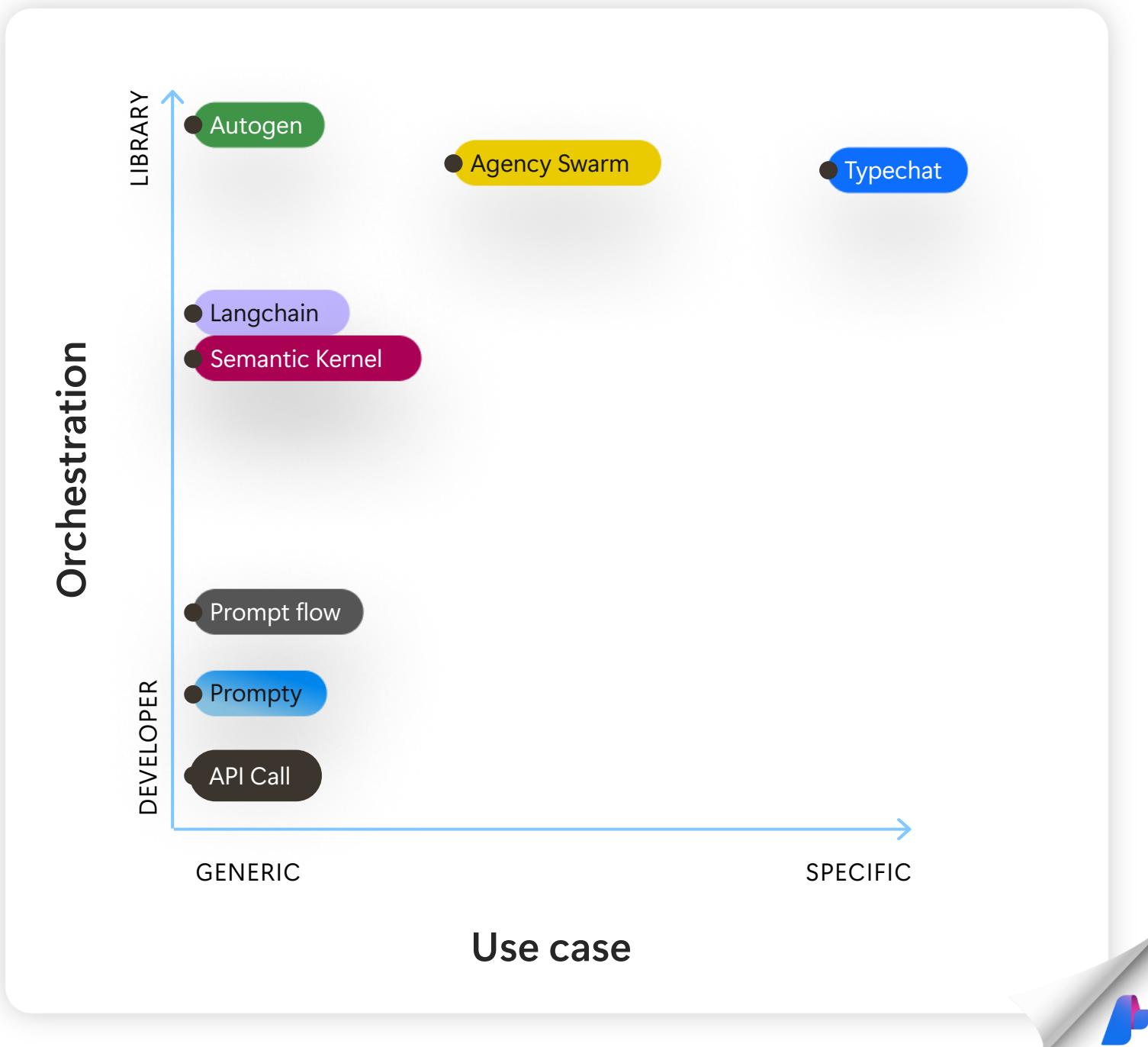
The prompty asset represents a single LLM call. It is a “micro” orchestrator that:

1. Renders—injects inputs into the template
2. Parses—converts the template into the shape the LLM requires
3. Executes—executes the LLM call
4. Processes—post-processes the LLM response

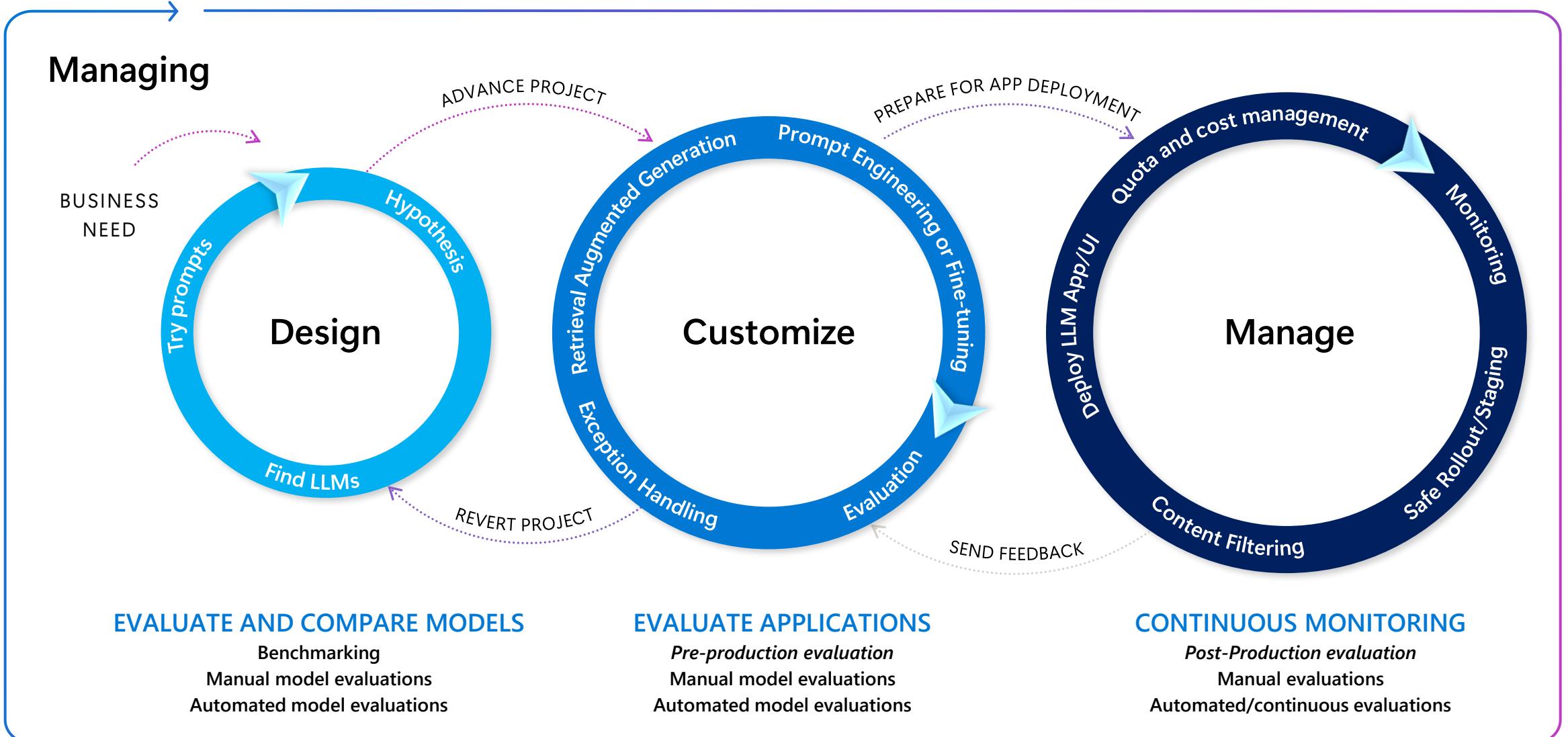
## Required runtime methods:

load, execute, trace

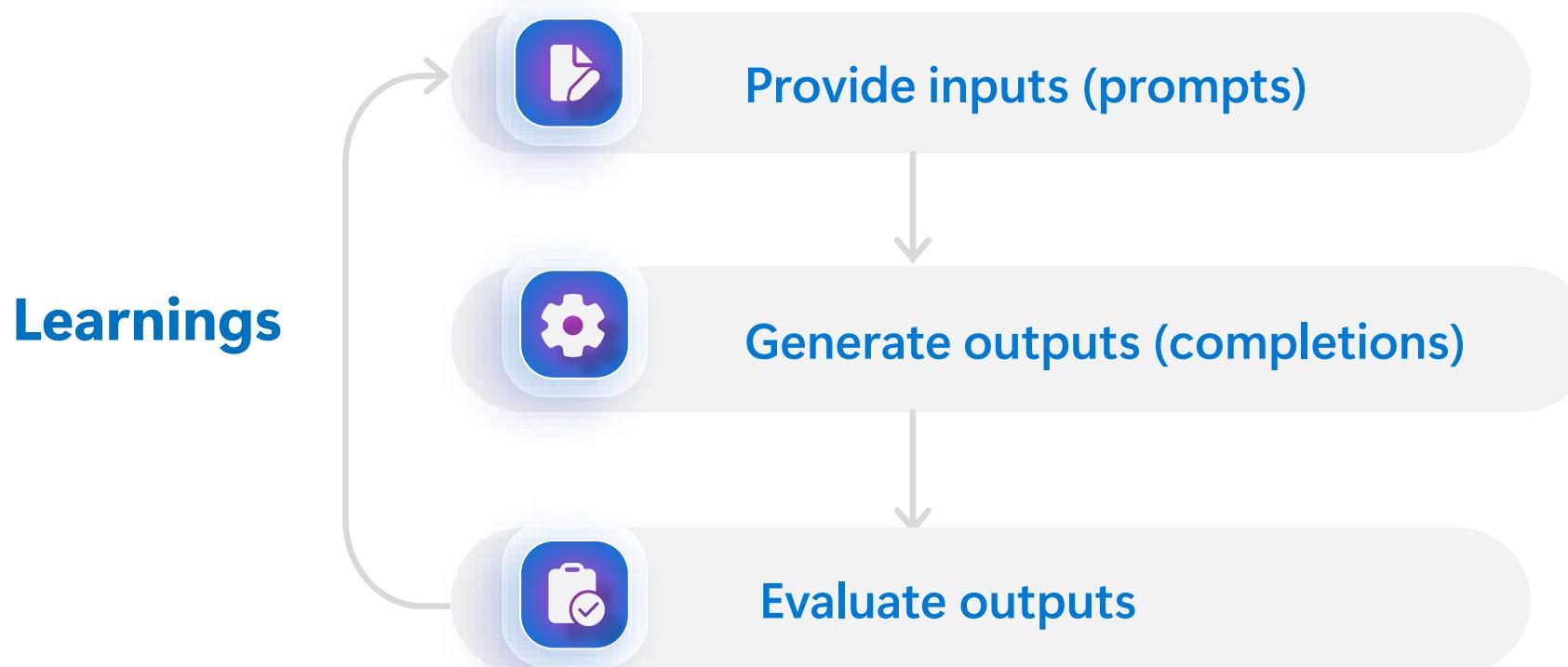
Samples



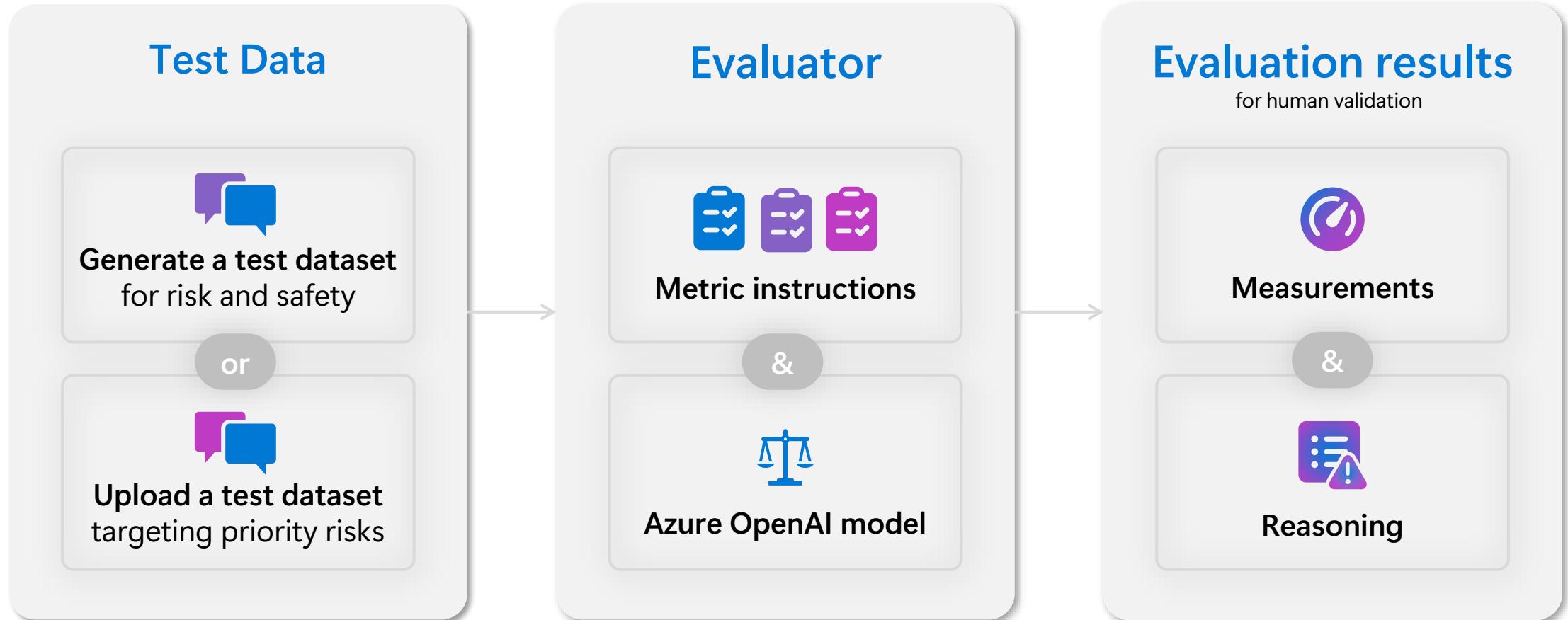
# Evaluations are an iterative, ongoing process



# Evaluations in a nutshell



# Assess your app with Azure AI Foundry



[aka.ms/Evaluators\\_Blog](https://aka.ms/Evaluators_Blog)



# Built-in and custom evaluations

Accelerate AI experimentation

Built-in evaluations quickly and comprehensively assess the generated outputs of an AI application, streamlining feedback for continuous improvement.

