



Microsoft Foundry Part 1

CSCI E-94

Fundamentals of Cloud Computing - Azure

Joseph Ficara

Copyright © 2013-2026

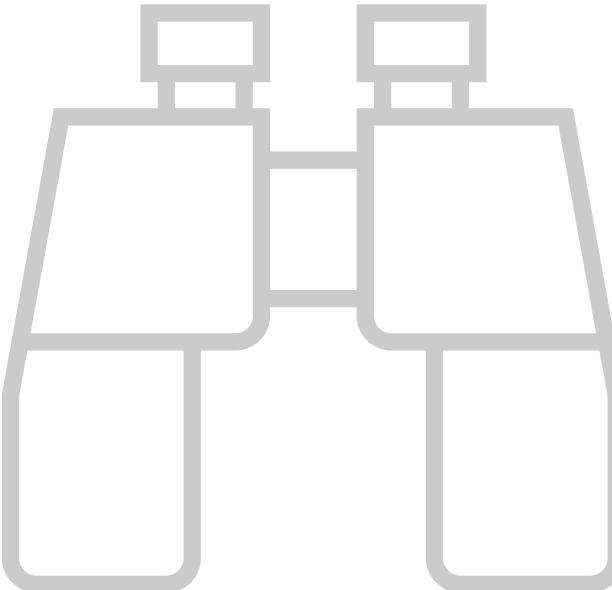


Agenda

- Microsoft Foundry
 - Overview
 - What is it?
 - Why do you care?
 - Essentials
 - Configuring in Azure
 - Simple Chatbot
 - Structured Output JSON



Overview





Microsoft Foundry - Overview

- What is it?
 - Unified AI Platform
 - End-to-end AI development
 - Low-code/no-code prompt flow workflows
 - SDKs
- Experimentation & Deployment
 - Test ideas in the playground
 - Deploy to the cloud

