



#Milano



# Build AI with the latest version of Microsoft Foundry

**Codrina Merigo**

Digital Product Manager @ Fresenius Medical Care



#Milano

improve



TD SYNnex

Grazie ai nostri sponsor 🙏



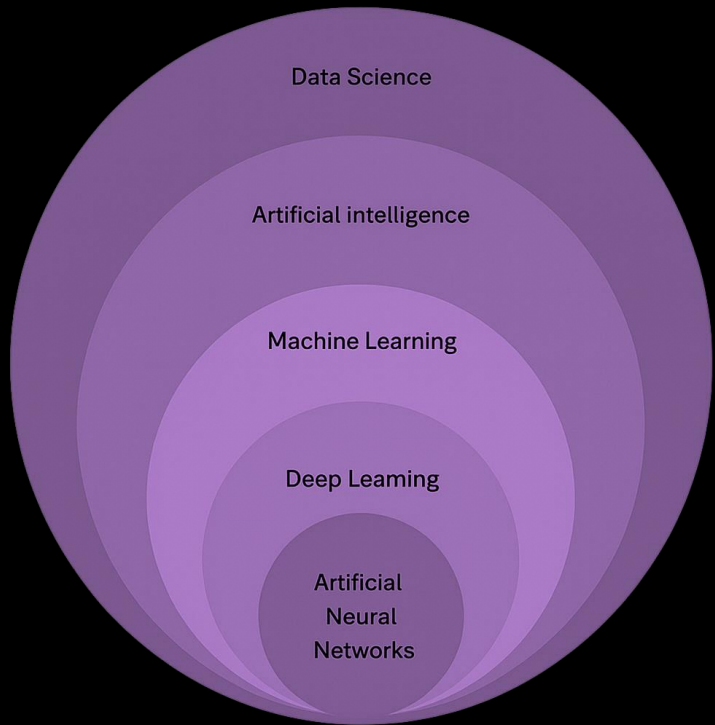
#Milano

**Slide e video:**  
**<https://www.globalazuremilano.it>**

AI IS TRANSFORMING OUR WORLD

The background features a complex network of glowing nodes and connecting lines in shades of white and light blue. The nodes vary in size and brightness, creating a sense of depth and connectivity. The lines are thin and light blue, forming a web-like structure that spans the right side of the image. The overall aesthetic is futuristic and technological.

# AI INTRODUCTION



**AI** or **Artificial Intelligence** is a subfield within computer science associated with constructing machines that can simulate human intelligence.

An **AI model** is a program or algorithm that utilizes a set of data that enables it to recognize certain patterns. This allows it to reach a conclusion or make a prediction when provided with sufficient information.



VECTOR DB



PIPELINES/ORCHESTRATION



DATA CHUNKING



INFRASTRUCTURE

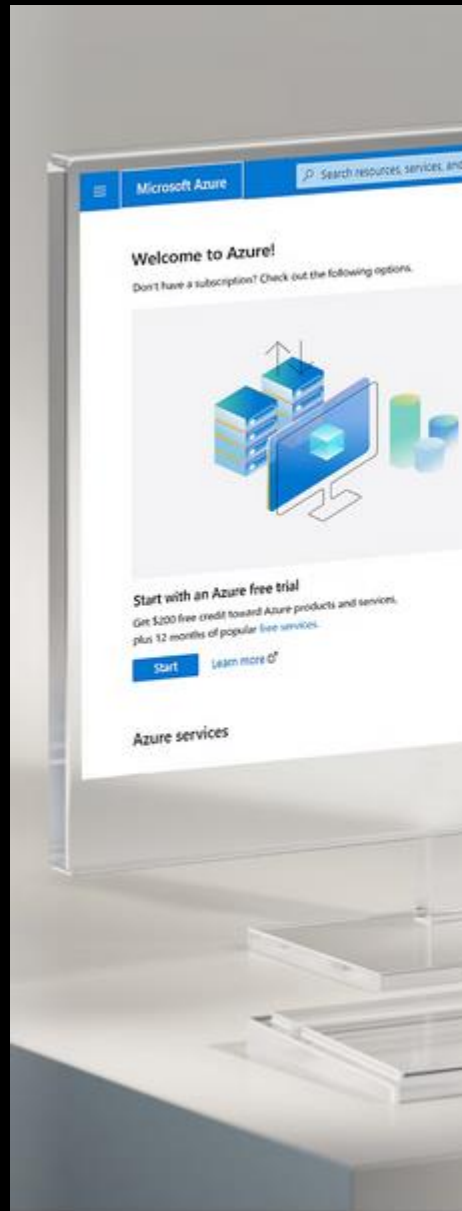


CUSTOM  
AI



# MICROSOFT AZURE

Azure is a trusted cloud platform for building, deploying, and managing innovative solutions.



AI-powered assistant  
Created with Copilot Studio

Azure is a trusted cloud platform for building, deploying, and managing innovative solutions. Launched in 2010, Azure marked a pivotal shift from on-premises datacenters to cloud computing. It offers more than 200 products and services across a global network of datacenters, providing businesses with extensive capabilities that go beyond simplifying infrastructure management. Azure supports AI, data, and application services, delivering a unified approach to cloud computing. It is designed to support each company's business strategy and stage of AI transformation. Azure is highly secure, with built-in security solutions and a dedicated team of security experts monitoring systems continuously [1].

[1]: [What is Azure? | Microsoft Azure](#)

Related:

[Try Azure for free](#)

Type a follow up message or select one below.

[Unlock potential with Azure cloud solutions.](#)

[Scale your business using Azure services.](#)

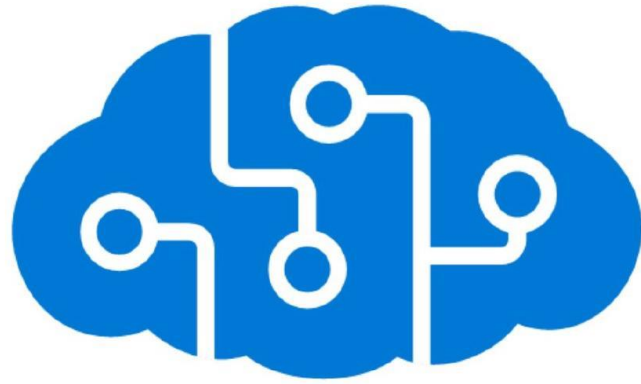
[Secure your data with Azure technology.](#)

Just now

AI-generated content may be incorrect

Ask a work question or make a request...





# AZURE COGNITIVE SERVICES



## Vision

From faces to feelings, allow your apps to understand images and video



## Speech

Hear and speak to your users by filtering noise, identifying speakers, and understanding intent



## Language

Process text and learn how to recognize what users want



## Knowledge

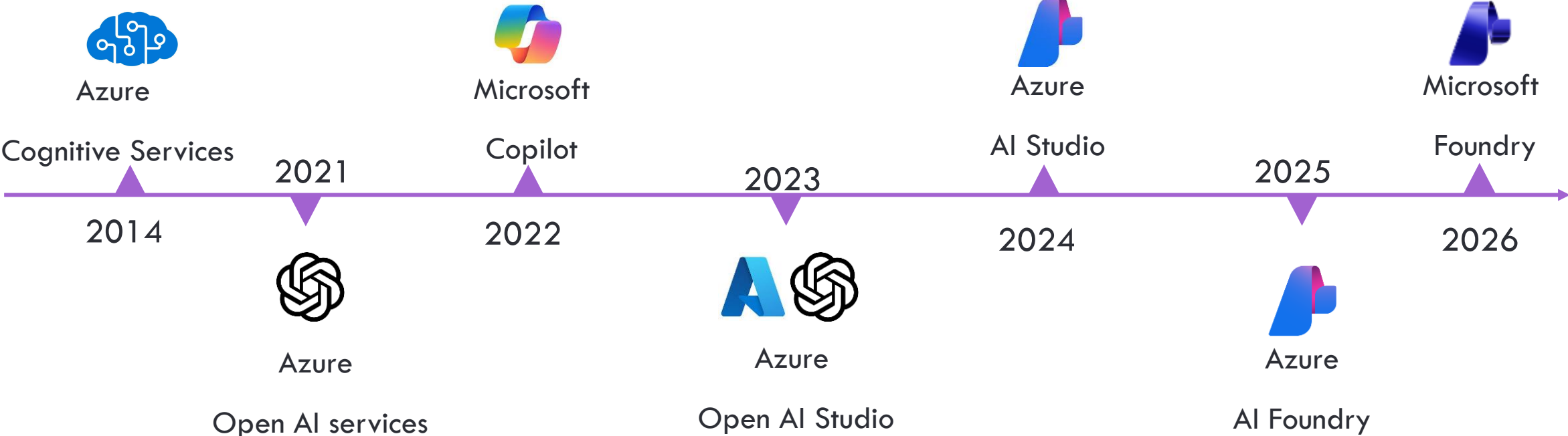
Tap into rich knowledge amassed from the web, academia, or your own data



## Search

Access billions of web pages, images, videos, and news with the power of Bing APIs

# EVOLUTION OF AZURE AI SERVICES



**Microsoft Foundry**  
**empowers you to be the**  
**FRONTIER ORGANIZATION**

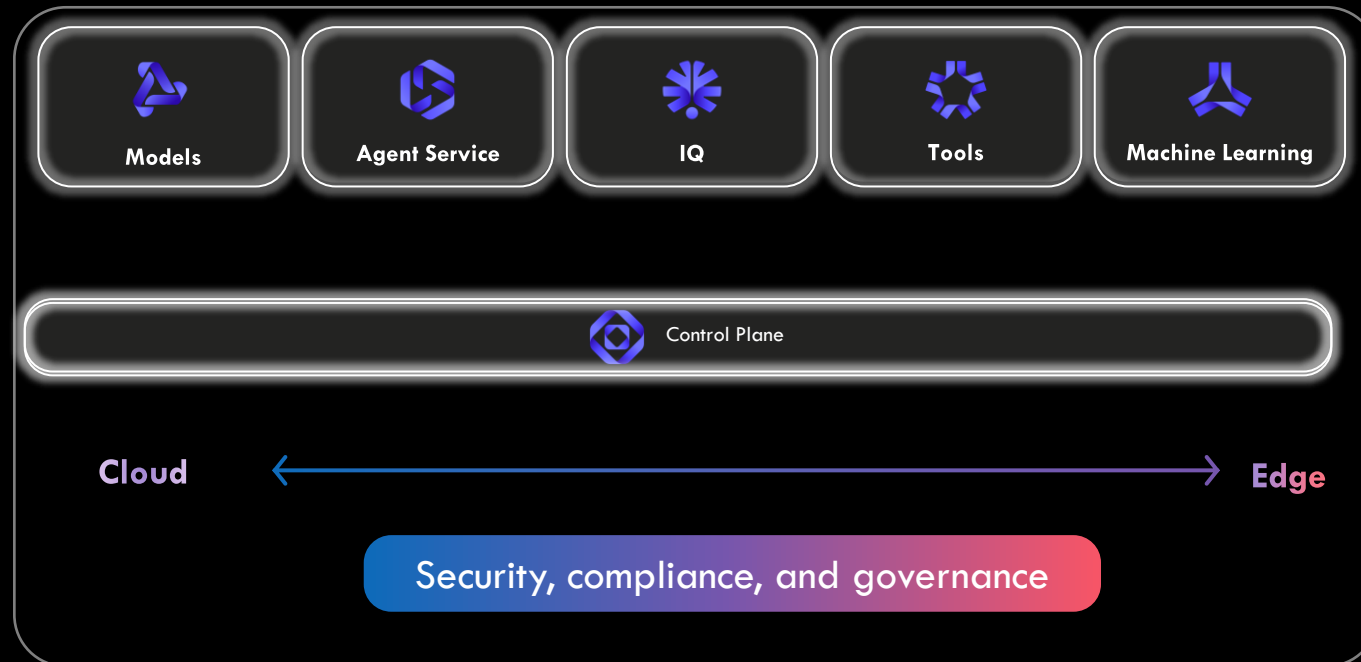


MICROSOFT FOUNDRY



# MICROSOFT FOUNDRY

## The AI app and agent factory



# MICROSOFT FOUNDRY

Streamline development with native IDE experiences



GitHub



Visual Studio



Visual Studio Code



Copilot Studio



Models



Agent Service



IQ

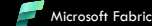


Tools



Machine Learning

Build context-aware and action-oriented agents with 1,400+ pre-built connections and MCP tools