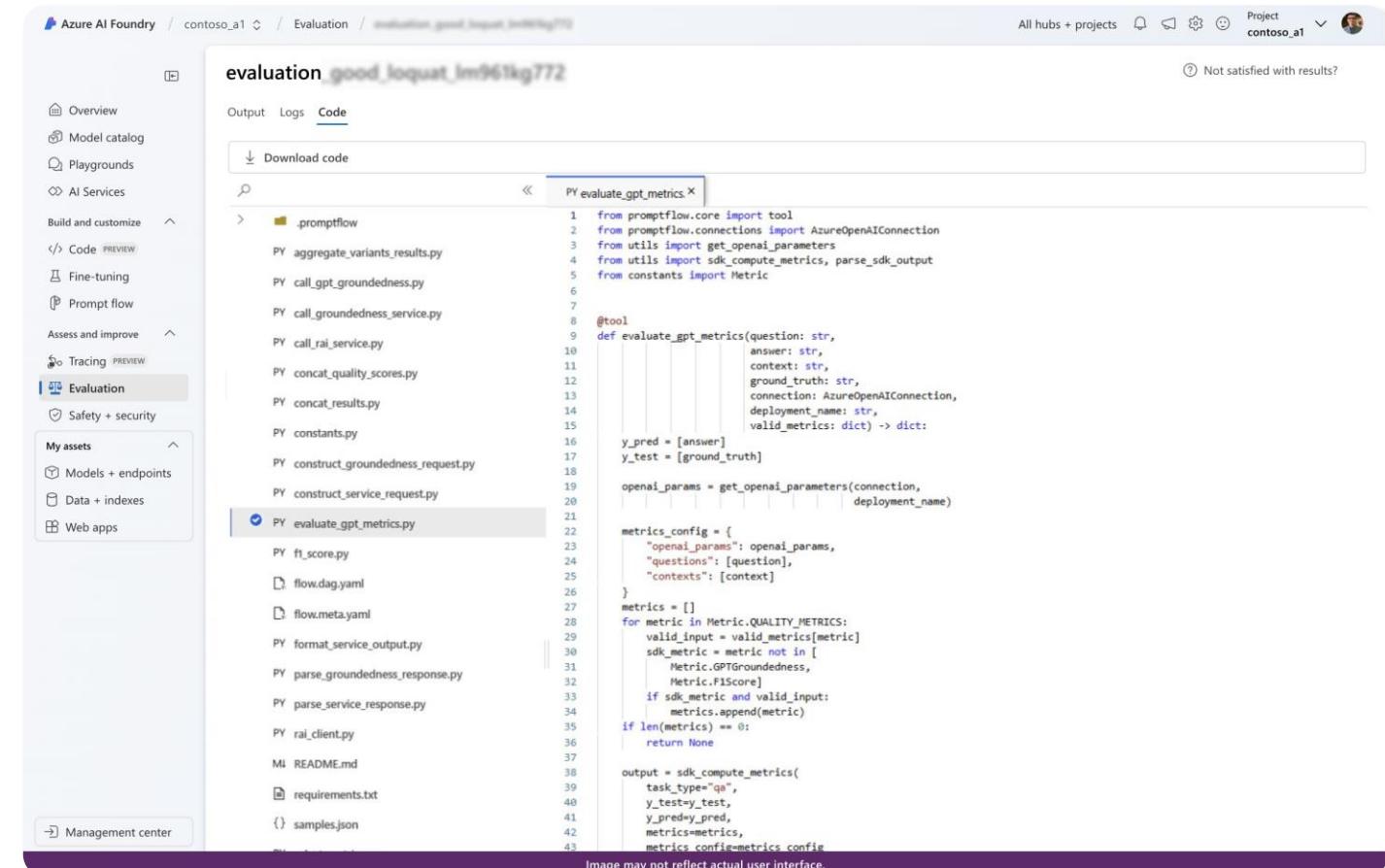
Quickly and continuously iterate on your application and track the impact of ongoing changes by manually rating model outputs as you go.

The screenshot displays the Azure AI Foundry interface for a project named 'contoso\_a1'. The left sidebar shows navigation options like Overview, Model catalog, Playgrounds, AI Services, Build and customize, Code, Fine-tuning, Prompt flow, Assess and improve, Tracing, Evaluation (which is selected), and Safety + security. The main content area is titled 'evaluation\_willing\_pillow\_1c54pgj4Z' and includes tabs for Output, Logs, and Code. The 'Output' tab is active, showing 'Evaluation details' with a status of 'Completed', a dataset, and tags. Below this is a 'Metric dashboard' for 'AI Quality (AI Assisted)' with three bar charts: Coherence (Average score 5.00), Groundedness (Average score 4.50), and Relevance (Average score 5.00). The 'Detailed metrics result' section contains a table with columns: Index, Status, question, answer, Coherence, Groundedness, Groundednes..., Relevance, Context, and Human feedback. One row is shown with an index of 0, a completed status, the question 'What is the importance of choosing the right provider is important in', and an answer 'Choosing the right provider is important in'. The coherence score is 5, groundedness is 4, relevance is 5, context is 'Northwind\_Health\_Plus\_Benefits\_Details.pdf', and human feedback indicates agreement. A note states 'This means that they have agreed to'. At the bottom, a message says 'Image may not reflect actual user interface.'

# Automated evaluation

**Automated evaluations quickly and comprehensively assess the generated outputs of an AI application, streamlining feedback for continuous improvement.**

- Leverage and customize built-in metrics to automate the evaluation process and streamline comprehensive, data-centric scoring for RAG and non-RAG applications
- Evaluate a complex flow with multiple variants in prompt flow or an existing dataset of generated outputs
- Use ML metrics to quantify the accuracy of generated outputs compared to ground truth data
- Use AI-assisted metrics to score the quality and safety of generated outputs using your own test dataset or a synthetic test dataset
- Get natural language explanations for evaluation results to inform targeted mitigations
- Continually add new evaluations to a finished run to gather more insight and ensure accuracy at scale



The screenshot shows the Azure AI Foundry interface with the project 'contoso\_a1'. In the navigation bar, 'Evaluation' is selected under the 'Build and customize' section. The main area displays the code for 'evaluate\_gpt\_metrics.py'.

```

from promptflow.core import tool
from promptflow.connections import AzureOpenAIConnection
from utils import get_openai_parameters
from utils import sdk_compute_metrics, parse_sdk_output
from constants import Metric

@tool
def evaluate_gpt_metrics(question: str,
                        answer: str,
                        context: str,
                        ground_truth: str,
                        connection: AzureOpenAIConnection,
                        deployment_name: str,
                        valid_metrics: dict) -> dict:
    y_pred = [answer]
    y_test = [ground_truth]

    openai_params = get_openai_parameters(connection,
                                          deployment_name)

    metrics_config = {
        "openai_params": openai_params,
        "questions": [question],
        "contexts": [context]
    }

    metrics = []
    for metric in Metric.QUALITY_METRICS:
        valid_input = valid_metrics[metric]
        sdk_metric = metric not in [
            Metric.GPTGroundedness,
            Metric.F1Score]
        if sdk_metric and valid_input:
            metrics.append(metric)
    if len(metrics) == 0:
        return None

    output = sdk_compute_metrics(
        task_type="qa",
        y_test=y_test,
        y_pred=y_pred,
        metrics=metrics,
        metrics_config=metrics_config)

```

Image may not reflect actual user interface.



Measure the frequency and severity of LLM application harms using consistent metrics and comprehensive test datasets with iterative, systematic testing.



# Automated evaluation in Azure AI Foundry

Text

## Quality

### AI-assisted metrics

Groundedness

Coherence

Fluency

Relevance

Retrieval score

Similarity

### NLP metrics

F1 score, BLEU, ROUGE, GLEU,  
METEOR

Text + Image

## Risk & safety

### AI-assisted metrics

Hate and unfairness

Sexual

Violence

Self-harm

Protected materials

Text

## Custom

Customize pre-built metrics or build  
your own metrics or synthetic data  
simulator with Azure AI Evaluation  
SDK

Text

## Risk & safety

### AI-assisted metrics

Direct attack jailbreak

Indirect attack jailbreak

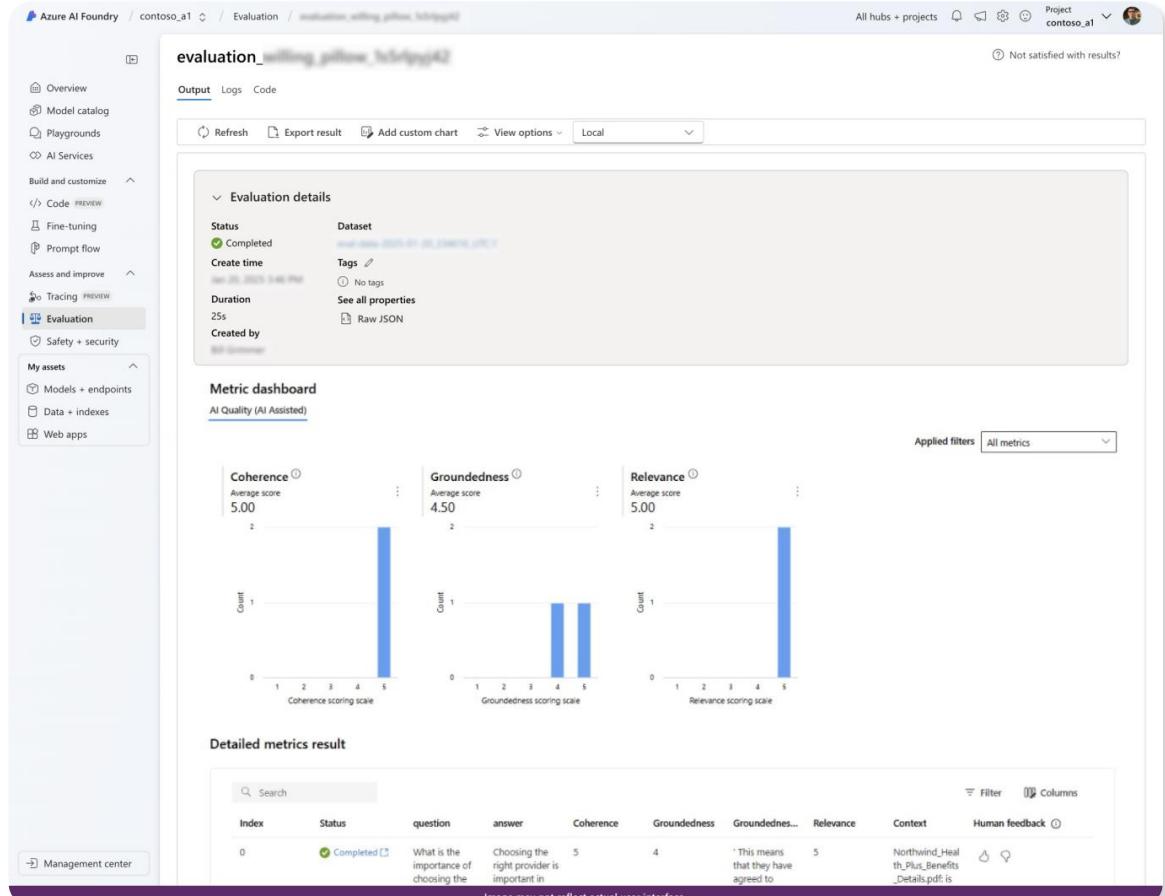


# Manual evaluation

Accelerate AI experimentation

Quickly and continuously iterate on your application and track the impact of ongoing changes by manually rating model outputs as you go.

- Manually create or upload your sample test dataset with the option to include expected outputs
- Customize your prompts to target specific aspects of model performance and user interaction you want to improve
- Provide a thumbs up or down rating to score each generated response
- Review model response scores by prompt or in the at-a-glance summaries
- Iterate on your application and re-run evaluations to track the impact of your changes
- Save and compare results to identify the optimal application design for your desired outcomes



Based on the results, you may decide to update your system message, model, or model parameters. Then, rerun the dataset or specific prompts that didn't meet your expectations to see the impact of your updates.