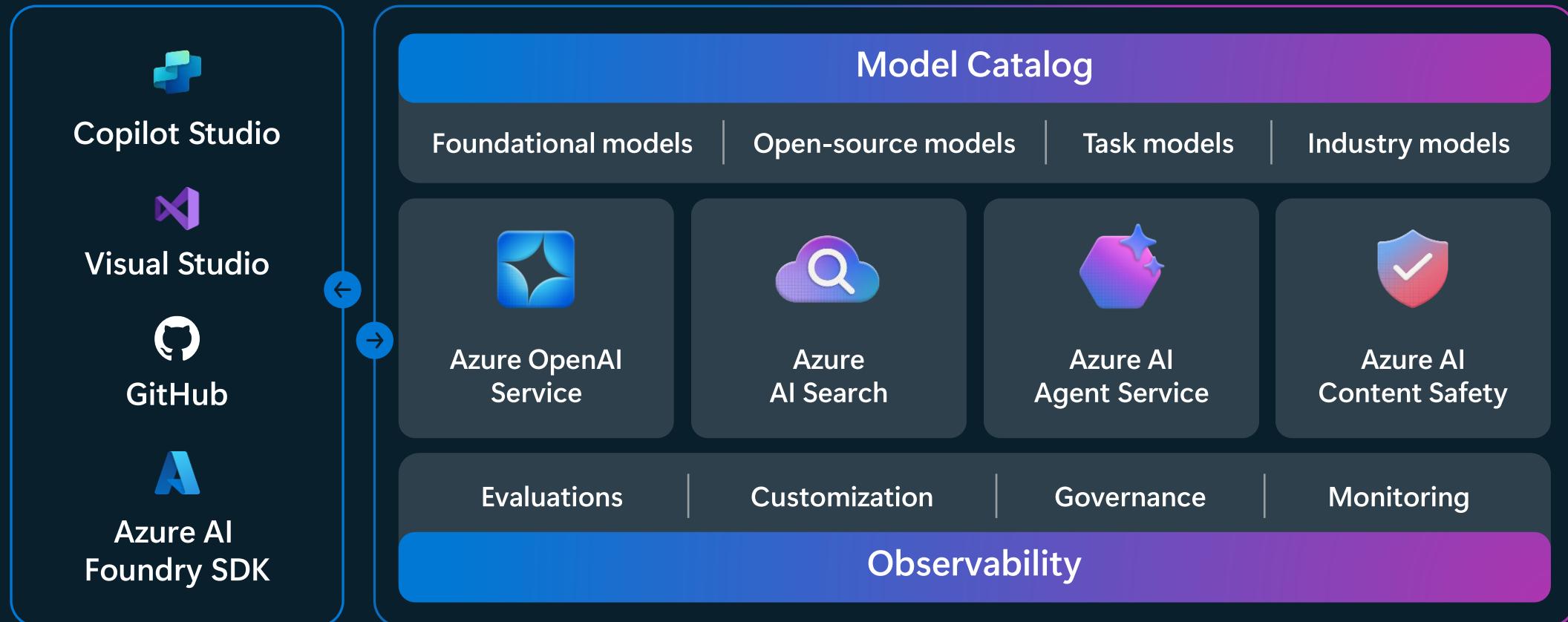


Microsoft Foundry - Overview

- Why do you care?
 - Accelerate development of AI powered solutions
 - Highly available infrastructure
 - One stop place to organize, manage and maintain
 - Your Azure based AI solutions



Microsoft Foundry Overview





Microsoft Foundry - Overview

What is it

- Collaborative Project Management
 - Projects
 - Manage containers, resources, connectivity and security
- Integration with Azure Services
 - AI services
 - Azure Storage
 - Databases



Microsoft Foundry Overview

Open and Modular Platform

New Collaborations

accenture

gretel

KPMG

scale

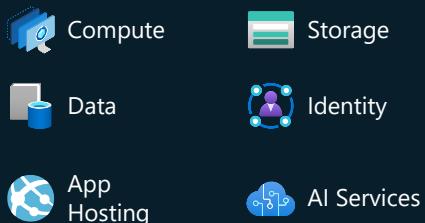
STATSIG

Weights & Biases

Ingest data



Provision cloud resources



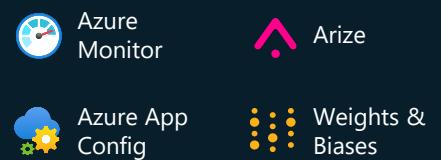
Orchestrate with OSS frameworks



Leverage Azure App Services



Monitor in production



Microsoft
Copilot Studio



Microsoft
Power Platform



Azure AI
Foundry portal



Azure Machine
Learning



Github
Enterprise



Visual
Studio Code



Azure
DevOps



Microsoft
Azure Portal



Microsoft Foundry - Overview

- Diverse AI models
 - Over 11,000 models
 - Anthropic Claude
 - DeepSeek
 - Grok
 - Hugging Face
 - OpenAI
 - Microsoft
 - Mistral
 - Other open source and public models



Microsoft Foundry Overview

Support for diverse selection of models

Now 11,299 + Frontier, task, and Open Models



OpenAI
Model Family
(available day 1)



Phi SLM
Model Family



Mistral AI
Model Family



Meta Llama 2
Model Family



Jais G42
Model Family



Cohere
Model Family



Databricks
Model Family



Hugging Face
Collection

New



New



New

Industry
AI Models



Microsoft Foundry - Overview

What is it

- Responsible AI
 - Content Safety
 - Input & Output filters
 - Safety validation