Control Plane

Leverage a complete signals management layer with Microsoft Security integrations



Microsoft Agent 365



Microsoft Defender



Microsoft Purview



Microsoft Entra

Security, compliance, and governance



# MICROSOFT FOUNDRY

## The full enterprise package

Model policy

99.9% uptime

99% latency SLA for provisioned

Data Zone for US and EU

Off peak capacity @ 50% for models

Priority processing for Azure OpenAI

Model router

Content safety by default for models

RAG through Azure AI Search

VNET

# AGENTIC AI

AI agents can plan, adapt to new information, and execute tasks independently, making them capable of handling complex, dynamic environments, decision-making and autonomous action. They can learn from feedback, adjust their strategies, and operate with minimal human oversight.



Reason over a provided business process

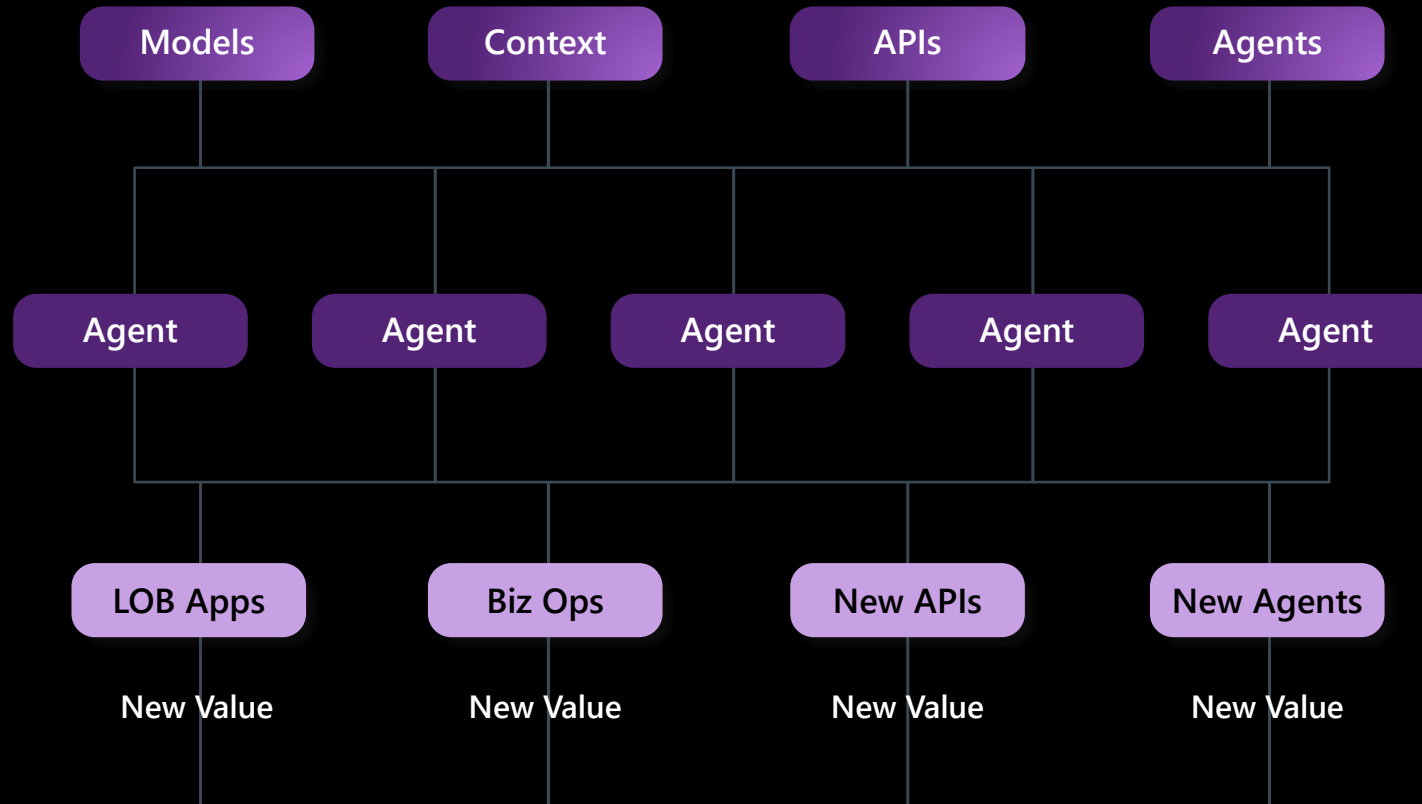


Retrieve context to complete the process

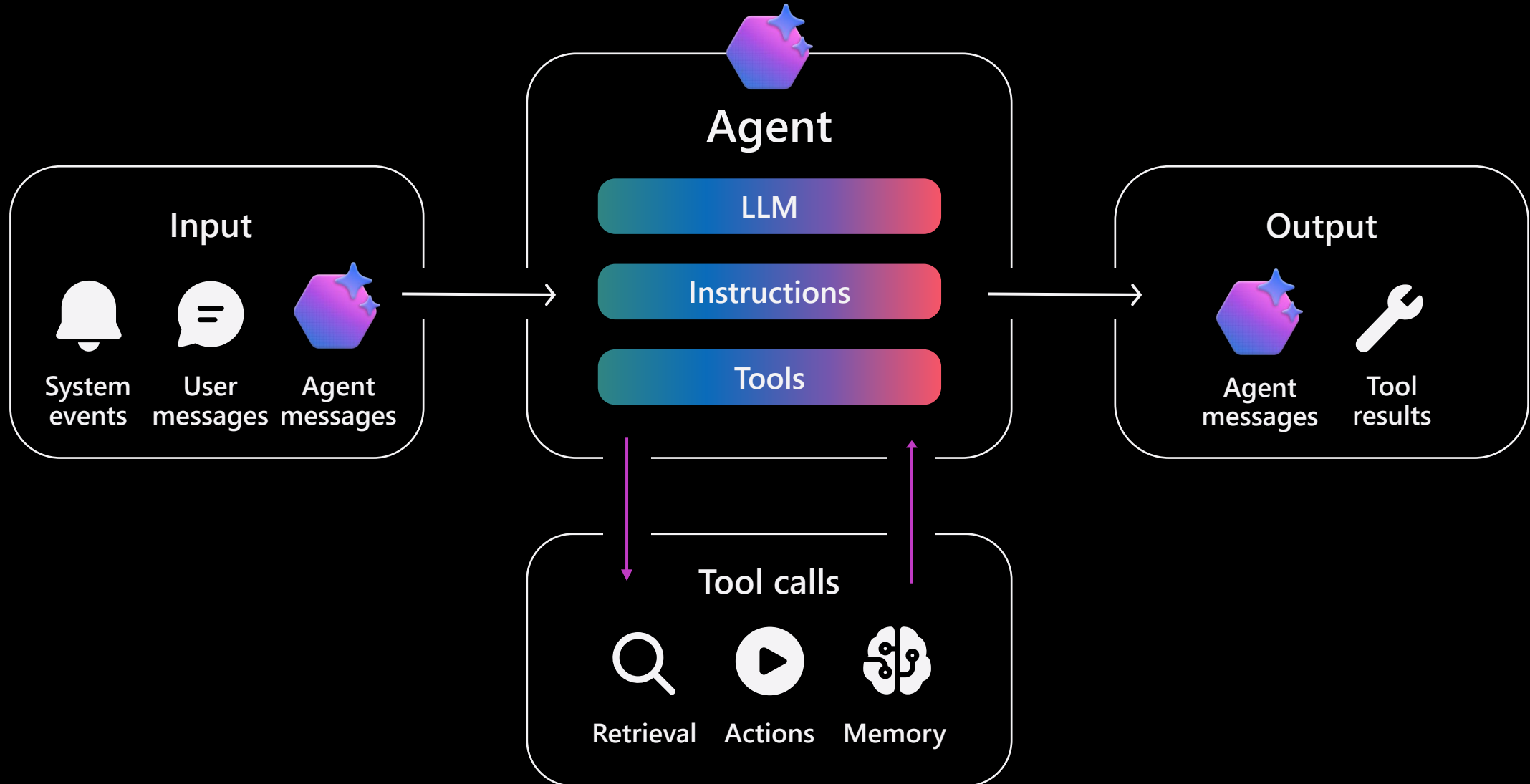


Perform an action for the end-user

# Foundry is "agent-first"

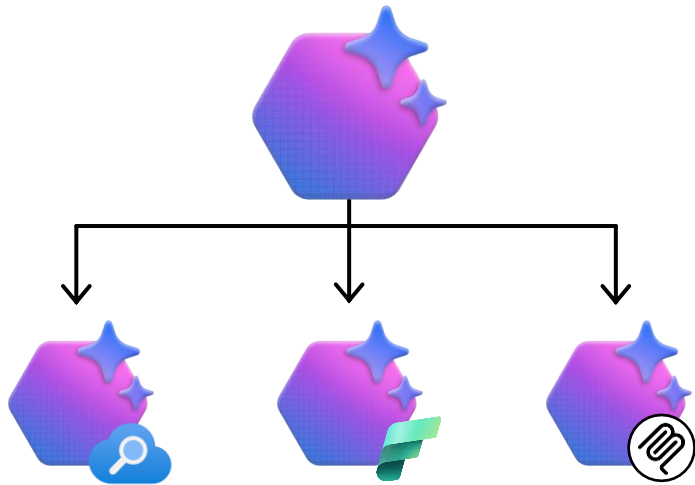


# What is an agent?



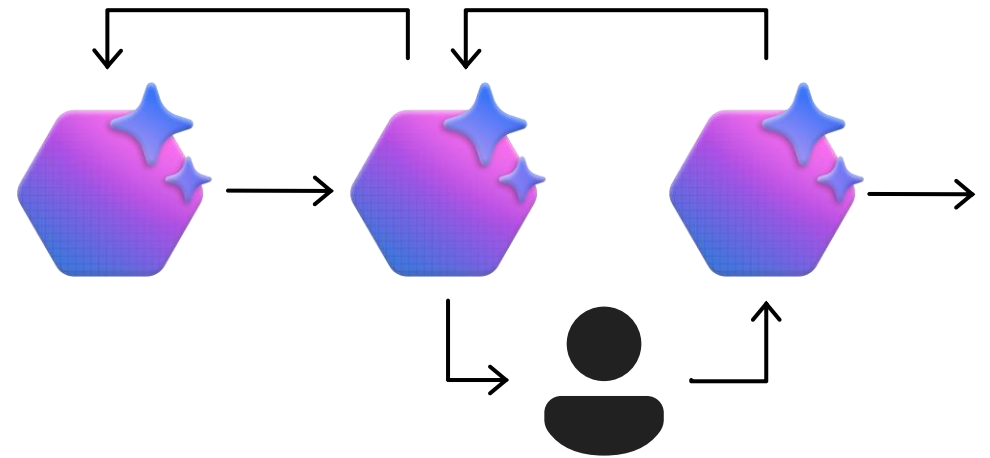
# Evolving into Multi-Agent Systems

## Agent orchestration



LLMs dynamically decide the control flow and direct their own processes and tool usage, maintaining control over how they accomplish tasks