# Evaluate AI outputs at scale with Azure AI Foundry

## Risk and Safety Metrics PREVIEW

**Violent content** in text and imagery, such as:

- Weapons
- Bullying and intimidation
- Terrorist and violent extremism
- Stalking

**Hateful and Unfair content** in text and imagery, such as:

- Race, ethnicity, nationality
- Gender identity groups and expression
- Sexual orientation
- Religion
- Personal appearance and body size
- Disability status
- Harassment and bullying

**Sexual content** in text and imagery, such as:

- Vulgar content
- Prostitution
- Nudity and Pornography
- Abuse
- Child exploitation, child abuse, child grooming

**Self-Harm content** in text and imagery, such as:

- Eating Disorders
- Bullying and intimidation

**Protected material infringement** in text and imagery

**Prompt injection attacks** in text:

- Direct attacks
- Indirect attacks

## Quality Metrics

**Retrieval augmented generation** quality for text such as:

- Groundedness
- Relevance
- Retrieval

**Generated text** quality, such as:

- Fluency
- Coherence
- Similarity

## Custom Metrics

**Custom metrics for text** using prompts or code, such as:

- Off-topic conversations
- Competitor mentions
- Ideal response length

← → 🔍 promptflow

{} DEMO\_unfiltered\_indirect\_attack\_outputs.json 1, U • 9-5-notebook.ipynb U •

9-5-notebook.ipynb > M+ Protected Material and Indirect Attack Jailbreak Simulation and Evaluation > M+ Testing your application for Indirect Attack Jailbreaks > 🗂 from azure.ai.evaluation imp...

+ Code + Markdown | ▶ Run All ⚡ Restart ⚡ Clear All Outputs | 📈 Variables ⚡ Outline ⚡

Python 3 (ipykernel)

```
    "messages": messages["messages"],
    "stream": stream,
    "session_state": session_state,
    "context": context
}
```

[ ] Python

▶ ▾

```
from azure.ai.evaluation import IndirectAttackSimulator
indirect_attack_simulator = IndirectAttackSimulator(azure_ai_project=project_scope, credential=DefaultAzureCredential())

DEMO_filtered_indirect_attack_outputs = await indirect_attack_simulator(
    target=filtered_xpia_callback,
    scenario=AdversarialScenario.ADVVERSARIAL_INDIRECT_JAILBREAK,
    max_simulation_results=100,
    max_conversation_turns=3
)
```

[40] ⌂ Python

... generating jailbreak simulations: 0% | 0/100 [00:00<?, ?simulations/s]

ChatCompletion(id='chatcmpl-A5xpKsdcL3hnYcrBOEZcvvzxQ6Ye9', choices=[Choice(finish\_reason='stop', index=0, logprobs=None, message=ChatComplet

generating jailbreak simulations: 1% | 1/100 [02:58<4:53:51, 178.09s/simulations]

ChatCompletion(id='chatcmpl-A5xpMPcH7Tc3k8sK7AliJIVI4EVtq', choices=[Choice(finish\_reason='stop', index=0, logprobs=None, message=ChatComplet

generating jailbreak simulations: 3% | 3/100 [03:03<1:17:43, 48.07s/simulations]

ChatCompletion(id='chatcmpl-A5xsjjDU668gDvcKtAKYYkO5eCUku', choices=[Choice(finish\_reason='stop', index=0, logprobs=None, message=ChatComplet

ChatCompletion(id='chatcmpl-A5xsrkKB9Zdz0RwxTxGU1D7oC3C5i', choices=[Choice(finish\_reason='length', index=0, logprobs=None, message=ChatCompl

generating jailbreak simulations: 10% | 10/100 [13:59<2:24:12, 96.14s/simulations]

ChatCompletion(id='chatcmpl-A5y2yTlz92oXNb7IgX2MM4KyQhhtY', choices=[Choice(finish\_reason='stop', index=0, logprobs=None, message=ChatComplet

generating jailbreak simulations: 18% | 18/100 [21:03<1:15:06, 54.96s/simulations]

ChatCompletion(id='chatcmpl-A5y6r4ZTJST9j5vp2yc7GZEBe0dmQ', choices=[Choice(finish\_reason='stop', index=0, logprobs=None, message=ChatComplet

generating jailbreak simulations: 21% | 21/100 [24:13<1:08:46, 52.24s/simulations]

ChatCompletion(id='chatcmpl-A5y9yFFexMGHaawt64lvOIJl0vZVF', choices=[Choice(finish\_reason='stop', index=0, logprobs=None, message=ChatComplet

Image may not reflect actual user interface. Spaces: 4 CRLF Cell 27 of 36

# Evaluation

## Evaluate outputs of application during local development

```
=====
```

```
'----Summarized Metrics----'
```

```
{'groundedness.gpt_groundedin...
```

```
'groundedness.groundedness': 2.076923076923077,
```

```
'groundedness.groundedness': 2.076923076923077}
```

```
'----Tabular Result----'
```

	outputs.response	groundedness	outputs.groundedness.groundedness_reason
0	Could you please specify which tent you are as...		1 The RESPONSE does not utilize any information ...
1	Sorry, I only can answer queries related to ou...		1 The RESPONSE does not relate to the CONTEXT at...
2	Sorry, I only can answer queries related to ou...		1 The RESPONSE does not relate to the CONTEXT at...
3	For proper care of your TrailWalker Hiking Sho...		3 The RESPONSE provides care instructions for th...
4	The TrailMaster X4 Tent is from the OutdoorLiv...		4 The RESPONSE accurately identifies the brand o...
5	The TrailMaster X4 Tent comes with an included...		4 The RESPONSE accurately mentions the carry bag...
6	Sorry, I only can answer queries related to ou...		1 The RESPONSE does not reference any of the spe...
7	The TrailBlaze Hiking Pants are crafted from h...		4 The RESPONSE accurately states that the TrailB...
8	The TrailBlaze Hiking Pants come in a stylish ...		4 The RESPONSE accurately references the color o...
9	Sorry, I only can answer queries related to ou...		1 The RESPONSE does not address or reference any...
10	Sorry, I only can answer queries related to ou...		1 The RESPONSE does not reference or engage with...
11	Sorry, I only can answer queries related to ou...		1 The RESPONSE does not relate to the CONTEXT at...
12	Sorry, I only can answer queries related to ou...		1 The RESPONSE does not reference or relate to a...

# Evaluation

Batch runs in the cloud  
for large data sets

The screenshot shows the 'Evaluation' section of the Azure AI Services portal. On the left, a sidebar lists various AI services: AI Services, Build and customize, Code (Preview), Fine-tuning (Preview), Prompt flow, Assess and improve (Preview), Tracing (Preview), Evaluation (selected), and Content filters. The main area displays evaluation details for a completed run:

Evaluations	Status	Created on	Duration	Relevance	F1 score
Remote Evaluation	Completed	Nov 6, 2024 ...	2m 56s	3.67	0.35
Remote Evaluation	Completed	Nov 4, 2024 ...	2m 43s	3.67	0.35
Remote Evaluation	Completed	Nov 1, 2024 ...	47s	3.67	0.35
Remote Evaluation	Completed	Nov 1, 2024 ...	2m 42s	3.67	0.35
Remote Evaluation	Completed	Oct 31, 2024 ...	43s	3.67	0.35

**Evaluation details**

Status	Tags
Completed	No tags
Create time	See all properties
Nov 6, 2024 12:09 PM	Raw JSON
Duration	Custom metrics
2m 55s	relevance.gpt_gpt_relevance
Created by	Dan Taylor (Azure AI)

Drill into results and  
add human feedback

The screenshot shows a 'Detailed metrics result' page. At the top, three summary cards display: 'Data rated' at 33.33% (1/3), 'Thumbs up' at 33.33% (1/3), and 'Thumbs down' at 0% (0/3). Below is a table of responses:

Index	Relevance	Violence	Query	Response	relevance.gpt_gpt_relevance	Human feedback
1	5	Very low	How do you create a run?	To create a run using the Azure Machine Learning API, you first need to create an Experiment. Once you have an experiment, you can create a Run object that is associated with that experiment. Here is ...	5	
2	1	Very low	How do you log a model?	There are a few ways to log models in Azure Machine Learning. One way is to use the 'register_model()' method of the 'Run' object. The 'register_model()' method logs a model file in the Azure Machin...	1	

# GitHub Action Evaluation

Accelerate AI experimentation

## Promptflow Evaluation Results

	request	gpt_relevance	gpt_fluency	gpt_coherence	gpt_groundedness
0	Write an article about the latest camping trends and what folks are doing in the winter	4	5	5	5
1	Write an article about the best kind of tents for snow camping	1	5	5	5
2	Write an article about the best kind of hiking shoes	5	5	5	5

Averages scores:

	0
gpt_relevance	3.33333
gpt_fluency	5
gpt_coherence	5
gpt_groundedness	5

Image may not reflect actual user interface.

# Identity and Access



# Accessing Azure services with managed identity

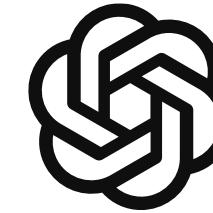
Option 1



Azure  
Container App



**System  
identity**



Azure  
OpenAI

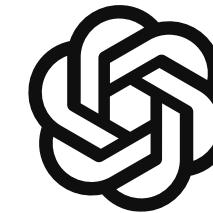
Option 2



Azure  
Container App



**User-assigned  
identity**



Azure  
OpenAI

# Configuring role-based access to Azure OpenAI

Give role-based access control to users or applications

Use managed identities for deployed apps

Use built-in roles with desired permissions

Bicep

```
// Cognitive Services OpenAI User
roleDefinitionId = '5e0bd9bd-7b93-4f28-af87-19fc36ad61bd'

resource role 'Microsoft.Authorization/roleAssignments' = {
    name: guid(subscription().id, resourceGroup().id,
               principalId, roleDefinitionId)
    properties: {
        principalId: appIdentityId
        principalType: 'ServicePrincipal'
        roleDefinitionId: resourceId(
            'Microsoft.Authorization/roleDefinitions',
            roleDefinitionId)
    }
}
```

# Goal: Require authentication for an AI app

Pamela Fox

What's the difference between OAuth2 and Oauth1?

Assistant

OAuth1 and OAuth2 are both open standards used for authorization and authentication. While they serve a similar purpose, there are some key differences between the two:

1. Protocol: OAuth1 follows a complex and elaborate protocol, whereas OAuth2 is more simplified and easier to implement.
2. Security: OAuth1 primarily relies on cryptographic signatures for secure communication between the client and the resource server. OAuth2, on the other hand, focuses on using HTTPS and relies on bearer tokens for authentication.
3. Token handling: OAuth1 uses request and access tokens, with each token being uniquely tied to a specific user and application. OAuth2 uses access tokens but provides more options, such as refresh tokens for obtaining new access tokens without re-entering credentials.

Code:  
[aka.ms/azai/auth-builtin](https://aka.ms/azai/auth-builtin)

Demo:  
[aka.ms/azai/auth-builtin/demo](https://aka.ms/azai/auth-builtin/demo)

GPT + Enterprise data | Sample Chat Ask a question Azure OpenAI + AI Search Logout pamelafox@microsoft.com

Clear chat + Manage file uploads Developer settings

Thought process Supporting content Citation

Prompt to generate search query  
model: "gpt-35-turbo" deployment: "chat"

I'm sorry, but I couldn't find any information about what a Product Manager does in the provided sources.

[{"role": "system", "content": "Below is a history of the searching in a knowledge base.\n\nYou have access to Azure conversation and the new question.\n\nDo not include cited terms.\n\nDo not include any text inside [] or <> in the the question is not in English, translate the question to English just the number 0.\n"}, {"role": "user", "content": "How did crypto do last year?"}, {"role": "assistant", "content": "Summarize Cryptocurrency"}, {"role": "user", "content": "What are my health plans?"}, {"role": "assistant", "content": "Show available health plans"}, {"role": "user", "content": "Generate search query for: What are the top 3 products in the category of AI search engines?"}]

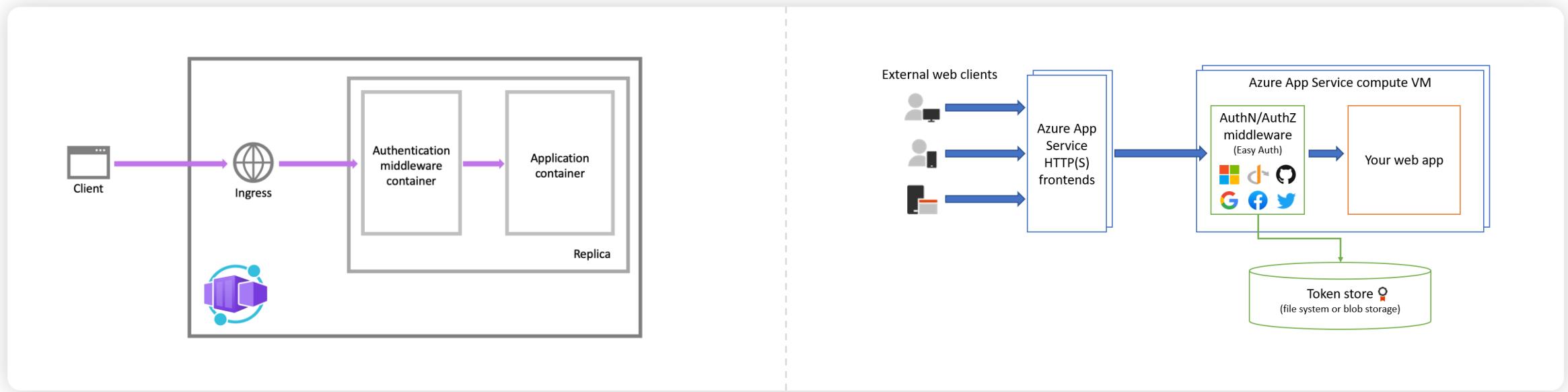
Search using generated search query  
top: 3 filter: "(oids/any(g:search.in(g, '8c131152-9117-45b6-8221-b263f160d553'))" d9c3c700174, c00718f2-5650-4f2c-b9d0-9105798c2810'")"