Microsoft Foundry Overview

Azure AI Content Safety



Harm
categories



Groundedness
detection



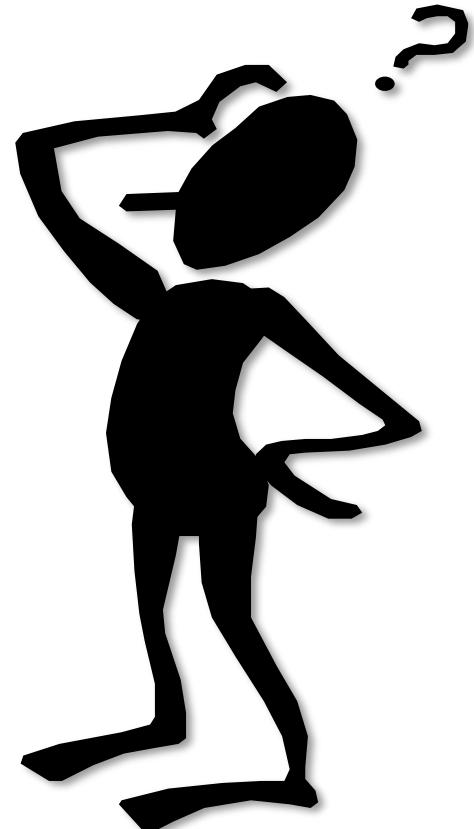
Prompt
shields



Protected
material
detection



Microsoft Foundry - Overview





Creating a Foundry Resource

- First register the `microsoft.insights` resource provider

Azure services

Create a resource Subscriptions Foundry

Home > Subscriptions >

Subscriptions

Joe Ficara Demo Student 2026 (jficara2026.onmicrosoft.com)

+ Add Manage Policies View Requests View eligible subscriptions Export to CSV

Global administrators can manage all subscriptions in this list by updating their policy setting [here](#).

Showing subscriptions in Joe Ficara Demo Student 2026 directory. Don't see a subscription? [Switch director](#)

Search for any field... Subscriptions : All (1 of 1) My role == all Status == all

Subscription name ↑↓	Subscription ID ↑↓
2026-Lab_Joe_Ficara Demo Student	2c68c4de-565b-47a2-aa40-70232287ff32

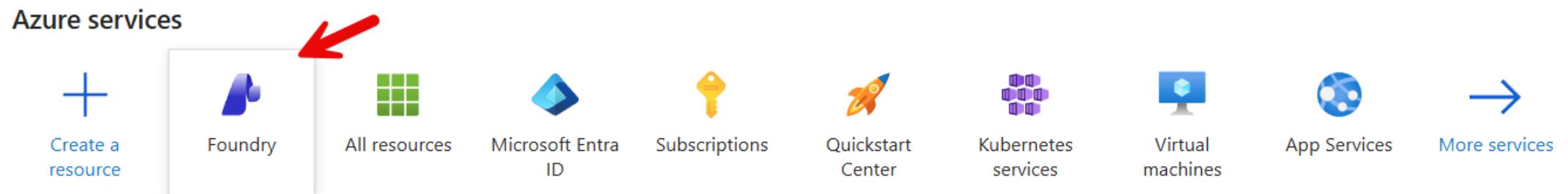
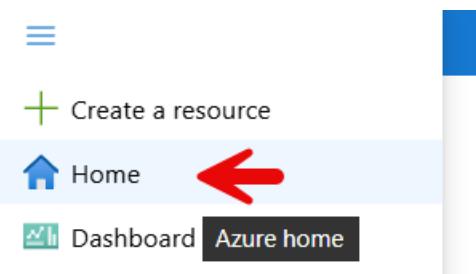
2026-Lab_Joe_Ficara Demo Student | Resource providers

Provider	Status	Registration Policy
Microsoft.Portal	Registered	RegistrationRequired
Microsoft.ResourceGraph	Registered	RegistrationFree
Microsoft.ResourceIntelligence	Registered	RegistrationFree
Microsoft.ResourceNotifications	Registered	RegistrationFree
Microsoft.Resources	Registered	RegistrationFree
Microsoft.Search	Registered	RegistrationRequired
Microsoft.SerialConsole	Registered	RegistrationFree
Microsoft.Storage	Registered	RegistrationRequired
Microsoft.Web	Registered	RegistrationRequired
microsoft.insights	Registered	RegistrationRequired
microsoft.support	Registered	RegistrationFree
ArizeAi.ObservabilityEval	NotRegistered	RegistrationRequired
Astronomer.Astro	NotRegistered	RegistrationRequired
Dell.Storage	NotRegistered	RegistrationRequired
Dynatrace.Observability	NotRegistered	RegistrationRequired
GitHub.Network	NotRegistered	RegistrationRequired