## Workflow orchestration



You decide the control flow orchestrating LLMs, Agents, tools and more via predefined code paths

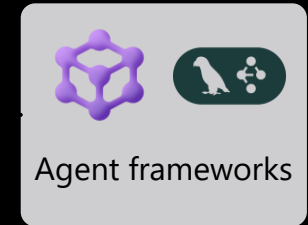
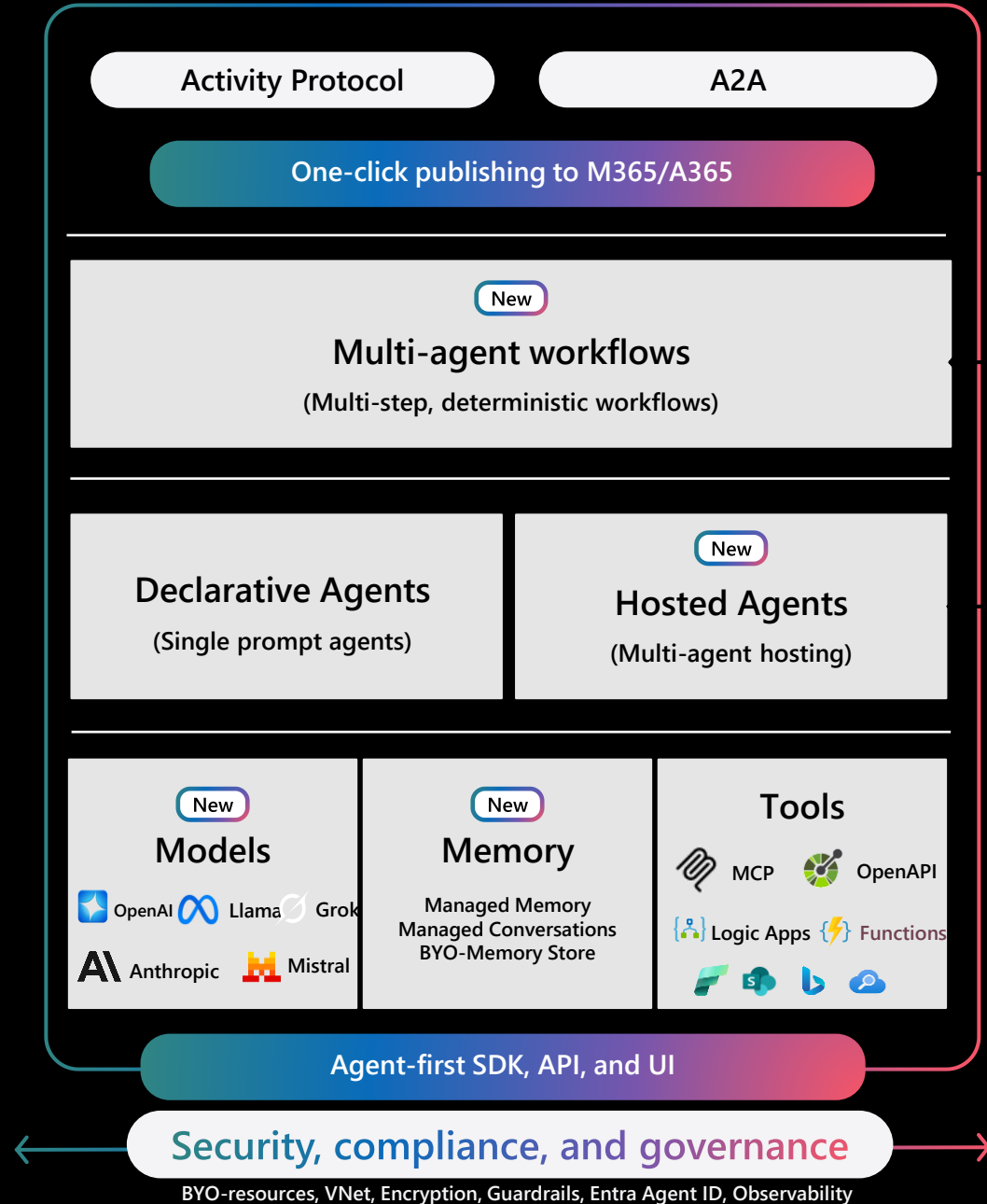


# Foundry Agent Service

Open by Design

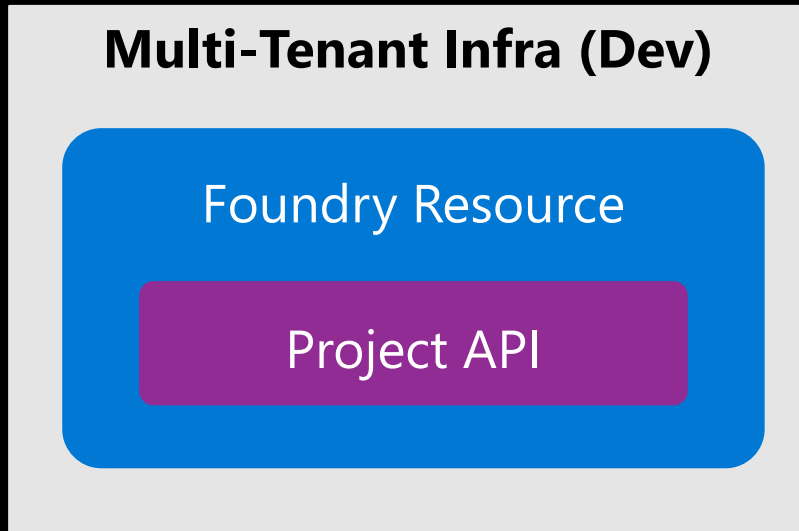
Connected Intelligence

Enterprise Security & Reach

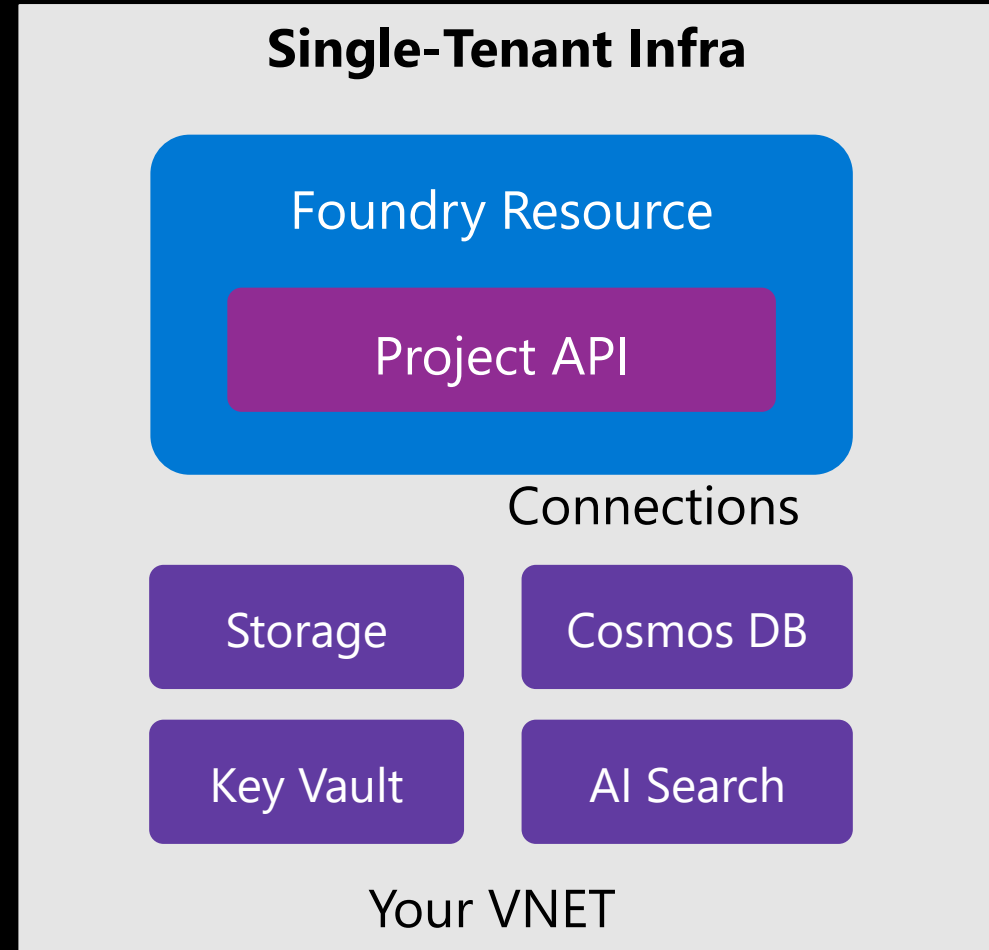


Python  
C#  
TypeScript  
Java

# Start quick in dev enterprise-grade in production



→  
No Code  
Changes



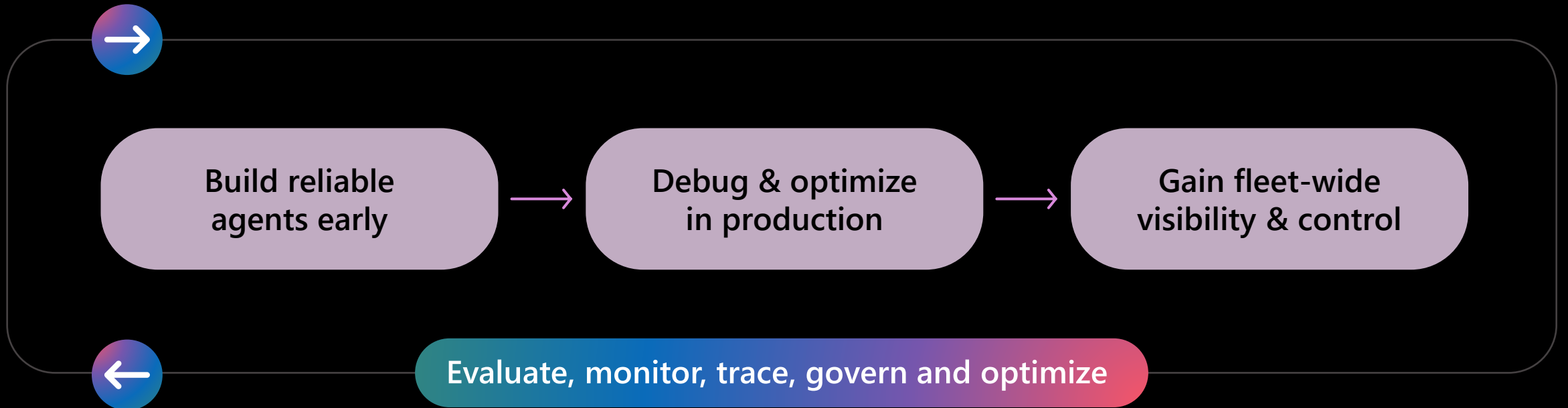
Project data:

- Agent definitions
- Conversation state
- Workflow checkpoints
- Vector stores
- Uploaded & generated files

Public Preview

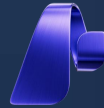
# Observability in Foundry Control Plane

Visibility, monitoring, and optimization across the AI agent lifecycle



# Observability in Foundry Control Plane: Key Capabilities

| Generally Available  | Public Preview Capabilities                                       |
|--|---|
| Evaluators (wide range for models + datasets), ability to run in cloud | Tracing, new evaluators, human evaluation, cluster analysis       |
|  | Monitoring, alerts, optimization capabilities like model upgrades |
|  | New agentic safety risks in AI Red Teaming Agent                  |



# MICROSOFT FOUNDRY

Simple

Trusted

Powerful



# MICROSOFT FOUNDRY

Simple

Trusted

Powerful

Keep building with Microsoft Foundry

View all resources

+ Create new

? Help

Resource name ▾

ai-codrina164234ai25062-project

projectUpdateC

apFoundry-project

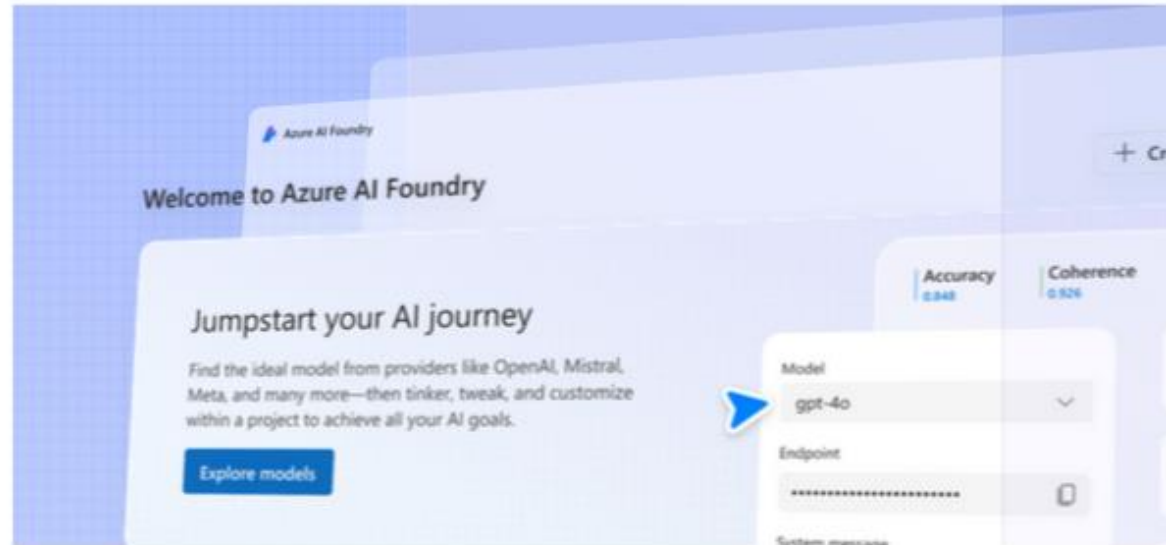
Created on ▾

Nov 4, 2025 4:40 PM

Nov 5, 2025 4:00 PM

Nov 7, 2025 9:18 AM

Explore models and ca



## Setting up Microsoft Foundry for you

You're almost ready to create and deploy AI solutions. We'll take you where you need to go once everything is ready.

Setting up your project, this may take a few minutes

Create smart  
Microsoft Fo

Design, customize, and manage powerful, adaptable AI

You are a helpful chatbot...

Here is the weather for today in the three biggest markets:

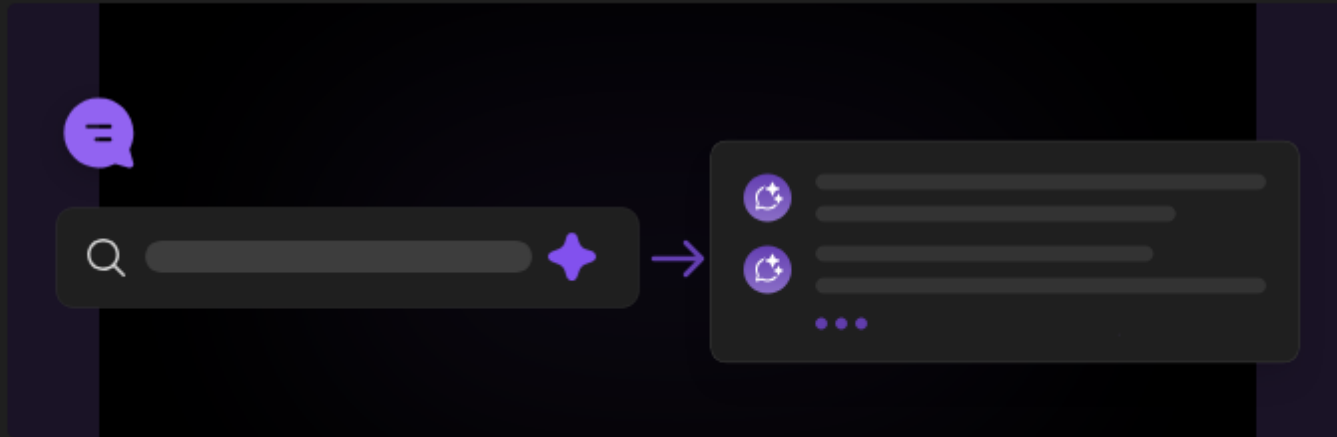
Get me the weather in the 3 biggest markets and

AI-generated content may be incorrect

# Welcome, Codrina .


Start building ▾

## Welcome to the new Microsoft Foundry



**Ask AI anything, anytime**

Use Search or Ask AI to find what you need in your project

 Ask AI or **Ctrl + K** to start search



Back

Create agent

Preview

# FOUNDRY IQ

## Knowledge for agents powered by Azure AI Search

The screenshot displays the Microsoft Foundry interface for managing knowledge. The main page is titled "Knowledge (0)" and includes a search bar and a "Connect knowledge" button. A sidebar on the left lists navigation options: Agents, Workflows, Models, Fine-tune, Tools, Knowledge, Data, Evaluations, and Guardrails. The "Knowledge" section is active, showing "Knowledge bases" and "Vector stores" tabs. Below these, there are search suggestions for "Ask AI".

A modal dialog titled "Choose a knowledge type" is open, providing options for configuring a knowledge base or via tools. The "Configure a knowledge base" section includes:

- Azure Blob Storage**: Enhance model output with web data.
- Azure AI Search index**: Enterprise scale search for app development.
- Grounding with Bing Search**: Enhance model output with web data.
- Grounding with Bing Custom**: Enhance model output with data from selected domains.
- SharePoint**: Securely integrate and manage internal data.

The "Configure knowledge via Tools" section includes:

- Azure Cosmos DB**: Enables Agents to interact with... [knowledge tool]
- Azure PostgreSQL**: Scalable, and secure relational... [knowledge tool]
- Azure SQL**: Relational database service bui... [knowledge tool]
- MongoDB**: MongoDB MCP Server allows a... [knowledge tool]
- Azure Managed Redis**: Fully managed, enterprise-grad... [knowledge tool]
- Databricks**: Databricks offers a unified platf... [knowledge tool]
- Elasticsearch**: Search, retrieve, and analyze...
- Microsoft Fabric**: End-to-end data and analytics...
- Pinecone**: Pinecone makes it easy to provi...

At the bottom of the dialog are "Configure" and "Cancel" buttons.

Preview

# MEMORY IN FOUNDRY AGENT SERVICE

Context-aware, personalized, and interactions

The screenshot displays the configuration interface for the Memory feature in Foundry Agent Service. On the left, a sidebar contains three expandable sections: Tools, Knowledge, and Memory. The Memory section is expanded, showing a description: "Memory retains context to personalize and improve interactions. A memory store will be auto-created for you. It will use the model you have deployed for this agent. [Learn more about cost](#)". Below this, there is a toggle switch labeled "On" and a dropdown menu for "Memory store" currently set to "Auto-create memory store". A note states: "Note: The agent will be able to add-to and modify this store." At the bottom, the "Scope" is set to "Signed-in user". On the right, a terminal window shows the command "1 | [project-name]-knowledge-base" and the output "25t | 0.5s | AI quality: 100% | Safety: 100%".

Preview



# FOUNDRY TOOLS

Agentic power, delivered as tools

The screenshot displays the Microsoft Foundry Tools interface. At the top, there is a search bar with the text "Search with AI (Ctrl + K)" and a refresh icon. The navigation bar includes "New Foundry" (with a toggle switch), "Home", "Discover", "Build", "Operate", "Docs", and notification icons. The left sidebar contains navigation options: "Overview", "Models", "Tools" (highlighted), and "Solution templates".

The main content area is titled "Tools (1404)" and features a search input and a "Sort by Featured" dropdown. Below this, there are filter sections for "Type", "Provider", and "Category".