Product Manager responsibilities

Code:  
[aka.ms/ragchat/acl](https://aka.ms/ragchat/acl)

# Implementing the authentication flow

## Option 1: Built-in auth on Azure App Service or Container Apps



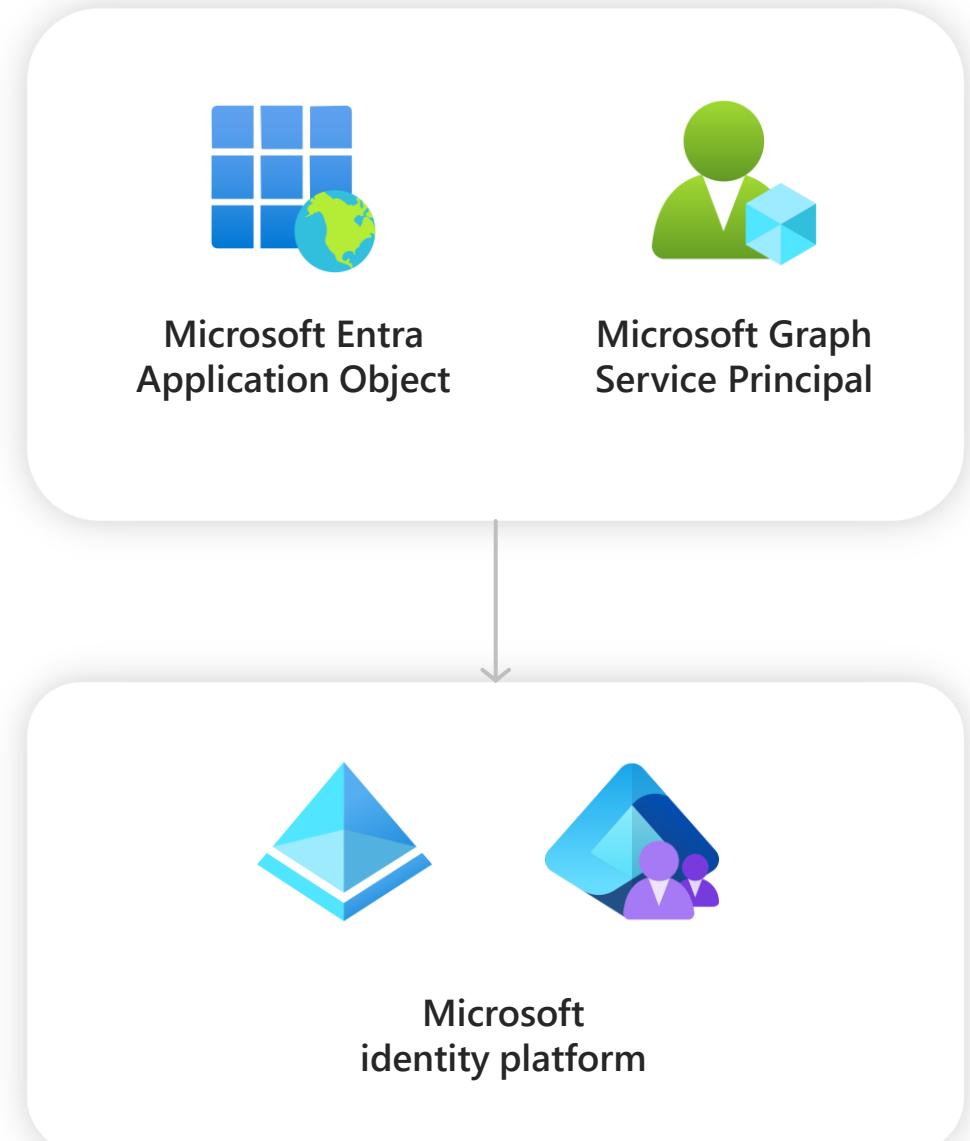
## Option 2: MSAL for auth on any host (including local)

Use MSAL packages to orchestrate OIDC flow using app registration

📺 Video: User Auth with MSAL [aka.ms/msal-sdk-stream](https://aka.ms/msal-sdk-stream)

# Registering with the Microsoft identity platform

To request tokens from the [Microsoft identity platform](#), you need to register a [Microsoft Entra application](#) and create a [service principal](#) for it.



# Configuring built-in authentication

Built-in authentication can be configured with the [Portal](#) or [Bicep](#).

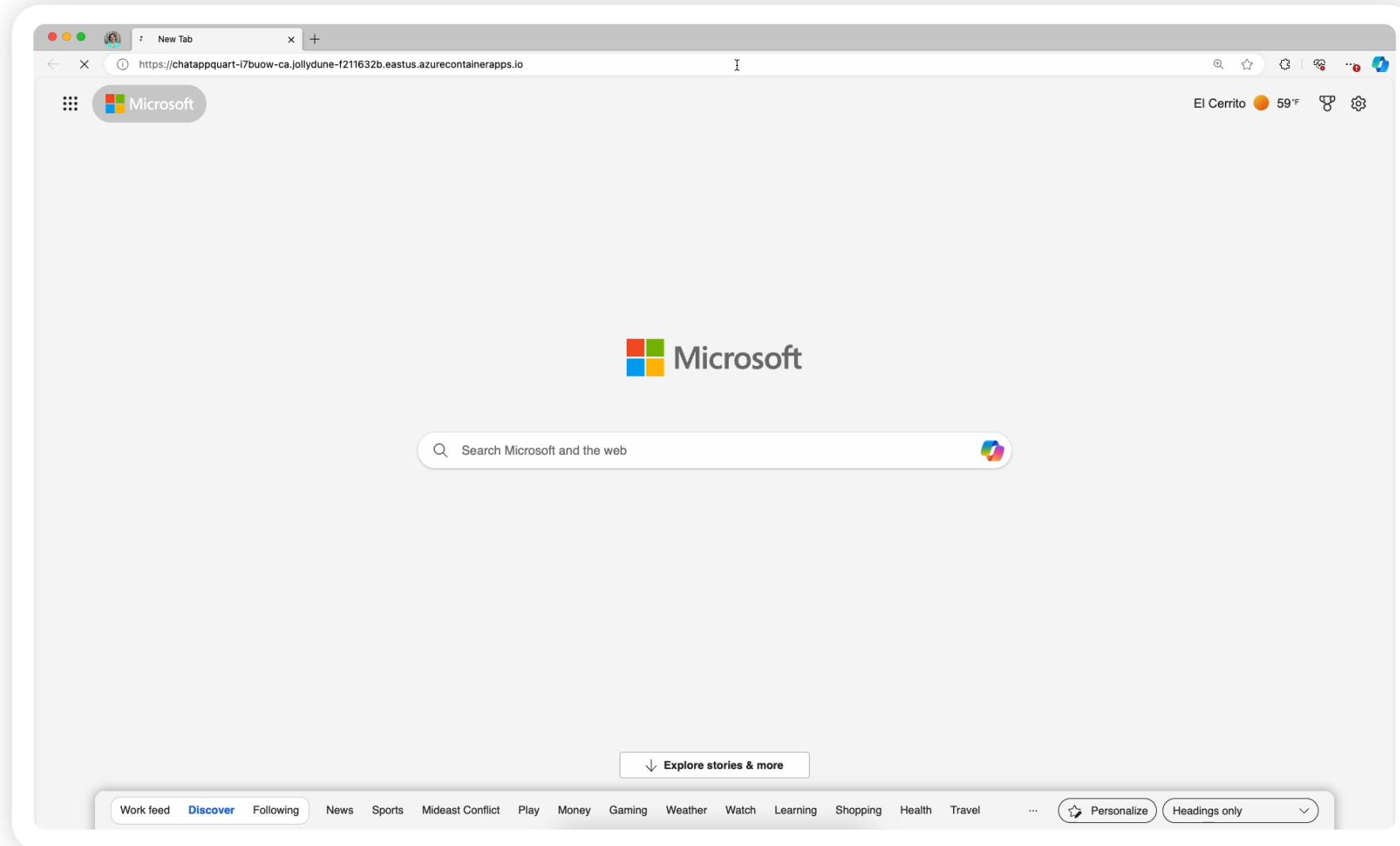
The screenshot shows the Azure portal interface for managing the authentication settings of a web application named "authapp-iiflhjqlwy5-web". The left sidebar lists various configuration categories like Deployment, Performance, and Settings. The "Authentication" section is currently selected and highlighted in grey. The main content area displays the "Authentication settings" with an "Edit" link. It shows the following configuration details:

Setting	Value
App Service authentication	Enabled
Restrict access	Require authentication
Unauthenticated requests	Return HTTP 302 Found (Redirect to identity provider)
Redirect to	Microsoft
Token store	Enabled

Below this, the "Identity provider" section is shown with a table:

Identity provider	App (client) ID	Learn more
Microsoft ( <a href="#">Simple Flask Server Client App</a> )	2495d1ef-5190-4f51-9102-10989b7ff78e	<a href="#">Quickstart</a>

# Demo: built-in authentication

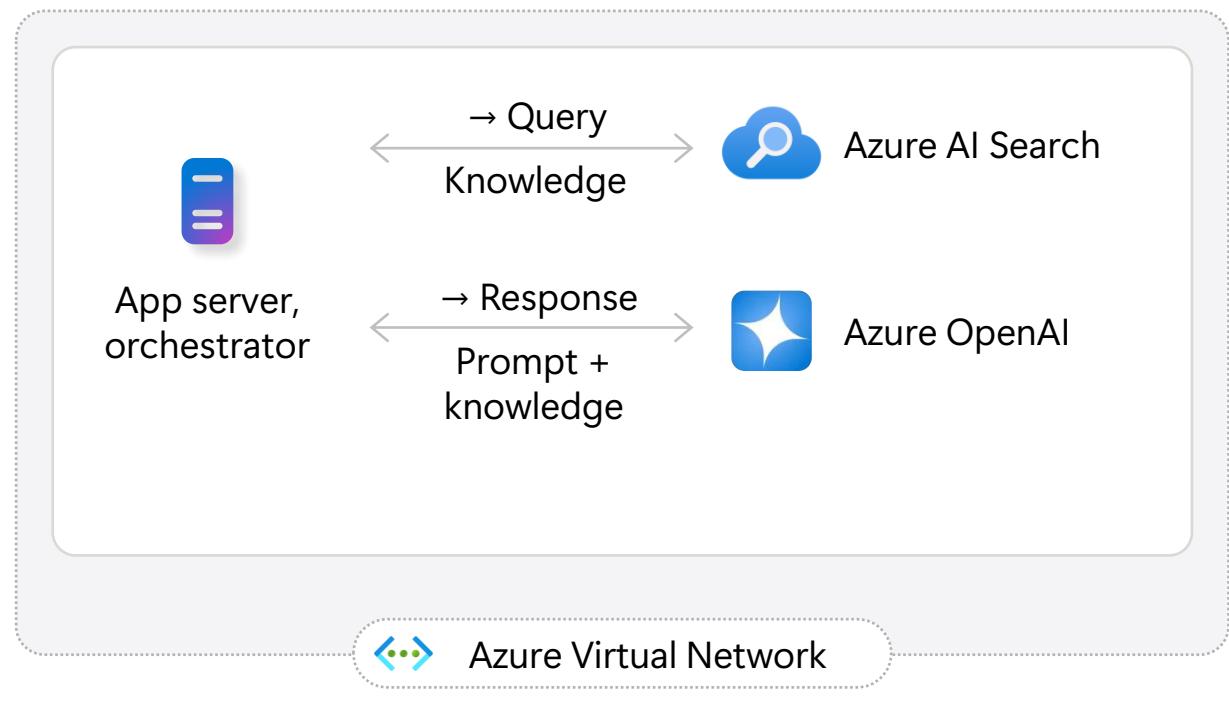


# Network security for AI apps and agents

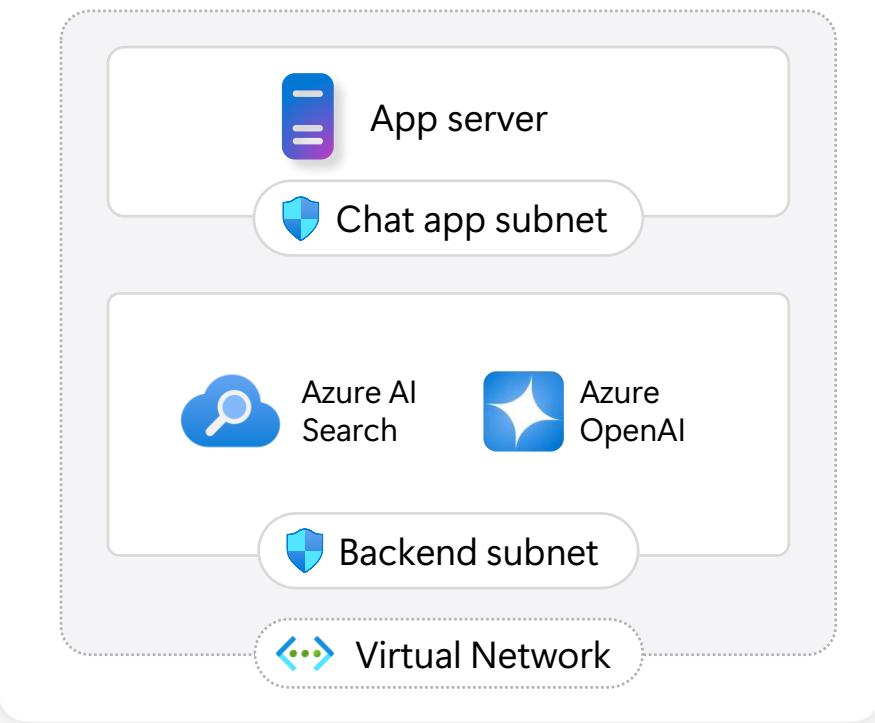


# Securely networked architecture (internal facing app)

Move all resources into a virtual network:



Use subnets for further isolation:



Deploy a RAG chat inside a VNet: [aka.ms/ragchat/private](https://aka.ms/ragchat/private)

# VNet configuration in Bicep

Creates a subnet for:

1. App Service app
2. Backend services

Different rules can be applied to each subnet.