Creating a Foundry Resource

- Next: Navigate to Home and choose Foundry





Creating a Foundry Resource

- Fill out the card
 - Resource group,
 - Name
 - Region
 - East US2
 - Newer models
 - Default project name

The screenshot shows the 'Create a Foundry resource' wizard in the Microsoft Azure portal. The 'Basics' tab is selected. The 'Instance Details' section includes fields for Subscription (2026-Lab_Joe_Ficara Demo Student), Resource group (1. (New) rg_aifoundry), Name (2. ai-cscie94-foundry-demo-01), and Region (3. East US). The 'Your first project' section shows a Default project name (4. 01-lecture). The 'Content Review Policy' section provides information about Microsoft's access to Azure OpenAI models. At the bottom, there are 'Previous' and 'Next' buttons, and the 'Review + create' button is highlighted with a red circle containing the number 5.



Creating a Foundry Resource

- Setup KeyVault
 - Create new

The screenshot shows the Microsoft Azure portal interface for creating a Foundry resource. On the left, the main page displays sections for Basics, Storage, Network, Identity, Encryption, Tags, and Review + create. The Storage section is currently selected. Below this, there's a note about credential storage and application logging, followed by a warning about default configurations for projects. The Agent service section includes options for Model Deployment, Cosmos DB, AI Search, and Storage Account. The Speech and Language service section also includes similar integration options. At the bottom, there are navigation buttons for Previous, Next, and Review + create, along with a URL bar containing <https://portal.azure.com/#>. On the right, a modal window titled "Select Key Vault" is open. It shows a message stating "Failed to load existing resources - you can still create new ones" and a button labeled "Create new". A red circle with the number 6 is placed over the message area, and another red circle with the number 7 is placed over the "Create new" button.



Creating a Foundry Resource

- Setup KeyVault
 - Create new

The screenshot shows the Microsoft Azure 'Create a Foundry resource' wizard and a 'Select Key Vault' modal.

Main Wizard:

- Storage Tab:** Selected. It includes sections for 'Credential storage and application logging' (Key Vault (preview) and Application Insights), a warning about default configurations, and an 'Agent service' section with 'Model Deployment' and other options.
- Bottom Navigation:** Previous, Next, Review + create.

Select Key Vault Modal:

- Name:** Failed to load existing resources - you can still create new ones.
- Create new:** Button with a red circle containing the number 6.
- Key Vault:** kv-cscie94-demo-01 (selected).
- Buttons:** Cancel, Create.

Bottom Buttons: Add (with a red circle containing the number 7), Cancel.



Creating a Foundry Resource

- Setup
 - App Insights
- Click
 - Review & create

The screenshot shows the Microsoft Azure 'Create a Foundry resource' wizard and a 'Select Application Insights' modal dialog.

Main Wizard:

- Storage Tab:** Selected. Shows 'Key Vault (preview)' set to 'kv-cscie94-demo-01' and 'Select Azure Key Vault'. Shows 'Application Insights' with 'Select Application Insights' link.
- Warning:** 'The configurations below will be the default for all projects created under this Foundry resource. You may override the storage location per project by using infrastructure templates. [Learn more](#)'
- Agent service:** Describes managing threads, messages, deployments, and files within the Foundry resource. Options include 'Select Resources' and 'Delete'.
- Speech and Language service:** Describes managing file uploads and outputs of Speech and Language tasks.
- Buttons:** 'Previous', 'Next', 'Review + create' (highlighted with a red circle containing the number 10).

Select Application Insights Modal:

- Header:** 'Select Application Insights'.
- Message:** 'Name * Failed to load existing resources - you can still create new ones'.
- Input:** 'Name' field containing 'appi-cscie94-foundry-demo' (highlighted with a red circle containing the number 8).
- Buttons:** 'Cancel' and 'Create'.