**Type filters:**

- Custom: 1362
- MCP: Remote: 37
- MCP: Local: 5

**Provider filters:**

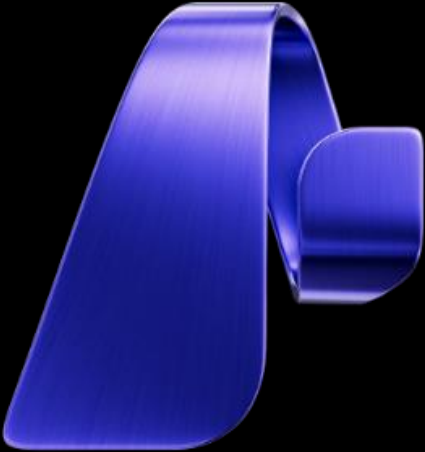
- Partner: 31
- Microsoft: 10

**Category filters:**

- Developer Tools: 7
- Databases: 6
- Analytics: 5
- Customer Service: 4
- Productivity: 4
- Web: 4
- Machine Learning: 3
- Finance: 2

The tool cards are arranged in a grid. Each card includes an icon, the tool name, a brief description, and a label indicating its MCP type (Remote or Local). The visible tools include:

- Amplitude MCP Server:** Search, access, and get insights on your... (MCP: Remote)
- Atlan:** The Atlan MCP server provides a set of tools th... (MCP: Remote)
- Atlassian:** Connect to Jira and Confluence for issue... (MCP: Remote)
- Azure AI Foundry Remote MCP server:** Connect to this MCP server to use Foundry... (MCP: Remote)
- Azure Cosmos DB:** Enables Agents to interact with and retrieve... (MCP: Local)
- Azure Database for PostgreSQL:** Enables Agents to interact with and retrieve... (MCP: Local)
- Azure Databricks Genie:** Azure Databricks Genie MCP server lets AI... (MCP: Remote)
- Azure Language in Foundry Tools:** The MCP server enables AI agents to access... (MCP: Remote)
- Azure Managed Redis:** Azure Managed Redis MCP Server provides a... (MCP: Local)
- Azure Speech MCP Server:** A hosted MCP server that exposes Azure... (MCP: Remote)
- Azure SQL MCP Server:** A secure, self-hosted MCP for interacting with... (MCP: Local)
- bria.ai:** Equip AI agents with Bria.ai image generation... (MCP: Remote)
- CAST Imaging MCP Server:** Deterministic mapping of application... (MCP: Remote)
- Celonis PI Graph MCP Server:** Agent toolkit that provides process intelligenc... (MCP: Remote)
- Elasticsearch:** Search, retrieve, and analyze Elasticsearch da... (MCP: Remote)
- Exa Web Search:** Exa MCP is a powerful web search and web... (MCP: Remote)
- Factory RCA MCP:** Toolset for manufacturing root-cause analysis... (MCP: Remote)
- GitHub:** Access GitHub repositories, issues, and pull... (MCP: Remote)
- Hugging Face MCP Server:** Search through millions of Hugging Face... (MCP: Remote)
- Infobip RCS MCP server:** Infobip RCS MCP server enables seamless... (MCP: Remote)
- Infobip SMS MCP server:** The Infobip SMS MCP server enables agentic... (MCP: Remote)



DEMO



# MICROSOFT FOUNDRY

Simple

Trusted

Powerful

# RESPONSIBLE AI PRINCIPLES



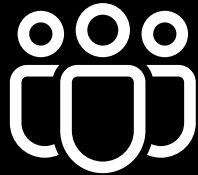
*Fairness*



*Reliability and safety*



*Privacy and security*



*Inclusiveness*



*Transparency*



*Accountability*

# BUILDING TRUSTWORTHY AI IS A TEAM SPORT



IT



Developer



Security

**Controls**

**Observability**

**Security**

**Fleet-wide  
Operations**

Preview



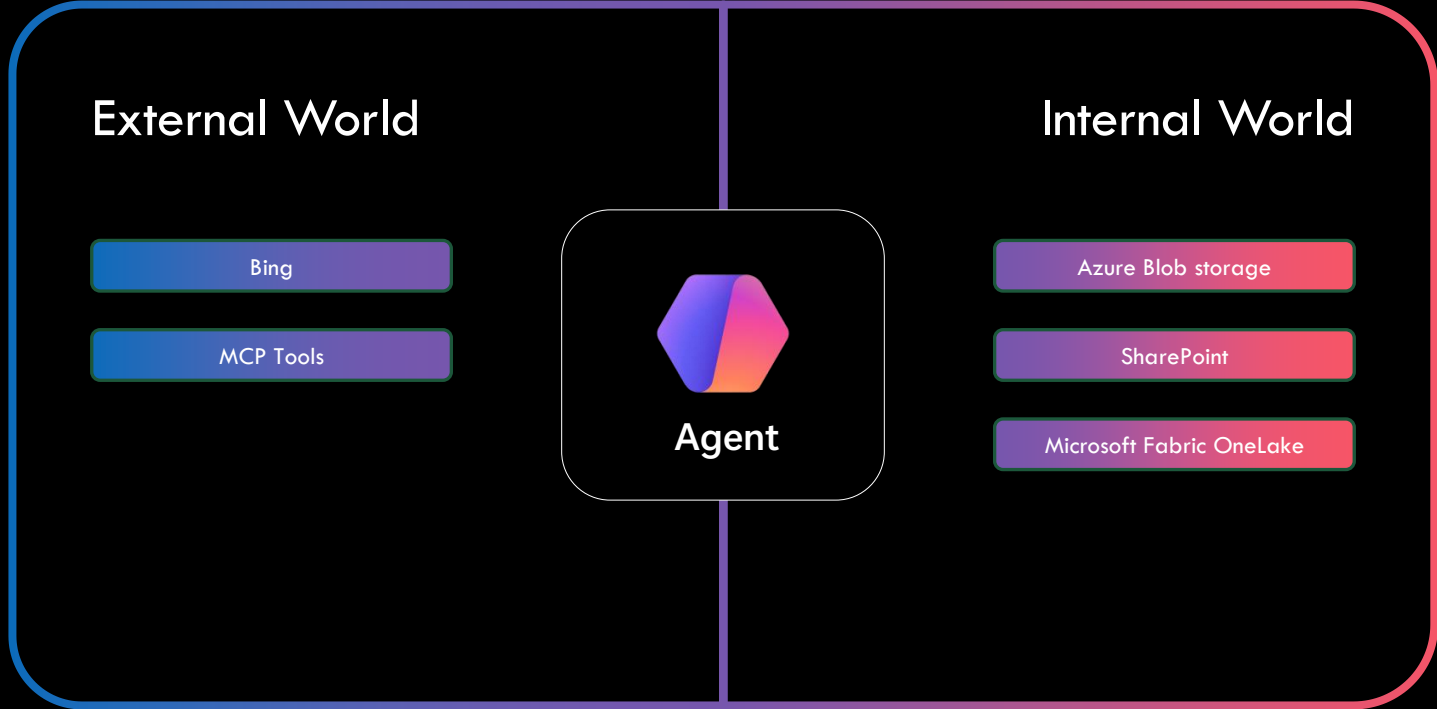
Foundry Control Plane

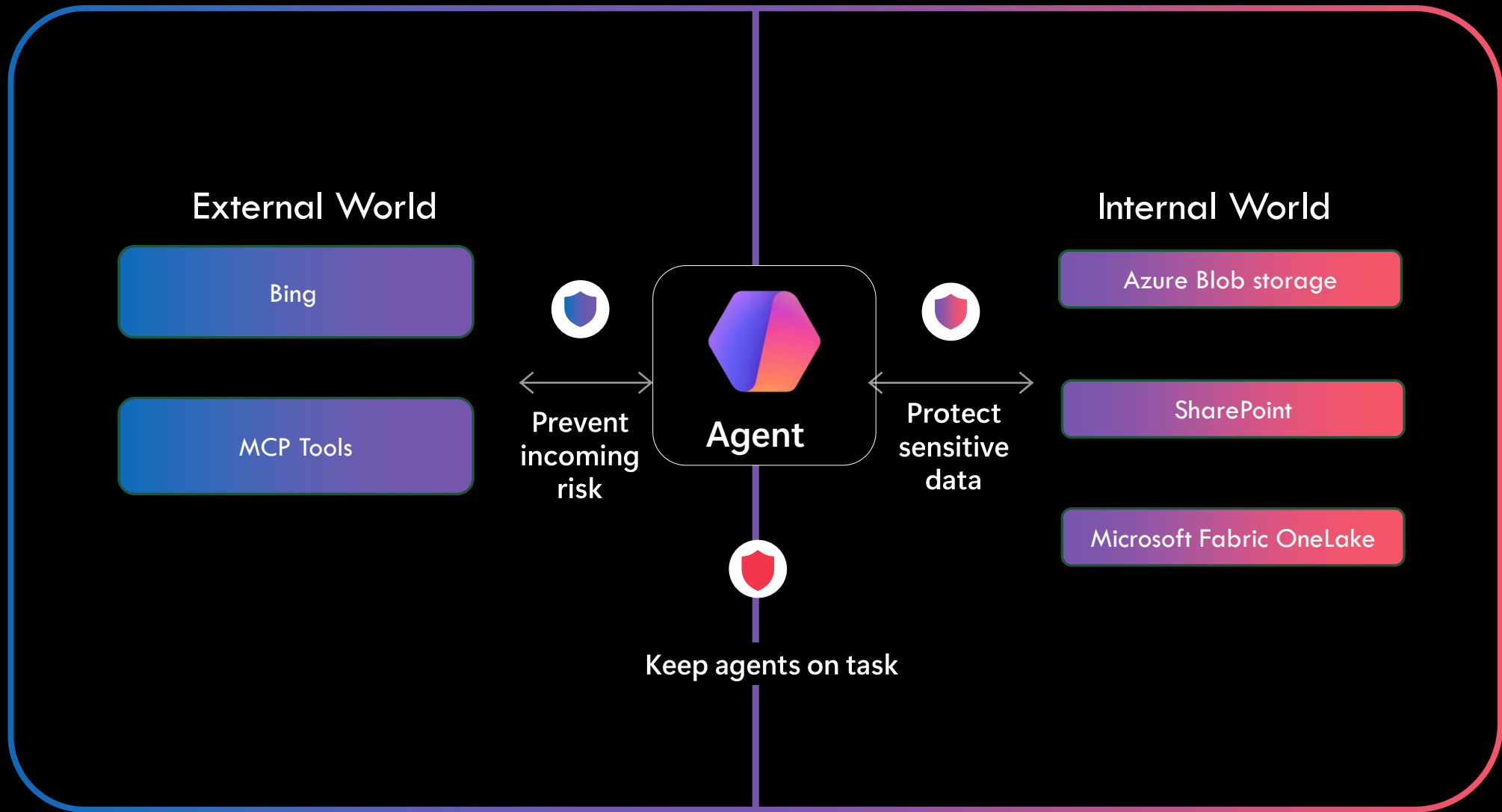
Controls

Observability

Security

Fleet-wide  
Operations





# CONTROLS

Keep agents safe and grounded

Azure AI Foundry / project-name123 Search with AI K Home Discover Build Operate Docs Profile

## ← Create guardrail

Add guardrails to reduce safety and security risks. Create, manage and apply them across models or agents for a safer user experience. [Learn more](#)

### 1 Add controls

Add individual controls to your guardrail. Controls define what to detect, where to detect it, and how to respond to flagged behavior. [Learn more](#)

Risk type \* ⓘ  
Spotlighting (Preview) ▼

Intervention point \* ⓘ

- User input Recommended
- Tool call (Preview)
- Tool response (Preview) Recommended
- Output

Action \* ⓘ

- On Recommended
- Off

[Add control](#)

[Back](#) [Next](#)

### Controls

Review your current configuration below, including both default and custom controls. Make edits by adding a new control.