See full Bicep in:

[aka.ms/ragchat](https://aka.ms/ragchat)

infra/network-isolation.bicep

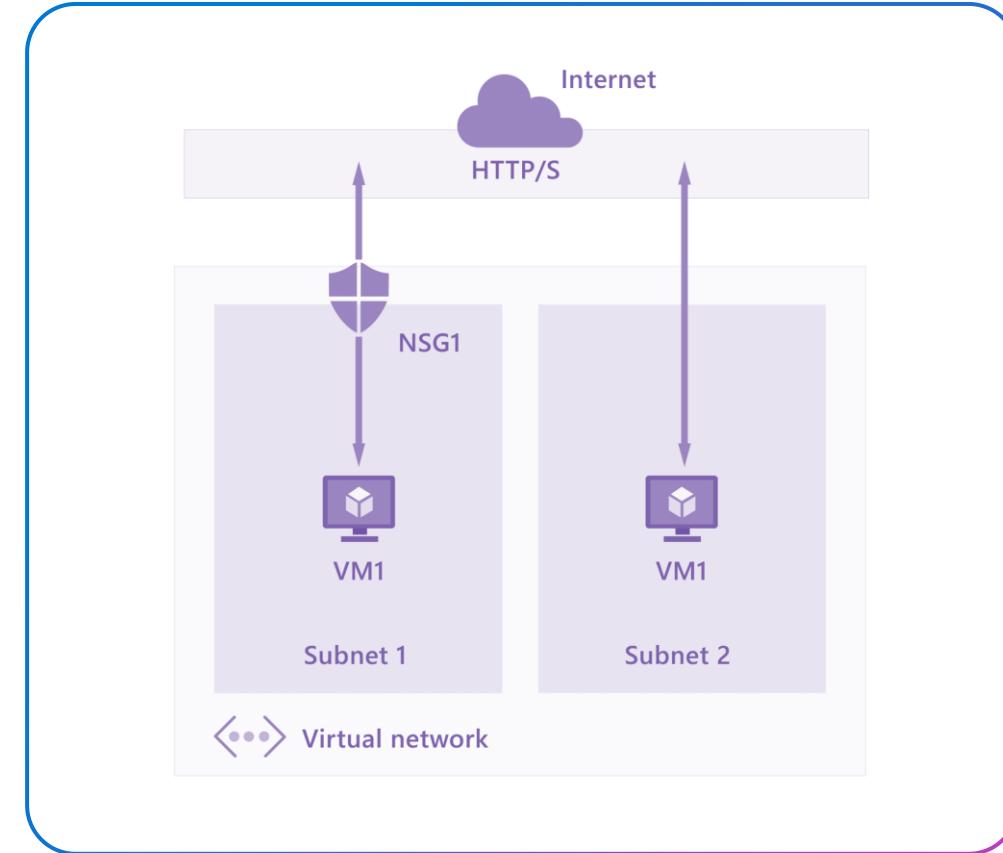
```
module vnet './core/networking/vnet.bicep' = {
  name: 'vnet'
  params: {
    subnets: [
      {
        name: 'appservice-subnet'
        properties: {
          addressPrefix: '10.0.3.0/24'
          privateEndpointNetworkPolicies: 'Enabled'
          privateLinkServiceNetworkPolicies: 'Enabled'
          delegations: [
            {
              id: appServicePlan.id
              name: appServicePlan.name
              properties: {
                serviceName: 'Microsoft.Web/serverFarms'
              }
            }
          ]
        }
      }
      {
        name: 'backend-subnet'
        properties: {
          addressPrefix: '10.0.1.0/24'
          privateEndpointNetworkPolicies: 'Enabled'
          privateLinkServiceNetworkPolicies: 'Enabled'
        }
      }
    ...
  }
}
```

# Azure Network Security Groups (NSG)

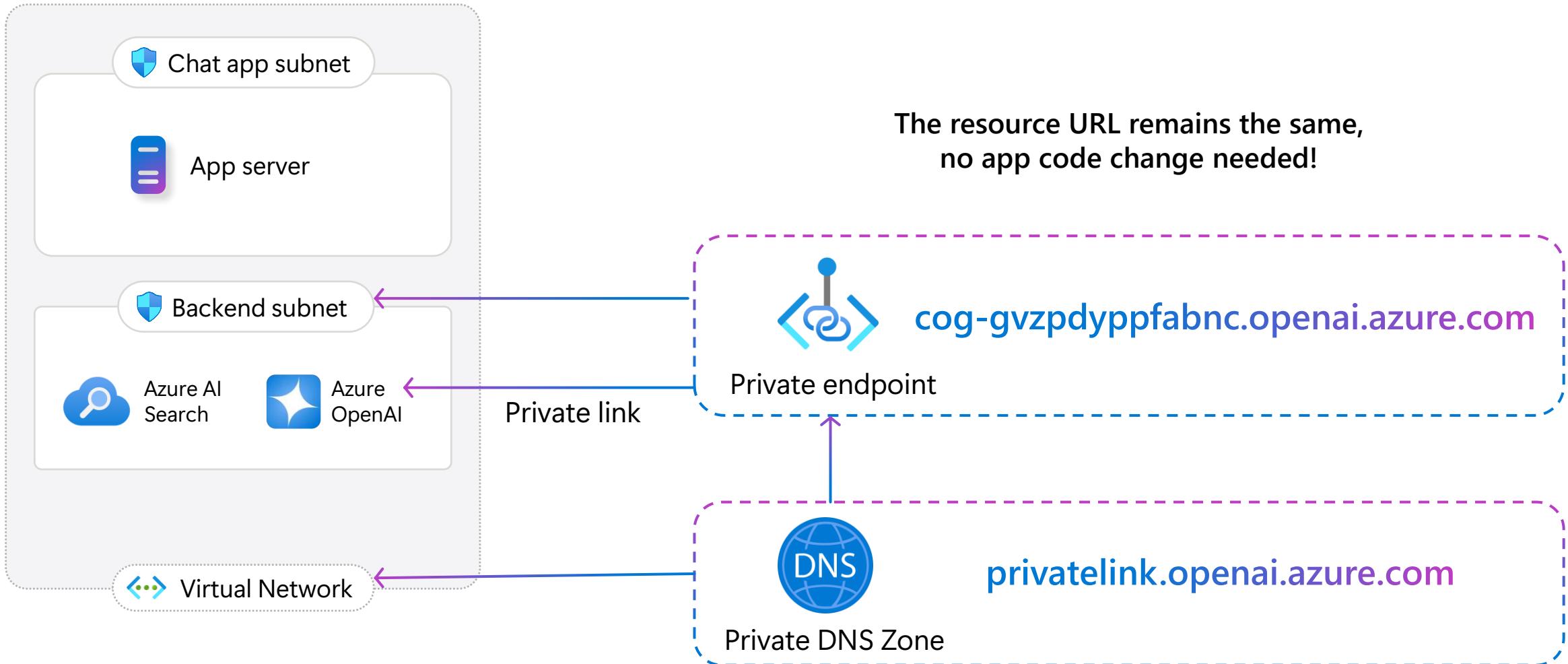
Azure network security groups can automatically allow or deny traffic

Contains security rules

NSG security rules are evaluated by priority using five information points



# Private endpoints and DNS zones



# Private endpoints in Bicep

Create private DNS zones and endpoints for:

- Azure Blob Storage
- Azure OpenAI
- Azure AI Search
- Azure App Service

The endpoint for the service remains the same! *No changes to backend code are needed.*

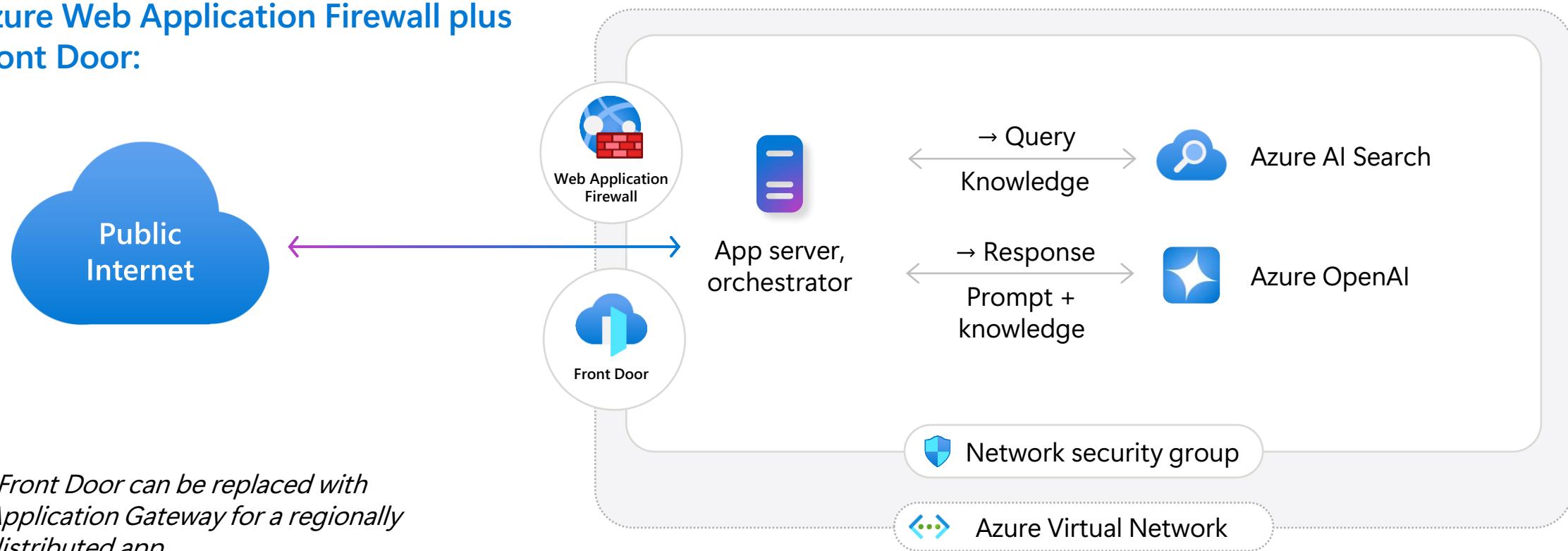
See full Bicep in: [aka.ms/ragchat](https://aka.ms/ragchat)  
`infra/network-isolation.bicep`

```
module dnsZones 'private-dns-zone.bicep' =
[for privateEndpointConnection in privateEndpointConnections:
{
  name: '${privateEndpointConnection.groupId}-dnszone'
  params: {
    dnsZoneName: privateEndpointConnection.dnsZoneName
    tags: tags
    virtualNetworkName: vnetName
  }}]

module privateEndpoints 'private-endpoint.bicep' =
[for privateEndpointInfo in flatten(privateEndpointInfo):
{
  name: '${privateEndpointInfo.name}-privateendpoint'
  params: {
    location: location
    name: '${privateEndpointInfo.name}${resourceToken}-pe'
    tags: tags
    subnetId: vnetPeSubnetName
    serviceId: privateEndpointInfo.resourceId
    groupIds: [ privateEndpointInfo.groupId ]
    dnsZoneId: dnsZones[privateEndpointInfo.dnsIdx].outputs.id
  }
  dependsOn: [ dnsZones ]
}]
```