Bottom Buttons:

- 'Add' (highlighted with a red circle containing the number 9)
- 'Cancel'



Creating a Foundry Resource

■ Review & Create - Accepts defaults for:

- Network
- Identify
- Encryption
- Tags

Home > Microsoft Foundry | Foundry >
Create a Foundry resource ...

Basics Storage Network Identity Encryption Tags Review + create

Inbound Access *

All networks, including the internet, can access this resource.
 Selected networks, configure network security for your Foundry resource.
 Disabled, no networks can access this resource. You could configure private endpoint connections that will be the exclusive way to access this resource.

⚠️ Agents APIs only support network injection for Standard Agent set-up. If you would like to deploy a network secured Agent set-up, deploy an Agent template. [Learn more about Agent networking configurations.](#)

Home > Microsoft Foundry | Foundry >
Create a Foundry resource ...

Basics Storage Network Identity Encryption Tags Review + create

Identity type *

System assigned
 User assigned

💡 We'll grant your user identity the **Azure AI User** role so you can develop with all projects under this resource.

To enable other users to use this project:
• Assign **Azure AI User** to develop with projects.
• Assign **Azure AI Project Manager** to develop and manage project settings.

[Learn more](#)

Home > Microsoft Foundry | Foundry >
Create a Foundry resource ...

Basics Storage Network Identity **Encryption** Tags Review + create

Data Encryption

Your data is encrypted by default using Microsoft-managed keys. For additional control over your data, you may choose to bring your own key for encryption. [Learn more about customer-managed key encryption.](#)

Encrypt data using a customer-managed key



Creating a Foundry Resource

- Review Summary
 - Click Create

Create a Foundry resource ...

Basics Storage Network Identity Encryption Tags Review + create

[View automation template](#)

Basics

Subscription	2026-Lab_Joe_Ficara Demo Student
Resource group	rg_aifoundry
Name	ai-cscie94-foundry-demo-01
Region	East US
Default project name	01-lecture

Storage

Key Vault (preview)	kv-cscie94-demo-01
Application Insights	appi-cscie94-foundry-demo
Storage for Agents service	0 item(s)
Storage Account (preview)	

Network

Inbound Access	All networks, including the internet, can access this resource.
----------------	---

Identity

Identity type	System assigned
---------------	-----------------

Previous Next **Create** 11



Configuring in Azure

- Once deployed navigate to the Foundry UI
 - Pick your foundry instance

Azure services

The screenshot shows the Azure services dashboard. A red arrow labeled '1' points to the 'Foundry' service tile. Other visible tiles include 'All resources' (grid icon) and 'Microsoft Entra ID' (blue ID icon).

Home > Microsoft Foundry

Microsoft Foundry | Foundry

Joe Ficara Demo Student 2026

Search

+ Create Manage deleted resources Manage view

You are viewing a new version of Browse experience. Click here to...

Filter for any field... Subscription equals all

Foundry

AI Hubs

Azure OpenAI

AI Search

More services

Classic AI services

Help

The screenshot shows the Microsoft Foundry UI. A red arrow labeled '2' points to the 'Foundry' item in the sidebar navigation. Another red arrow labeled '3' points to the first item in the list of foundry instances: 'ai-cscie94-eastus-foundry-demo-01'.

Name ↑
<input type="checkbox"/> ai-cscie94-eastus-foundry-demo-01
<input type="checkbox"/> ai-cscie94-eastus2-foundry-demo-01



Configuring in Azure

- Deploy a model
 - Pick gpt-5-mini

The screenshots illustrate the steps to deploy a model in Microsoft Foundry:

1. In the left sidebar, click on **Models + endpoints**.
2. Click on the **+ Deploy model** button.
3. Click on **Deploy base model**.
4. In the search bar, type **gpt-5-mini**.
5. Select the **gpt-5-mini** model from the list.
6. Click on the **Confirm** button.