▼ Content safety

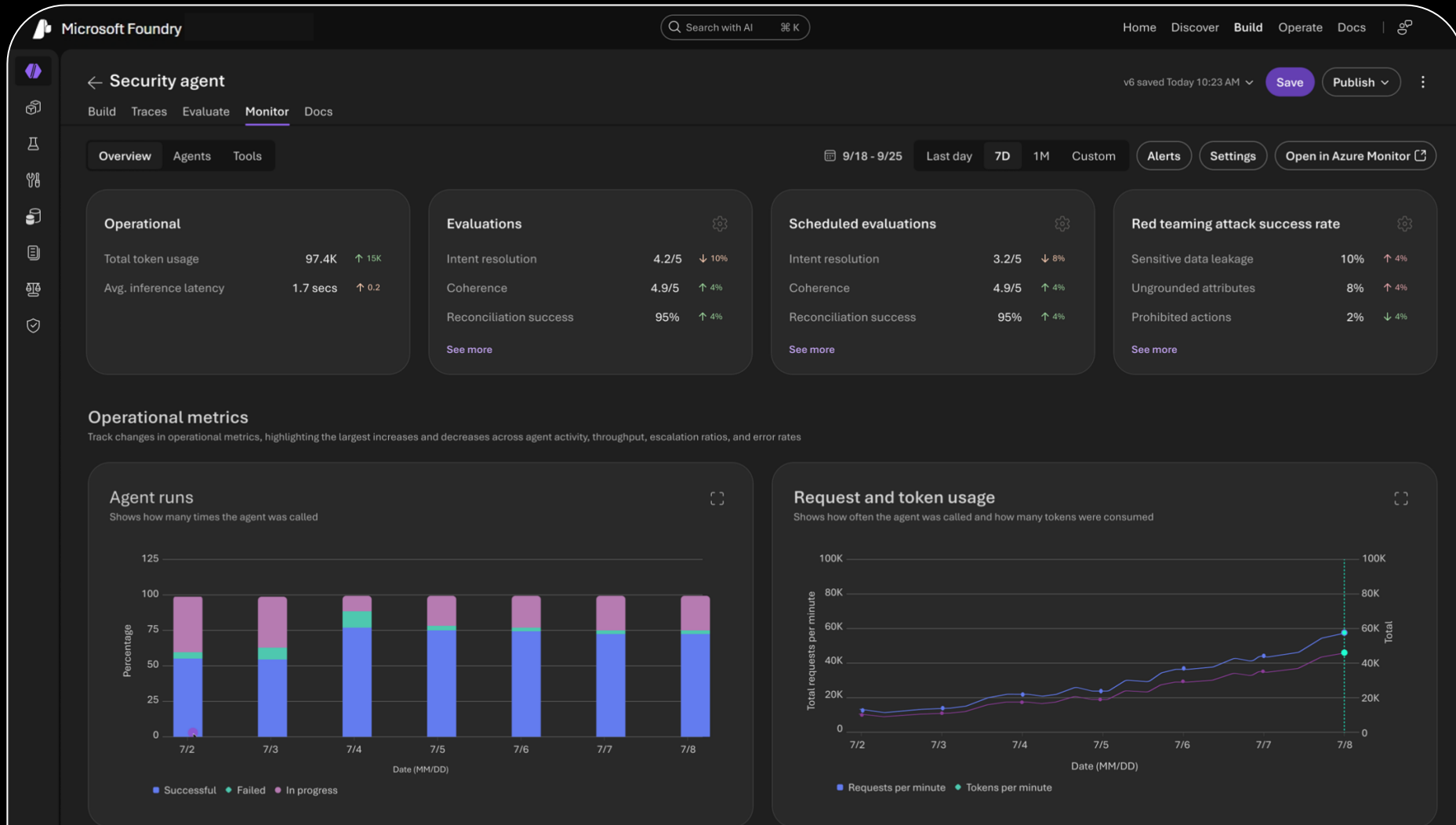
| <input type="checkbox"/> Risk type                  | Intervention point | Action             |
|---|--------------------|--------------------|
| <input type="checkbox"/> Hate: Lowest blocking      | User input, Output | Annotate and block |
| <input type="checkbox"/> Self-harm: Lowest blocking | User input, Output | Annotate and block |
| <input type="checkbox"/> Sexual: Lowest blocking    | User input, Output | Annotate and block |
| <input type="checkbox"/> Violence: Lowest blocking  | User input, Output | Annotate and block |

▼ Protected materials

| <input type="checkbox"/> Risk type      | Intervention point | Action             |
|---|--------------------|--------------------|
| <input type="checkbox"/> Protected code | User input, Output | Annotate only      |
| <input type="checkbox"/> Protected text | User input, Output | Annotate and block |

# OBSERVABILITY

Evaluate and monitor agents for highest quality before and after deployment



# SECURITY

Integrated with Microsoft Security, protecting every agent at every step

The screenshot displays the Azure AI Foundry interface. At the top, the navigation bar includes 'Azure AI Foundry', a search bar, and links for 'Home', 'Discover', 'Build', 'Operate', and 'Docs'. The main content area is titled 'Compliance' and has tabs for 'Policies', 'Guardrails', and 'Security'. A 'Create policy' button is visible in the top right. Below the tabs, there are dropdown menus for 'Subscription: Contoso-subscription2025' and 'Project: mona-kane-9-18-2025'. The 'Microsoft Purview' section is active, showing a toggle switch and a description of its capabilities. Below that is the 'Microsoft Defender for Cloud' section, which includes a search bar and a table of recommendations. A pink arrow points to the first recommendation in the table. To the right, a detailed view of the selected recommendation is shown, including the affected resource, resource type, and risk level, along with a list of steps to resolve the issue.

**Subscription:** Contoso-subscription2025 **Project:** mona-kane-9-18-2025

### Microsoft Purview

Data security posture management

(Preview) Enable Microsoft Purview for your subscription to govern AI apps and agents with enterprise-grade data security, compliance, and governance. Once enabled, app interaction data will be sent to your tenant's Purview account, and policies created in Purview will be enforced in Foundry. Enablement requires the 'Cognitive Services Security Integration Admin' Azure role.

**Prerequisites:** Microsoft Purview in the tenant. Billing is based on Purview pay-as-you-go meters and policies created in Purview. If billing isn't set up, only Purview Audit will be supported. [Learn about Microsoft Purview](#)

### Microsoft Defender for Cloud

Security recommendations

Enable Microsoft Defender for Cloud to monitor your Azure resource configurations for potential security vulnerabilities and to suggest remediation actions. [Learn about Microsoft Defender](#)

Search recommendations

| Recommendation  | Affected resource | Resource type | Risk level |
|---|-------------------|---------------|------------|
| <input checked="" type="radio"/> Microsoft Entra ID should be used when connected to dat... | my-first-project  | CustomKeys    | High       |
| <input type="radio"/> Microsoft Entra ID should be used when connected to dat...            | my-first-project  | CustomKeys    | High       |
| <input type="radio"/> Microsoft Entra ID should be used when connected to dat...            | my-first-project  | CustomKeys    | High       |
| <input type="radio"/> Microsoft Entra ID should be used when connected to dat...            | my-first-project  | CustomKeys    | High       |

### Microsoft Entra ID should be used when connected to stores on Azure AI Foundry project.

High

[Take action in Azure Portal](#) [View docs](#)

**Affected resource**  
my-first-project

**Resource type**  
CustomKeys

**Risk level**  
High

### Recommendation

To configure identity-based access when connecting to data stores in AI Foundry project:

1. In the Azure Portal, navigate to your AI Foundry project.
2. From the menu, open the **Overview page**.
3. In the *Start building in Azure AI Foundry* section, click **Launch studio**.
4. From the Overview page in *Manage project settings*, click **Open in management center**.
5. From the *Connected resources* page, select the connected data resources in which the authentication type is Credentials based.
6. From Access details, click the **Edit** icon.
7. From the *Update custom shared connection* page, change the Authentication method to **Microsoft Entra ID based**.

# FLEET-WIDE OPERATIONS

One dashboard for every agent: health, performance, cost, compliance

Azure AI Foundry

Search with AI

Home Discover Build Operate Docs

## Overview

View in Ops365 Register assets

Subscription: Contoso-su... Project: mona-kane-9-18-... Asset type: All

9/18 - 9/25 Last day 7D 1M Custom

### 3 Active alerts

1 Critical 2 Warnings 9 Improvements

Sort by: Priority - High to low

| Alert  | Asset                | Category              | Action |
|--|----------------------|-----------------------|--------|
| <span>!</span> User access blocked due to a critical error—immediate at... | Entra ID             | Security: Access      | Review |
| <span>!</span> Agent at risk for decommission in current re...             | Agent: DataProcessor | Policy: Compliance    | Review |
| <span>!</span> Payment method expires soon                                 | Subscription         | Cost: Billing         | Update |
| <span>i</span> Cost-saving opportunity detected                            | Model: Model router  | Cost: Budget          | Review |
| <span>i</span> Performance optimization available                          | Model: GPT-5         | Performance: Respo... | Review |

View all alerts

### Active vs. inactive ag...

9/9 agents

View all agents

### Agents in compliance

8/9 agents

One or more agents...

View all policies

### Behaviors prevented

47

↑ 2.5%

Compared to previous 7 days

View Guardrails

### Cost

3,202 USD

Models Tools Agents

↑ 2.5%

Compared to previous 7 days

### Agent runs

2,847

↑ 2.5%

Compared to previous 7 days

### Agent success rate

87.3 %

↑ 2.5%

Compared to previous 7 days

### Agent run volume over time

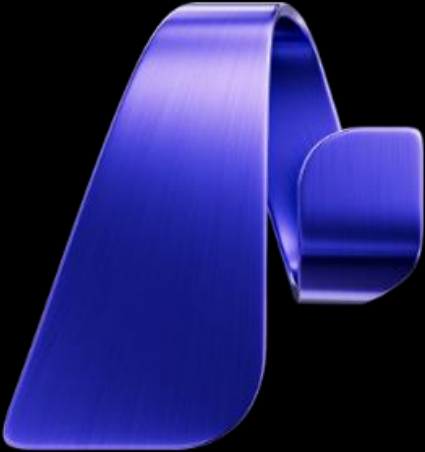
Track how often agents are triggered throughout the week

| Runs     |
|----------|
| 264 runs |

### Agent run volume week-over-week

Track the top increases and decreases in agent activity from the previous week

| Top increases   |                     |        | Top decreases |                     |       |
|-----------------|---------------------|--------|---------------|---------------------|-------|
| Agent           | Change (Agent runs) |        | Agent         | Change (Agent runs) |       |
| 1 DataProcessor | 180 → 435           | ↑ 142% | 1 LegacySync  | 220 → 121           | ↓ 45% |



DEMO



# MICROSOFT FOUNDRY

Simple

Trusted

Powerful

1

Models

2

Hosted  
Agents

3

Developer  
stack & tools

1

Models

2

Hosted  
Agents

3

Developer  
stack & tools



# FOUNDRY MODELS

core42  
A G42 company

gretel

SDAIA  
Sofisticated Data & AI Academy

NTT DATA

Azure OpenAI

Cohere

Databricks

DeepSeek

Hugging Face

11,000+ models

NVIDIA

Phi

Snowflake

XAI

Black Forest Labs

stability.ai

BRIA

d. Deci

NIXTLA

# INNOVATE FASTER AND SCALE RELIABLY WITH AZURE OPENAI

Latest innovations with same-day access

Connect to your *Microsoft Data Estate*

Content safety is guardrails

Flexible offerings

Enterprise Azure promises

Your data is private and secure

Developer-friendly integration

Model choice

# MODEL ROUTER

Real-time routing, cost control, and built-in governance

Microsoft Foundry

Search with AI (Ctrl + K)

New Foundry Home Discover Build Operate Docs

Overview Models Tools Solution templates

## model-router

Deploy

Details Deployments Benchmarks Responsible AI License

Model router is a deployable AI model that is trained to select the most suitable large language model (LLM) for a given prompt.