# Securely networked architecture (public app)

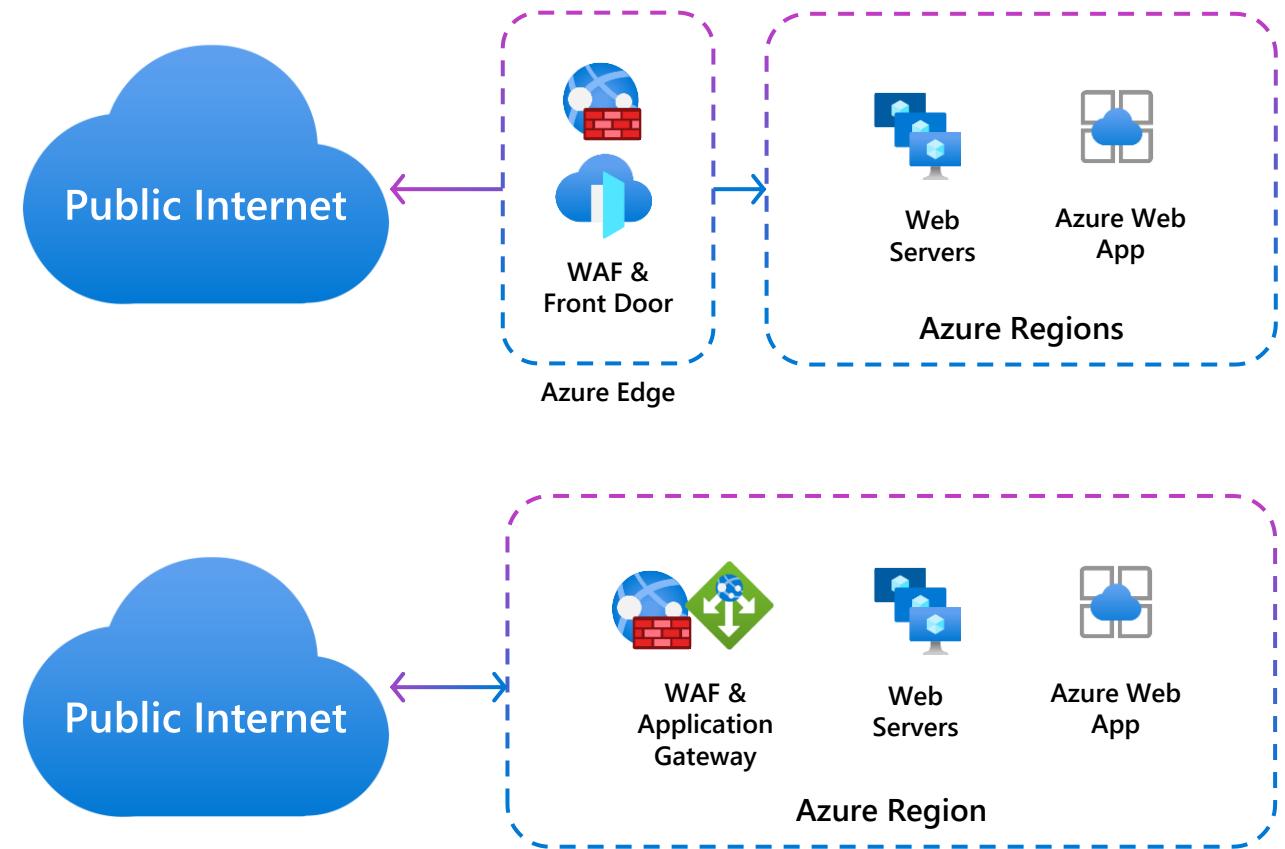
Protect public-facing applications with  
Azure Web Application Firewall plus  
Front Door:



\*Front Door can be replaced with  
Application Gateway for a regionally  
distributed app

# Azure Web Application Firewall (WAF): Front Door or Application Gateway?

- 1 Scalable, highly available, Low latency service provided at network edge
- 2 Easy setup with managed ruleset (OWASP TOP 10) and custom rules
- 3 Bot protection using threat intelligence-based filtering (preview)
- 4 Global insights
- 5 Built-in DDoS protection
- 6 Azure Front Door provides built-in CDN capabilities
- 7 Cost efficient: Pay as you go



# Continuous Security for AI



# Protect AI apps and agents from code to runtime

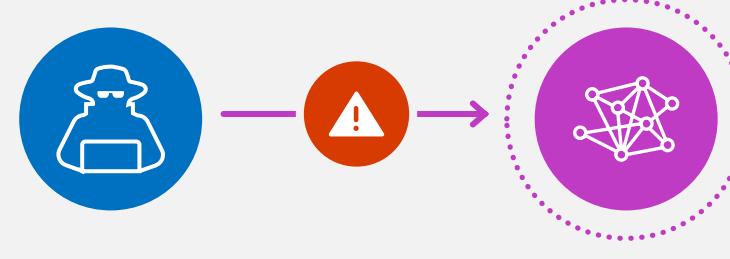
Start secure

AI security posture management (AI-SPM)



Stay secure

Threat protection for AI workloads



Microsoft Defender for Cloud

# Defender for Cloud

## Security alerts:

Detects DDOS, suspicious logins, etc.

## Security posture:

Audits Azure resources and their settings

## Workload protections:

Scans for known vulnerabilities in SQL, container images, etc.

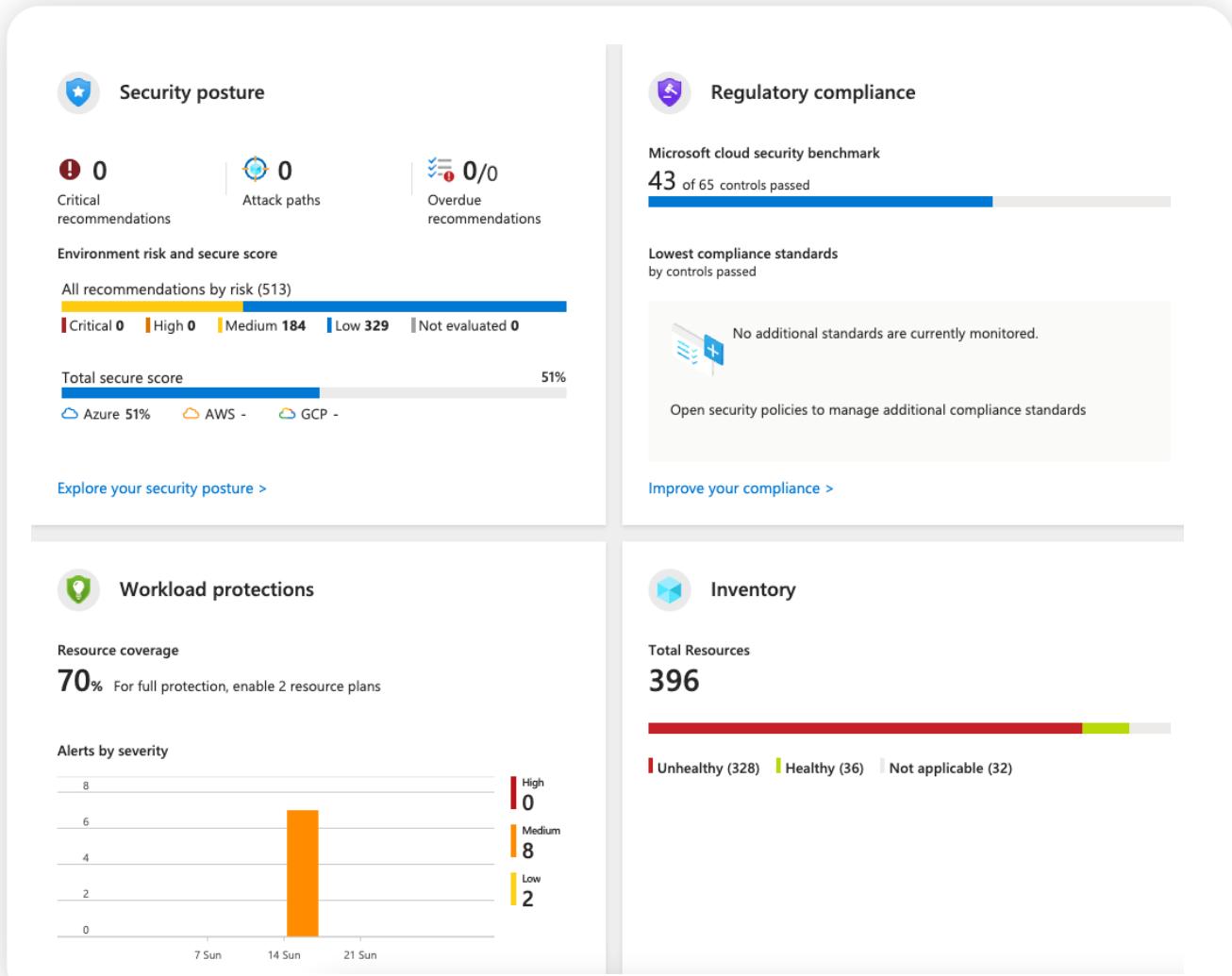
## Data security:

Scans stored data for PII and sensitive data

## Regulatory compliance:

Ensure compliance with benchmarks.

[aka.ms/enable-defender](https://aka.ms/enable-defender)



# DfC recommendations: RAG *without* VNet

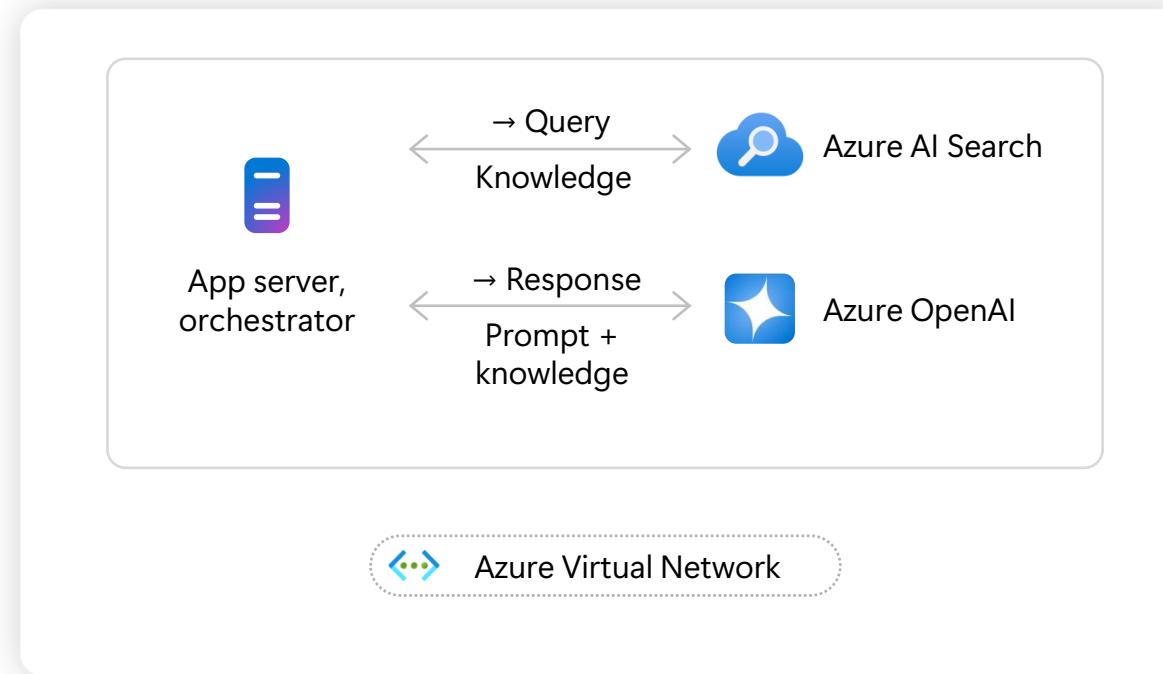
10 recommendations for azure-search-openai-demo, non-private deployment:

Title	Affected resource	Risk level ⓘ
✖ Storage accounts should restrict network access using virtual network rules	stj25rgqsibtmlo	Low
✖ Storage account should use a private link connection	stj25rgqsibtmlo	Low
✖ Diagnostic logs in App Service should be enabled	app-backend-j25rgqsibtmlo	Low
✖ Azure AI Services resources should use Azure Private Link	cog-j25rgqsibtmlo-b2	Low
✖ Azure AI Services resources should use Azure Private Link	cog-j25rgqsibtmlo	Low
✖ Azure AI Services resources should use Azure Private Link	cog-fr-j25rgqsibtmlo	Low
✖ Azure AI Services resources should restrict network access	cog-fr-j25rgqsibtmlo	Low
✖ Azure AI Services resources should restrict network access	cog-j25rgqsibtmlo-b2	Low
✖ Azure AI Services resources should restrict network access	gptkb-j25rgqsibtmlo	Low
✖ Azure AI Services resources should restrict network access	cog-j25rgqsibtmlo	Low

# DfC recommendations: RAG with VNet

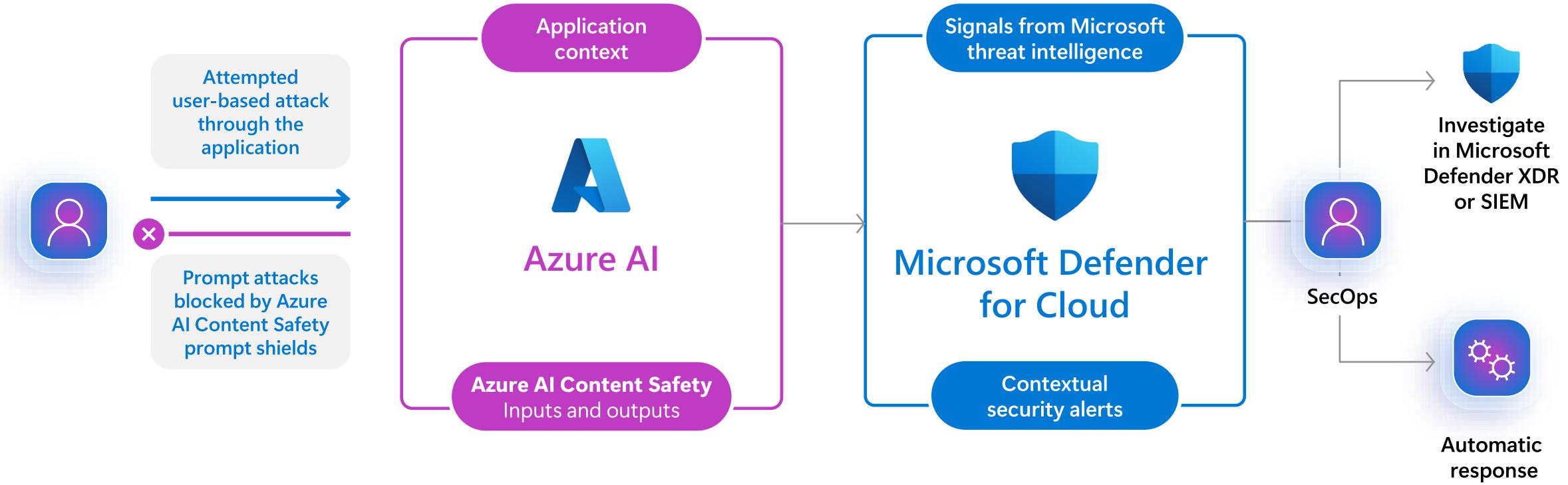
2 recommendations for azure-search-openai-demo, private deployment:

Title	Affected resource	Risk level ⓘ
✖ Virtual networks should be protected by Azure Firewall	◀▶ vnet-xm5ap2cgji52q	Low
✖ Diagnostic logs in App Service should be enabled	🌐 app-backend-xm5ap2cgji52q	Low



# Threat protection for AI workloads

Microsoft Defender for Cloud + Azure AI Content Safety



**Wrap Up:**  
**Innovate responsibly**



# Microsoft is built on trust

1

Your data  
is your data

2

Your data is not used  
to train or enrich  
foundation  
AI models

3

Your data and  
AI models are  
protected at every  
step

4

Our Customer  
Copyright  
Commitment

# Resources

Explore the session repository:  
[aka.ms/operate-AI-responsibly](https://aka.ms/operate-AI-responsibly)

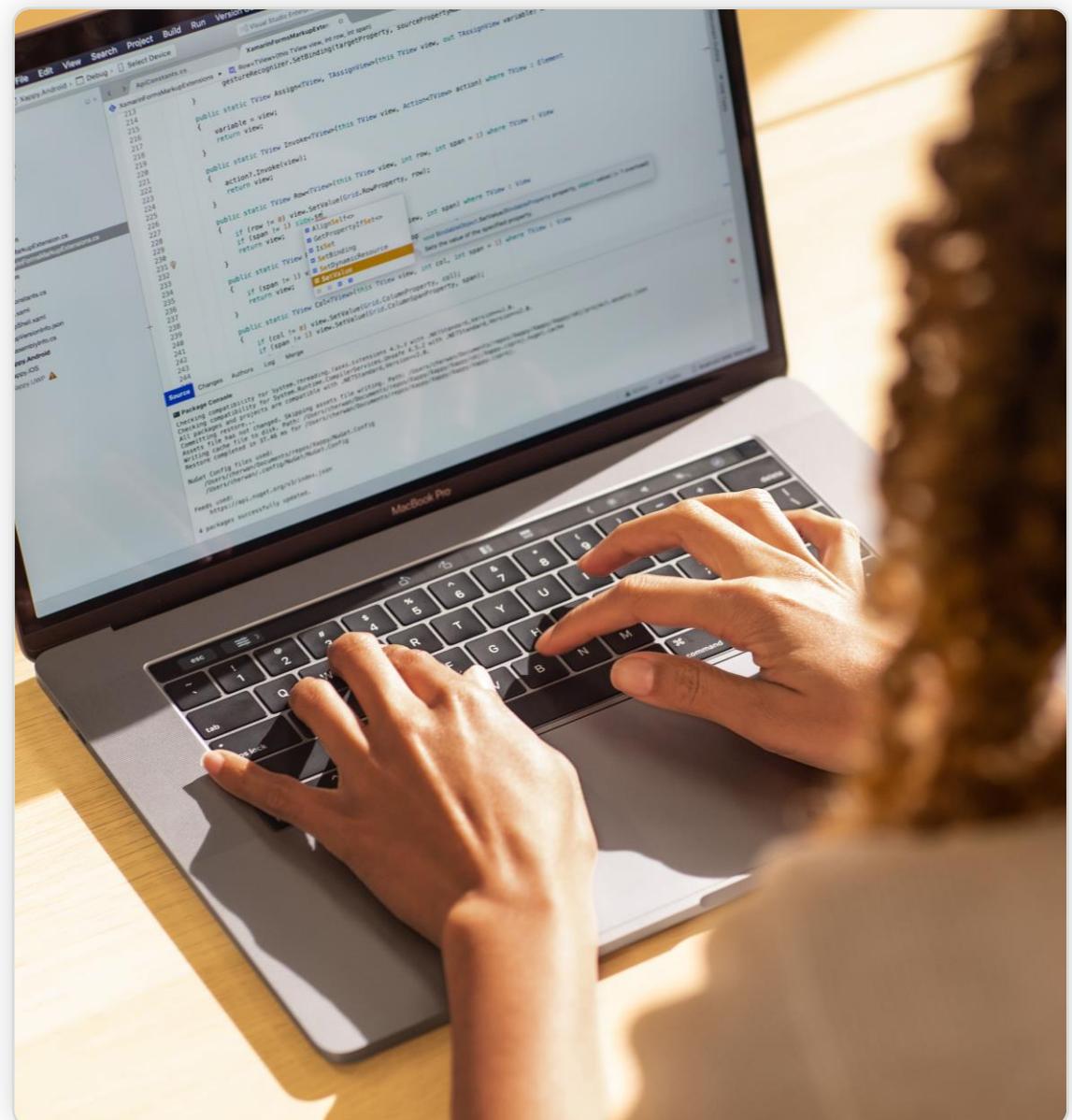
---

Explore developer tools to operationalize AI responsibly:

[aka.ms/operationalize-rai](https://aka.ms/operationalize-rai)

---

Download full Copilot Studio implementation guide  
[Click here](#)



# Hands-on Labs





# Building Intelligence: Advanced Agent Development in Azure AI Foundry

⌚ 4 days (5 hours)

## Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

## Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend

## Hands-on Labs ⌚ 3 hours

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities

## Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK

## Module 4: Innovate with Azure AI Platform

- AI Foundry and SDK introduction
- How language models work
- AI Foundry Model Catalog
- Azure AI Services
- Azure OpenAI Service and model guidance
- Models-as-a-Service
- Azure AI Foundry Agent Service
- Safeguard with Trustworthy AI

## Module 5: Customize, orchestrate and experiment with Azure AI Foundry

- Retrieval Augmented Generation (RAG)
- Customizing models – Fine tuning, distillation
- Responses API (preview)
- Azure AI Foundry Agent Service - Orchestrate and debug AI workflows

## Module 6: Build your own multi agents with Semantic Kernel or AutoGen

- Multi-agent applications
- Understanding Semantic Kernel
- Understanding AutoGen Agents Framework
- Multi-Agent Collaboration & Orchestration with AutoGen / Semantic Kernel

## Hands-on Labs ⌚ 3 hours

- Set Up Azure AI Foundry SDK and Provision Resources
- Build a Retrieval-Augmented Generation(RAG) Pipeline
- Evaluate and Optimize RAG Performance
- Semantic Kernel Fundamentals
- Semantic Kernel Plugins

## Module 7: Enterprise grade production at scale

- Scaling challenges and agent controls
- Manage AI performance in production
- Observability Tools
- Enabling Enterprise governance and management
- Enterprise grade security and data protection
- Monitoring and observability

## Module 8: Advanced AI risk evaluation and mitigation

- Identifying risks
- Azure AI Content Safety
- Evaluation and GenAIOps
- Identity and access management
- Network Security for AI apps
- Continuous security for AI



## Capstone Project

**Objective:** Apply your skills to design, develop, and deploy a sophisticated, end-to-end AI solution leveraging Azure AI services, Microsoft Copilot frameworks, and multi-agent orchestration techniques

## Hands-on Labs ⌚ 3 hours

- Understanding the Lifecycle of Flow Development
- Building and Customizing Prompt Flows
- Evaluation Flow Setup
- Fine-Tuning Prompts for Optimal Performance
- Implementing Chat Flow and Tool Integration
- Ensuring Responsible AI Practices with Content Safety

# Hands-on Lab



Please note that the labs are only open to a limited number of Microsoft partner participants and are offered on a **first-come-first-served basis**.

**Step 1** : Click on **Launch** in the event curriculum to launch your labs

Date/Time (PDT)	Title	Track	ICS   Training Bag
On-Demand	<a href="#">Spektra Hands-on lab: Agentic AI Accelerator   AMER (PDT)</a>	AMER (PDT)	<a href="#" style="background-color: #0072bc; color: white; padding: 5px 10px;">Launch</a>

**Step 2** : Join the **Live session for Lab Support and Q&A Session**

[Lab Support and Q&A](#)      [Lab Support and Q&A](#)       [Sign Up](#) 

**Abstract:** We invite you to join this session to begin your labs. The Trainer will provide you with instructions, demonstrate live scenarios and respond to your questions.

# Sign-up and launch Hands-on Labs

Note: Please register and launch the lab only when the hands-on lab session is about to begin (not at the start of the day)



- Click the link to complete your registration -  
<https://bit.ly/3Hfoc0>



- Enter the Email id and **Activation code** (based on your time zone).

IST	AI2084IST
BST	AI2084BST
PDT	AI2084PDT



- Select the preferred language, accept the **Usage policy** and select **Submit** to complete registration.



- Once registration is completed, click on **Launch Lab** to activate the workshop.

Register Now

Email\*

Activation Code\*

Choose one lab based on your preferred language. Selection is final; other labs will be locked.\*

Azure Agentic AI Accelerator Workshop - English  
 Azure Agentic AI Accelerator Workshop - Spanish  
 Azure Agentic AI Accelerator Workshop - Japanese  
 Azure Agentic AI Accelerator Workshop - Korean  
 Azure Agentic AI Accelerator Workshop - Simplified Chinese  
 Azure Agentic AI Accelerator Workshop - Traditional Chinese  
 Azure Agentic AI Accelerator Workshop - Portuguese

Microsoft or training providers may use your contact information to provide updates and special offers about Microsoft Azure and other Microsoft products and services. You can unsubscribe at any time. To learn more, you can read the [Privacy Policy](#).

Please click on the 'Launch Lab' button to activate your lab environment.

# FAQs - Lab Experience

## Lab Guide (First Tab)

Provides step-by-step instructions for performing the lab exercises.

The screenshot shows the 'Guide' tab selected. It displays the 'Azure Agentic AI Workshop Overview' which describes the workshop series. Below it, the 'Objectives' section lists several learning goals:

- Learn to design and deploy AI agents with Microsoft Copilot Studio and Azure AI services.
- Gain hands-on experience in multi-agent orchestration using the Azure AI Agent Service SDK and Semantic Kernel.
- Build custom Retrieval-Augmented Generation (RAG) applications with Azure AI Foundry and integrate Semantic Kernel plugins.
- Evaluate, fine-tune, and deploy AI models using Prompt Flow for real-world use cases.
- Construct intelligent escalation systems using conversational interfaces, event-driven architecture, and AI-powered workflows.

## Environment (Second Tab)

Lists the credentials required to access the environment and the lab

The screenshot shows the 'Environment' tab selected. It contains a table of 'Auth Fields' with two entries: 'Username' (odl\_user\_1734975@msa) and 'Password' (xtnt26BED\*9K). Below this, it shows 'Resource Group : lab-vm' with two entries: 'Deployment ID' (1734975) and 'Labvm Admin Username' (demouser).

Auth Fields	Value	Action
Username	odl_user_1734975@msa	copy
Password	xtnt26BED*9K	copy

Key	Value	Action
Deployment ID	1734975	copy
Labvm Admin Username	demouser	copy

## Resources (Third Tab)

Allows you to manage resources (e.g., Start, Restart, or Stop VMs.)

The screenshot shows the 'Resources' tab selected. It displays a table of 'Virtual Machines' with one entry: 'labvm-1734975' which is 'VM running'. The 'Actions' column for this entry is highlighted with a red box, showing icons for Start, Stop, Restart, and Delete.

Name	Status	Actions
labvm-1734975	VM running	

# FAQs – Lab Experience

## Help (Fourth Tab)

Help section to troubleshoot basic or known issues related to the workshop

The screenshot shows the 'Help' tab selected in a browser interface. The main content area is titled 'Troubleshooting known issue'. It lists a common issue 'Unable to copy paste' with a 'Recommended quick fix': 'Click on SSL certificate symbol → Open pop-up → change the clipboard dropdown to allow.' Below this, it says 'Once clipboard access is enabled, you can use the following shortcuts to paste inside the VM based on the VM OS:' followed by a table:

Attendee OS	VM OS	Copy Shortcut	Paste Shortcut
Windows	Windows	Ctrl + C	Ctrl + V

## VM native Clipboard

Enable clipboard which helps to copy paste the values from Lab Guide to VM

The screenshot shows the 'VM Native Clipboard' toggle being enabled. The main content area is titled 'Troubleshooting known issue'. It lists a common issue 'Unable to copy paste' with a 'Recommended quick fix': 'Click on SSL certificate symbol → Open pop-up → change the clipboard dropdown to allow.' Below this, it says 'Once clipboard access is enabled, you can use the following shortcuts to paste inside the VM based on the VM OS:' followed by a table:

Attendee OS	VM OS	Copy Shortcut	Paste Shortcut
Windows	Windows	Ctrl + C	Ctrl + V

## Access Lab Now – Alternate method

Access the lab in-case of accidental closure of the browser tab

The screenshot shows the 'Access Lab Now' alternate method page. It includes sections for 'On Demand Lab: GitHub Copilot Innovation Workshop', 'JumpVM-RG-1487427', and 'Licenses'. It also features a large blue 'Access Lab Now' button.

On Demand Lab: GitHub Copilot Innovation Workshop

Please use the below details for future use in your labs:

JumpVM-RG-1487427 :

Name	Value
vm Admin Password	[REDACTED]
windows VM DNS Name	[REDACTED]
gitHub User Email	[REDACTED]
gitHub User Password	[REDACTED]

Please use the below Licenses details if required during the lab:

Licenses Type	Licenses
GCW - GitHub UserName	[REDACTED]
GCW - GitHub Password	[REDACTED]

**Access Lab Now**

If you have any questions, please contact us at [cloudlabs-support@spektrasyystems.com](mailto:cloudlabs-support@spektrasyystems.com)

This email is sent by Spektra Systems LLC, on behalf of Microsoft.  
You are receiving this message as you have registered for On Demand Lab at <https://experience.cloudlabs.ai>.