Microsoft Version:

Welcome to the version of model router that supports **gpt-5** models and three new models: **Llama-3.3-70B-Instruct**, **gpt-oss-120b**, and **Llama-4-Maverick-17B-128E-Instruct-FP8** in addition to the previously supported **gpt-4** and **o4-mini** models.

Model Router dynamically selects the optimal large language model (LLM) for a specific query or task in real time. By evaluating factors like query complexity, cost, and performance, it efficiently routes requests to the most suitable model, ensuring high quality results while minimizing costs. Context length for model router is dependent on the underlying model that's being used for each prompt.

For more information, reference the [Model router documentation](#)

### Model version in the router

| Router Version | Model        | Model version | Availability    | Lifecycle         |
|----------------|--------------|---------------|-----------------|-------------------|
| xxxx-xx-xx     | gpt-5*       | xxxx-xx-xx    | Global standard | General available |
|                | gpt-5-mini   | xxxx-xx-xx    | Global standard | General available |
|                | gpt-5-nano   | xxxx-xx-xx    | Global standard | General available |
|                | gpt-5-chat   | xxxx-xx-xx    | Global standard | Preview           |
|                | gpt-4.1      | xxxx-xx-xx    | Global standard | General available |
|                | gpt-4.1-mini | xxxx-xx-xx    | Global standard | General available |
|                | gpt-4.1-nano | xxxx-xx-xx    | Global standard | General available |
|                | o4-mini      | xxxx-xx-xx    | Global standard | General available |

#### Quick facts

|                |                              |
|----------------|------------------------------|
| Model provider | Microsoft                    |
| Type           | Chat completion              |
| Lifecycle      | Preview                      |
| Input type     | text, image                  |
| Output type    | text                         |
| Context        | 32768 output                 |
| Pricing        | <a href="#">View pricing</a> |

#### Others who deployed this model also used

- gpt-5**  
Chat completion
- codex-mini**  
Responses
- DeepSeek-R1-0528**  
Chat completion

# FINE-TUNING IN FOUNDRY

Effortless model customization with support for new models, synthetic data, and quick-start training

The screenshot displays the Microsoft Foundry interface for fine-tuning models. The top navigation bar includes the Microsoft Foundry logo, a search bar with the text "Search with AI (Ctrl + K)", and navigation links for "New Foundry", "Home", "Discover", "Build", "Operate", "Docs", and user profile icons. A left sidebar contains navigation options: Agents, Workflows, Models, Fine-tune (highlighted), Tools, Knowledge, Data, Evaluations, and Guardrails.

The main content area is titled "Fine-tuning (2)" and features a "Fine-tune" button in the top right. Below the title is an "Ask AI" section with three prompts: "When should I consider fine-tuning?", "What RBAC roles do I need to fine-tune a model?", and "What regions are supported for fine-tuning?".

A search bar labeled "Search jobs" is positioned above a table of fine-tuning jobs. The table has the following columns: Name, Creation date, Base model, Status, and Customization method.

| Name  | Creation date | Base model              | Status    | Customization method |
|---|---------------|-------------------------|-----------|----------------------|
| <input checked="" type="radio"/> gpt-4.1-mini-2025-04...65d69 |               | gpt-4.1-mini-2025-04-14 | Completed | Supervised           |
| <input type="radio"/> gpt-4.1-2025-04-14.f...e5ea9            |               | gpt-4.1-2025-04-14      | Completed | Supervised           |

Below the table, it shows "1-2 of 2" items with "Prev" and "Next" navigation arrows.

The right sidebar provides details for the selected job "gpt-4.1-mini". It includes a "Completed" status badge, a "Deploy" button, and a "Continuous fine-tuning" button. The "Model attributes" section shows "gpt-4.1-mini-2025-04-14" and "Supervised".

The "Task parameters" section lists:

- Batch size: 1
- Learning rate multiplier: 2
- Number of epochs: 3
- Seed: 909805774

The "Metrics" section shows:

- Final train loss: 0.26
- Final train mean token accuracy: 0.9

A "View all metrics" link is located at the bottom right of the metrics section.

# FOUNDRY LOCAL

Run, customize, and control AI—right on the device

```
C:\>foundry model run qwen2.5-0.5b
```

```
Model qwen2.5-0.5b-instruct-ovintino-npu:1 was found in the local cache.
```

```
Interactive Chat. Enter /? or /help for help.
```

```
Press Ctrl+C to cancel generation. Type /exit to leave the chat.
```

```
Interactive mode, please enter your prompt
```

```
> Write a sentence with the word 'sun'.
```

```
🧠 Thinking...
```

```
🤖 In the vast expanse of the universe, where the sun's warmth and brightness are reflected on the vast expanse of the cosmos, the sun's light and warmth are reflected on the vast expanse of the cosmos, creating a breathtaking and awe-inspiring scene that is both beautiful and profound.
```

# Azure Foundry Local: l'AI all'Edge che cambia le regole del gioco

Lingua: Italiano

📍 Track 1 (Envisioning Theater - 1° piano)

🕒 Orario: 11:30 - 12:15

## Abstract

L'AI non vive più solo nel cloud: si avvicina ai dati, e Azure Foundry Local è un passo decisivo in questa direzione. Scopriremo come portare modelli avanzati, orchestrazione e sicurezza enterprise direttamente sull'edge, ottenendo inferenze immediate anche offline. Analizzeremo vantaggi e sfide dell'AI on the edge — latenza minima, controllo dei dati, resilienza, requisiti hardware e costi — attraverso esempi e pattern che mostrano quando l'edge è la scelta vincente e come integrarlo in una strategia AI ibrida moderna.

## Speaker



[Marco Dal Pino](#)

Technology Consultant -  
Microsoft

1

Models

2

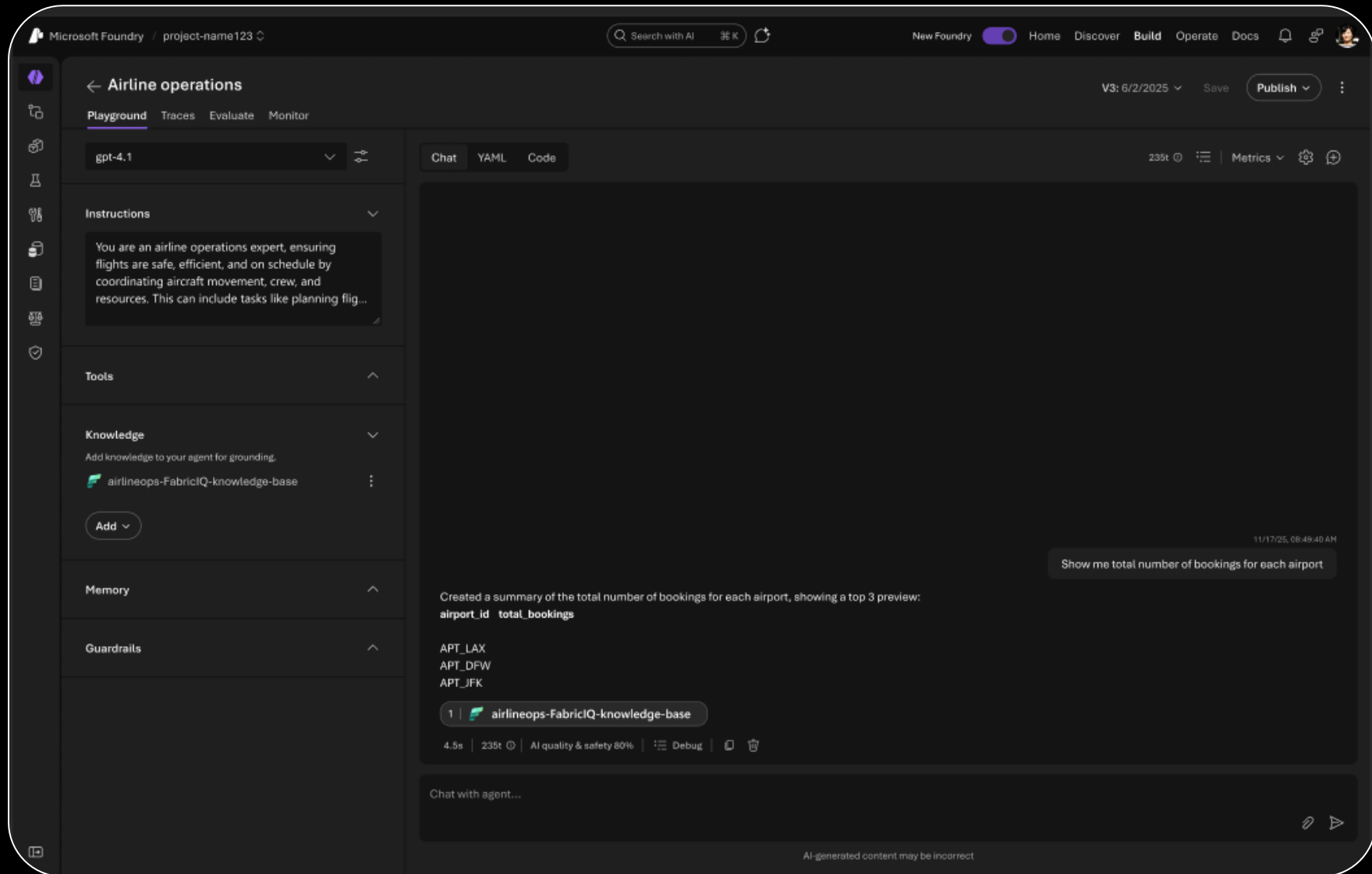
Hosted  
Agents

3

Developer  
stack & tools

# **FOUNDRY AGENT SERVICE**

Securely build, connect, and scale agents with workflows, built-in memory, and serverless hosting



The screenshot shows the Foundry Agent Service interface for an agent named "Airline operations". The interface is dark-themed and includes a sidebar on the left with navigation options: Playground, Traces, Evaluate, and Monitor. The main area is divided into several sections:

- Model:** gpt-4.1
- Instructions:** You are an airline operations expert, ensuring flights are safe, efficient, and on schedule by coordinating aircraft movement, crew, and resources. This can include tasks like planning flight...
- Tools:** (Empty)
- Knowledge:** Add knowledge to your agent for grounding. Includes a knowledge base named "airlineops-FabricIQ-knowledge-base" with an "Add" button.
- Memory:** (Empty)
- Guardrails:** (Empty)

The chat history shows a user prompt: "Show me total number of bookings for each airport" (timestamped 11/17/25, 08:48:40 AM). The agent's response is: "Created a summary of the total number of bookings for each airport, showing a top 3 preview:" followed by a table:

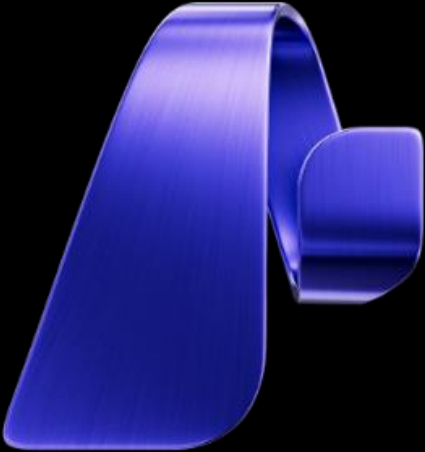
| airport_id | total_bookings |
|------------|----------------|
| APT_LAX    |                |
| APT_DFW    |                |
| APT_JFK    |                |

Below the table, a citation is shown: "1 | airlineops-FabricIQ-knowledge-base". The response also includes a timestamp of 4.5s, a token count of 235t, and a quality/safety score of 80%. The interface also features a "Chat with agent..." input field at the bottom and a disclaimer: "AI-generated content may be incorrect".

# MICROSOFT AGENT FRAMEWORK

Open-source engine for building  
and orchestrating intelligent AI agents

The screenshot displays the Microsoft Agent Framework Dev UI for a 'Content Review Workflow'. The interface is dark-themed and includes a top navigation bar with 'Dev UI' and a dropdown menu for 'Content Review Workf...'. Below the navigation, the workflow name 'Content Review Workflow' is shown with a refresh icon and a 'Run Again' button. A description reads: 'Multi-agent content creation workflow with quality-based routing (Writer → Reviewer → Editor/Publisher)'. The main area features a flowchart with five nodes: 'Writer', 'Reviewer', 'Publisher', 'Editor', and 'Summarizer', all labeled as 'AgentExecutor'. The flow starts with 'Writer', goes to 'Reviewer', then branches to 'Publisher' and 'Editor', and finally to 'Summarizer'. On the right, a panel shows 'Events' with 77 (1784 raw) items. The event log includes: 'completed' (Response complete (5235 tokens)), 'output\_item.done' (response.output\_item.done), and several 'output\_text.delta' events containing text snippets like '- The interplay between theory and experiment continues to d...', '- Significant theoretical challenges persist, particularly i...', '- Experimental physics is pushing quantum effects toward mac...', '- Quantum technologies are rapidly advancing, notably in com...', and '3. Key Highlights and Takeaways: - Quantum mechanics conti...'. At the bottom left, the 'Last Executor' section shows 'Summarizer' at 4:31:16 PM with the status 'Executor started'. A green 'Workflow Complete' notification is displayed, showing a preview of the generated text: 'discusses core principles such as superposition and entanglement, recent theoretical advancements in quantum field theory and quantum information science, as well as the ongoing debate around the interpretation of quantum mechanics. The report highlights cutting-edge technological applications including quantum computing, quantum cryptography, and quantum sensing. Ex progress is reviewed, noting advances in emerging'. A 'View Full' link is visible. At the bottom right, there is a link to the 'Deployment Guide for Content Review Workflow'.



DEMO

# Building Intelligent Agents on Azure: from local dev to production with Aspire

Lingua: Italiano

📍 Track 1 (Envisioning Theater - 1° piano)

🕒 Orario: 16:30 - 17:15

## Abstract

Negli ultimi mesi il concetto di AI Agent è passato rapidamente dalla sperimentazione locale a scenari distribuiti e produttivi. Ma come si costruisce un'architettura di agenti completa, coerente e manutenibile, senza perdersi tra tool e complessità operative?

In questa sessione esploreremo un percorso end-to-end per lo sviluppo di agenti su Azure, partendo dal local development con Microsoft Agent Framework, passando dall'integrazione con agenti hostati su Azure AI Foundry, fino al deployment cloud-native tramite .NET Aspire.

## Speaker



**[Tommaso Stocchi](#)**

Cloud Solution Architect -  
Microsoft

1

Models

2

Hosted  
Agents

3

**DEVELOPER**  
stack & tools

# AGENT DEVELOPMENT IS DIFFERENT

## Traditional apps

Fixed rules, manual updates

Rigid, predefined steps

Infrastructure-heavy,  
platform-specific code

Coding & backend

Logic

Workflow

Development  
focus

Skills

## AI agents

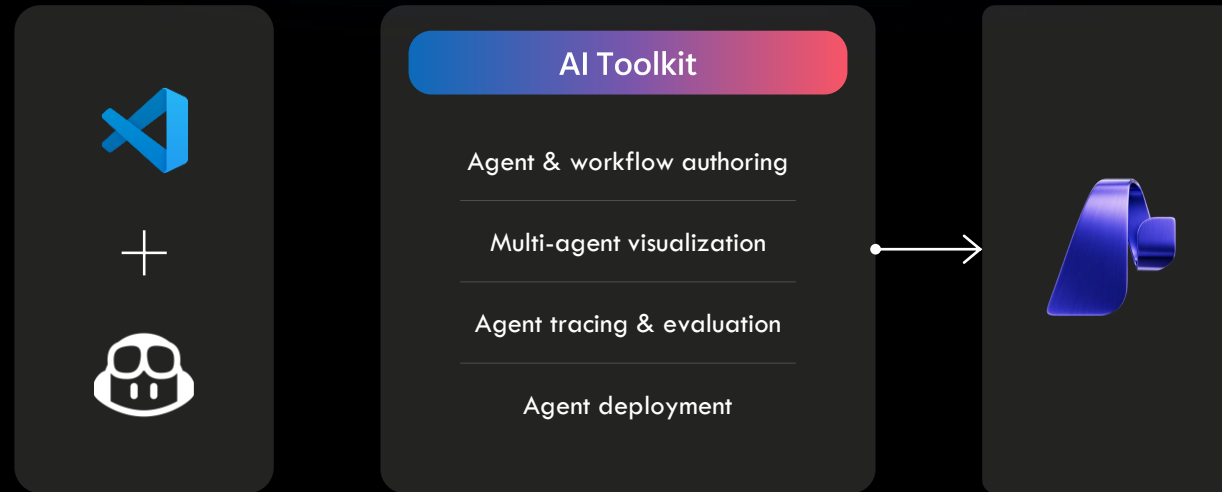
Autonomous, adaptive reasoning

Dynamic, iterative loops

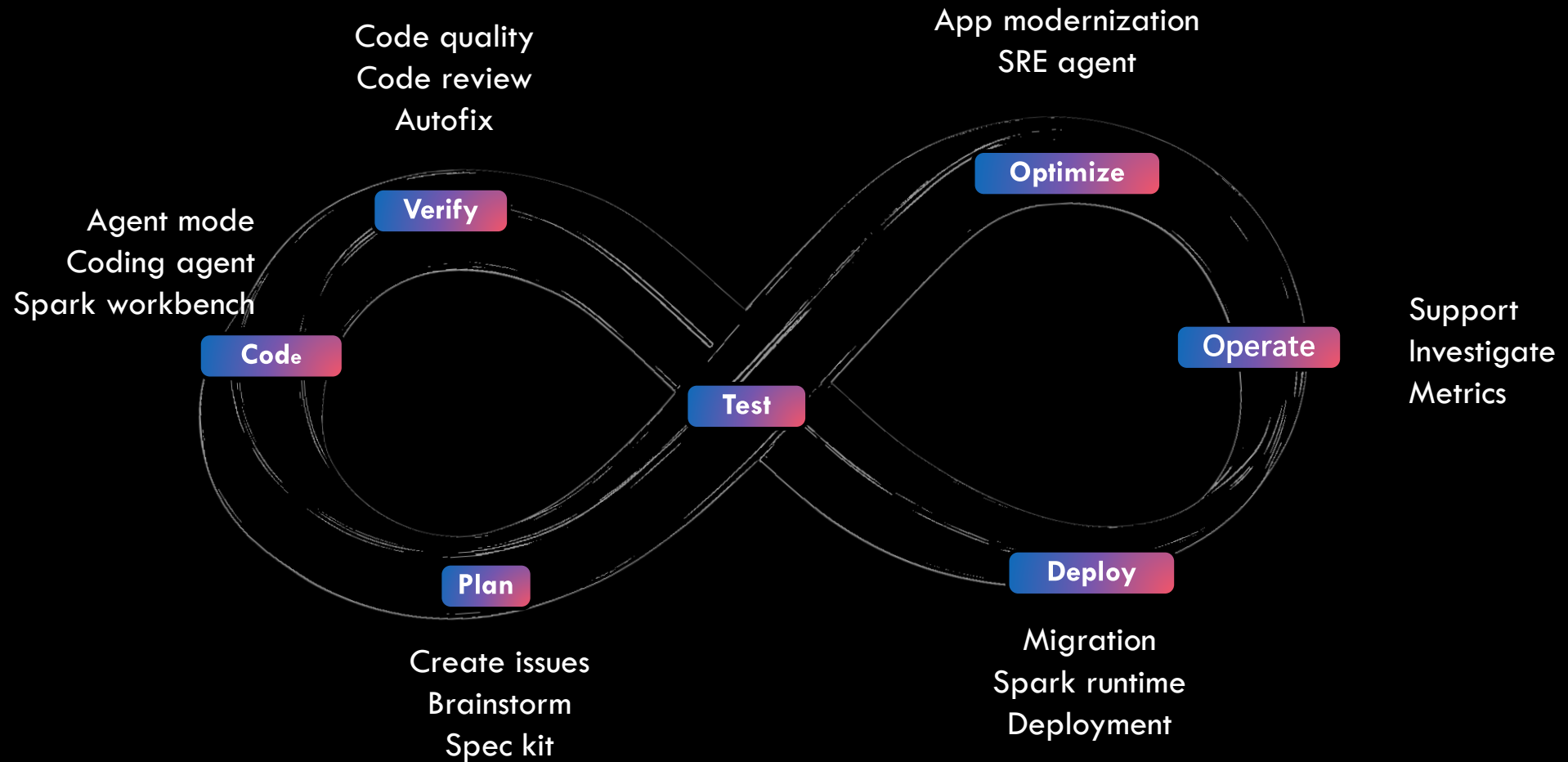
Intelligence, orchestration,  
adaptability

Specification-driven, orchestration

# AI TOOLKIT FOR VS CODE



# AGENTS ACROSS THE ENTIRE SOFTWARE LIFECYCLE





①

## PRE-BUILT & CUSTOM MODELS

Model Catalog · Router · Fine-tuning · Foundry Local

DEMO 1

②

## RESPONSIBLE AI AND BEST PRACTICES

Guardrails · Evaluations · Control Plane · Security

DEMO 2

③

## INTEGRATE AI INTO .NET

VS Code AI Toolkit · GitHub Copilot · Multi-agent patterns

DEMO 3

# GET STARTED

Define scope and requirements

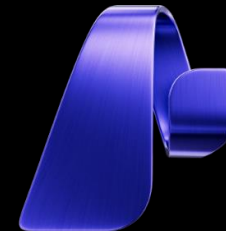
Prepare data

Build, train and evaluate

Review and approve app for production

Deploy and monitor

Govern and monitor continuously



[ai.azure.com](https://ai.azure.com)

# WITH YOU EVERY STEP OF THE WAY



**Explore AI use cases  
and solutions**

<https://azure.microsoft.com/solutions/ai/>



**Envision Envisioning  
workshop**

<https://Microsoft.com/hub>



**Execute Azure AI  
template gallery**

<https://aka.ms/ai-apps>

# ABOUT ME



**Digital Product Manager – *FME Italy***

**Former Microsoft MVP – *Developer Technologies***

**Coach - <https://codingcoach.io/>**

**Mentor**

**Book Author**

**<https://dotnet.social/@codrina>**

**<https://bsky.app/profile/codrina.bsky.social>**



# GRAZIE!



[\\_Codrina\\_](#)



[codrinamerigo](#)



[codrinamerigo](#)