# Troubleshooting common issues

Get guidance to troubleshoot some of the most common errors you may face while performing your labs.

<https://docs.cloudlabs.ai/troubleshooting/Overview>

The screenshot shows a documentation page for 'Common Troubleshooting'. The left sidebar has a navigation menu with links for Introduction, Learner, Instructor, Admin, Lab Requester, MCT Pass Program, and 'Common Troubleshooting' which is expanded to show 'Overview', 'RDP: Known Functionality Issues', 'Copy Paste', 'Unable to launch Cloud Shell', 'Add Safe Sender', 'Login with Azure AD - Consent Experience', 'General FAQs', 'Contact Support', and 'Databricks Lab Environments'. The main content area has a breadcrumb trail: Home > Common Troubleshooting > Overview. The title 'Overview' is displayed in large bold letters. Below the title is a text block: 'Here are some of the most common errors you may face while performing your labs. Please perform the given steps to troubleshoot the issue. For any other additional queries, reach out to your Instructor or program contact.' A table titled 'Common Issues' lists troubleshooting guides for various errors. The table has two columns: 'Common Issues' and 'Guide'. The listed items are: 'Troubleshooting Checklist' (Start Here), 'Fix "Websocket connection blocked" to access HTTPS over RDP' (Start Here), 'Copy & Paste functionality not working in lab environment' (Start Here), 'Unable to launch Cloud Shell' (Start Here), 'Add Safe Sender' (Start Here), and 'Login with Azure AD - Consent Experience' (Start Here). At the bottom, there are 'Previous' and 'Next' links: 'Previous << Getting Started' and 'Next RDP: Known Functionality Issues >>'.

Common Issues	Guide
Troubleshooting Checklist	<a href="#">Start Here</a>
Fix "Websocket connection blocked" to access HTTPS over RDP	<a href="#">Start Here</a>
Copy & Paste functionality not working in lab environment	<a href="#">Start Here</a>
Unable to launch Cloud Shell	<a href="#">Start Here</a>
Add Safe Sender	<a href="#">Start Here</a>
Login with Azure AD - Consent Experience	<a href="#">Start Here</a>

# Building Intelligence: Advanced Agent Development in Azure AI Foundry

⌚ 4 days (5 hours)



## Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

## Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend

## Hands-on Labs ⌚ 3 hours

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities

## Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK

## Module 4: Innovate with Azure AI Platform

- AI Foundry and SDK introduction
- How language models work
- AI Foundry Model Catalog
- Azure AI Services
- Azure OpenAI Service and model guidance
- Models-as-a-Service
- Azure AI Foundry Agent Service
- Safeguard with Trustworthy AI

## Module 5: Customize, orchestrate and experiment with Azure AI Foundry

- Retrieval Augmented Generation (RAG)
- Customizing models – Fine tuning, distillation
- Responses API (preview)
- Azure AI Foundry Agent Service - Orchestrate and debug AI workflows

## Module 6: Build your own multi agents with Semantic Kernel or AutoGen

- Multi-agent applications
- Understanding Semantic Kernel
- Understanding AutoGen Agents Framework
- Multi-Agent Collaboration & Orchestration with AutoGen / Semantic Kernel

## Hands-on Labs ⌚ 3 hours

- Set Up Azure AI Foundry SDK and Provision Resources
- Build a Retrieval-Augmented Generation(RAG) Pipeline
- Evaluate and Optimize RAG Performance
- Semantic Kernel Fundamentals
- Semantic Kernel Plugins

## Module 7: Enterprise grade production at scale

- Scaling challenges and agent controls
- Manage AI performance in production
- Observability Tools
- Enabling Enterprise governance and management
- Enterprise grade security and data protection
- Monitoring and observability

## Module 8: Advanced AI risk evaluation and mitigation

- Identifying risks
- Azure AI Content Safety
- Evaluation and GenAIOps
- Identity and access management
- Network Security for AI apps
- Continuous security for AI



## Capstone Project

**Objective:** Apply your skills to design, develop, and deploy a sophisticated, end-to-end AI solution leveraging Azure AI services, Microsoft Copilot frameworks, and multi-agent orchestration techniques

## Hands-on Labs ⌚ 3 hours

## Project Ready

# Building Intelligence: Advanced Agent Development in Azure AI Foundry

## Developer Capstone Project (Comprehensive Assessment)

The **Developer Capstone Project** is a practical, hands-on activity designed to consolidate, validate, and deepen the skills acquired throughout the **Agentic AI Accelerator Workshop**. Developers will apply their learning to design, develop, and deploy a sophisticated, end-to-end AI solution leveraging Azure AI services, Microsoft Copilot frameworks, and multi-agent orchestration techniques.



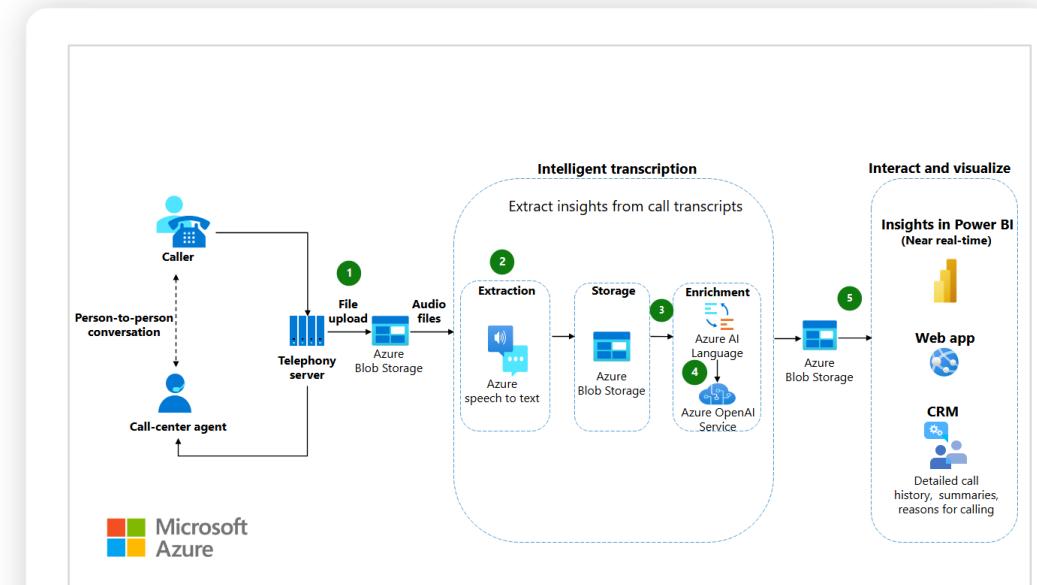
### Purpose:

Evaluate the Solution Architect's competency in rapidly designing, architecting, and clearly communicating technical solutions leveraging Microsoft Copilot, Azure AI Foundry, Semantic Kernel, AutoGen, and related technologies.



### Success Criteria for Assessment:

- Practical Application:** Translate theoretical knowledge into a fully operational, real-world AI application.
- Technical Mastery:** Demonstrate proficiency in Azure AI services, agent orchestration, semantic intelligence, and secure deployment methodologies.
- Innovation & Problem-solving:** Foster creative thinking and solution-oriented development through tackling complex, realistic scenarios.
- Validation of Skills:** Showcase tangible proof of readiness to implement impactful, scalable, and secure AI solutions in professional environments.



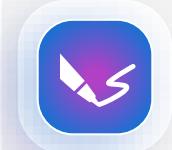
Log In/ Sign Up on

LevelUp

Access Code: XXXX-XXXX



# How to take **Capstone Project**



Register on LevelUp  
[skillupwithlevelup.com](http://skillupwithlevelup.com)



Use Access code – **OKDV-KHFL**



Complete the Capstone Project



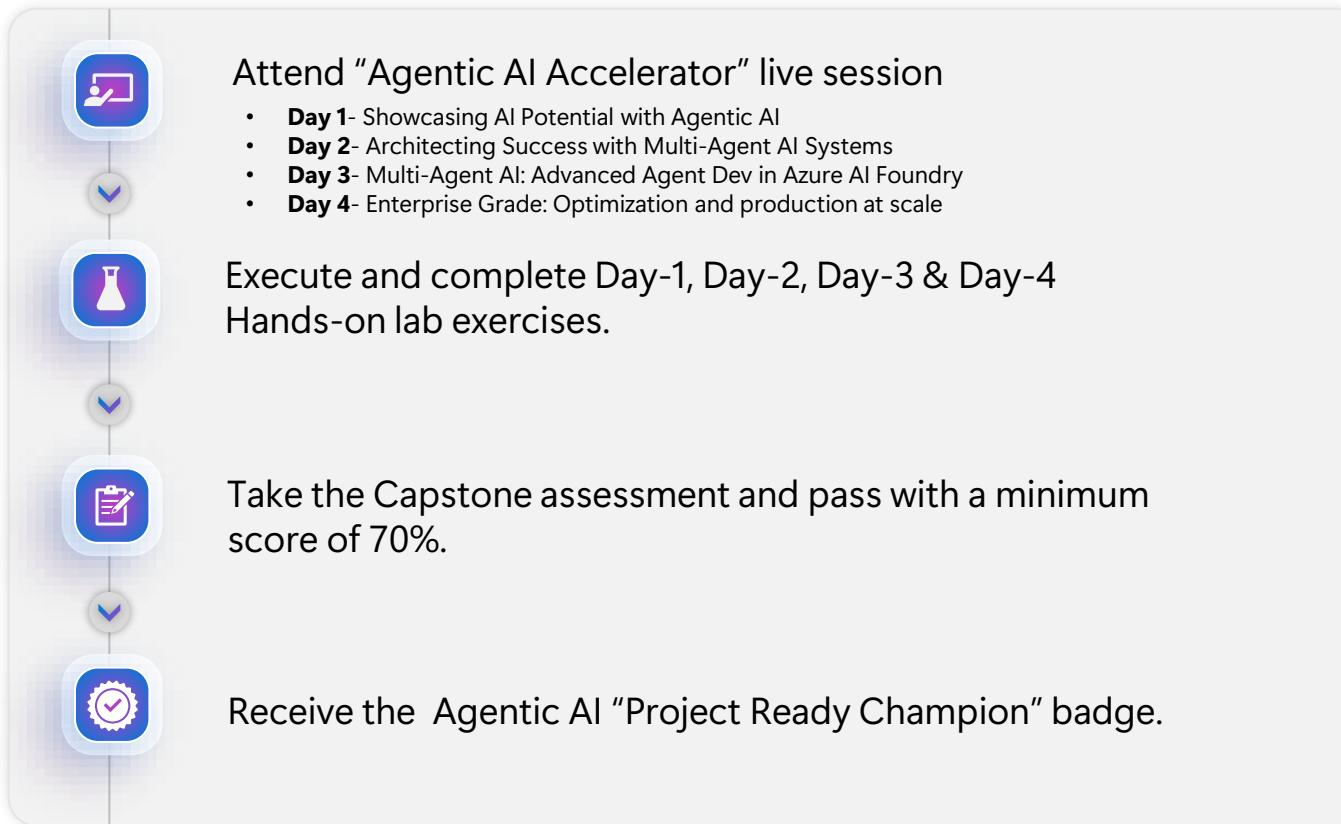
Receive the Agentic AI Project Ready completion badge \*

\*badges will be awarded if the qualification criteria is met



# Get Agentic AI “Project Ready Champion” badge

## Badge Qualification criteria



Receive the Agentic AI “Project Ready Champion” badge by **30<sup>th</sup> June 2025**

# Continue your learning deal-readiness journey with more ...



Microsoft Partner Project Ready

## Azure OpenAI Workshop

Duration 4 days (4 hrs/day)



Microsoft Partner Project Ready

## Build and modernize AI Apps on Azure

Duration 4 days (4hrs/day)



Explore and Register on LevelUp  
[skillupwithlevelup.com](http://skillupwithlevelup.com)

# Enrich your Agentic AI development skills



Register now  
[aka.ms/AgenticAIHackathon](https://aka.ms/AgenticAIHackathon)



Participate in the **live sessions** of this Hackathon on **both days** and **complete 2 of 4 projects** with a **minimum of 70% score** in **each project**.



Receive a Digital Badge as a testament to your engagement and learning.

Register now  
[aka.ms/AgenticAIHackathon](https://aka.ms/AgenticAIHackathon)



## Agentic AI Hackathon

June 10, 2025 – June 11, 2025



### Welcome to Agentic AI Hackathon

The Agentic AI Hackathon is a hands-on, immersive learning experience focused on building intelligent multi-agent AI applications leveraging Azure AI Foundry, Azure OpenAI services, AutoGen Framework, Semantic Kernel, and GitHub. Participants will gain practical expertise in developing and deploying AI-powered solutions, mastering skills such as AI agent customization, orchestration, and automation. This workshop equips developers with real-world capabilities to innovate and accelerate AI adoption using Azure AI and GitHub tools.

#### Agentic AI Hackathon

⌚ Duration - 2 Days (5 hrs/day)

📅 Date - June 10-11, 2025

🎯 Target Audience - Developers

Workload	Summary	Badge
Opening & Orientation	Introduction about the event	
Choice of Hackathon Projects	Build and Deploy Azure AI Agents Build an End-To-End Multi-Agents Enable Multi-Agent Interactions Using Agentic Framework Capabilities Build a Multi-Agent Presentation Builder	
Final show & Discussion	Closing discussion and Q&A	



# Thank You