Model deployments

Try the new Microsoft Foundry experience
Build and manage agents and workflows with access to a rich catalog of MCP serverless services.

Overview Model catalog Playgrounds Build and customize Observe and optimize Protect and govern Azure OpenAI Stored completions Batch jobs My assets Data + indexes Models + endpoints 1

+ Deploy model 2 Refresh

Model deployments

Try the new Microsoft Foundry experience
Build and manage agents and workflows with access to a rich catalog of MCP serverless services.

Overview Model catalog Playgrounds Build and customize Observe and optimize Protect and govern Azure OpenAI Stored completions Batch jobs My assets Data + indexes Models + endpoints 3

+ Deploy model Refresh

Select a model

Choose a model to create a new deployment. For flows and other resources, create a deployment from their respective list. [Go to model catalog](#).

Models 85 Inference tasks Show description

Model	Description	Status
gpt-5-mini	Chat completion, Responses	Selected 4
grok-5-mini	Chat completion	Available 5
Phi-3.5-mini-instruct	Chat completion	Available
Phi-4-mini-reasoning	Chat completion	Available
Phi-4-mini-instruct	Chat completion	Available
AI21-Jamba-1.5-Mini	Chat completion	Available
Phi-3-mini-128k-instruct	Chat completion	Available

gpt-5-mini

Task: Chat completion Task: Responses

Direct from Azure models

Direct from Azure models are a select portfolio curated for their market-differentiated capabilities:

- Secure and managed by Microsoft: Purchase and manage models directly through Azure with a single license, consistent support, and no third-party dependencies, backed by Azure's enterprise-grade infrastructure.
- Streamlined operations: Benefit from unified billing, governance, and seamless PTU portability across models hosted on Azure - all part of Microsoft Foundry.
- Future-ready flexibility: Access the latest models as they become available, and easily test, deploy, or switch between them within Microsoft Foundry; reducing integration effort.
- Cost control and optimization: Scale on demand with pay-as-you-go flexibility or reserve PTUs for predictable performance and savings.

Learn more about [Direct from Azure models](#).

6 Confirm Cancel



Configuring in Azure

- Deploy a model
 - Enter a Deployment Name
 - Choose a deployment type

Deploy gpt-5-mini

Deployment name * ←

Deployment type ←

Global Standard: Pay per API call with the highest rate limits. Learn more about [Global deployment types](#).

Data might be processed globally, outside of the resource's Azure geography, but data storage remains in the AI resource's Azure geography. Learn more about [data residency](#).

Deployment details Customize

Model version 2025-08-07	Authentication type Key
Capacity 150K tokens per minute (TPM)	Resource location East US
Content safety DefaultV2	Version upgrade policy Once a new default version is available

→ Deploy Cancel



Test in playground

- You can now experiment in the playground

The screenshot shows the Microsoft Foundry web interface. At the top, the navigation bar includes the Microsoft Foundry logo, the path '01-lecture / Deployments / gpt-5-mini', and a banner encouraging users to try the new Microsoft Foundry experience. On the left, a sidebar menu lists 'Overview', 'Model catalog' (which is currently selected and highlighted with a green cursor icon), 'Playgrounds', 'Build and customize', 'Observe and optimize', 'Protect and govern', 'Azure OpenAI', 'Stored completions', and 'Batch jobs'. The main content area displays a deployment named 'gpt-5-mini'. It features tabs for 'Details' and 'Metrics', with 'Details' being the active tab. A prominent blue button labeled 'Open in playground' is highlighted with a large red arrow pointing to it. Below this button, there are links for 'Request quota', 'Edit', and 'Delete'. The 'Endpoint' section contains a 'Target URI' field with the value 'https://ai-cscie94-eastus-foundry-demo-01.cognitiveservices.azure.com/openai/respo...' and a 'Key' field containing a redacted string of dots. To the right, partially visible sections include 'Lang' (language), 'Pyt' (Python), 'Ge' (General), and 'Bel' (Belo).



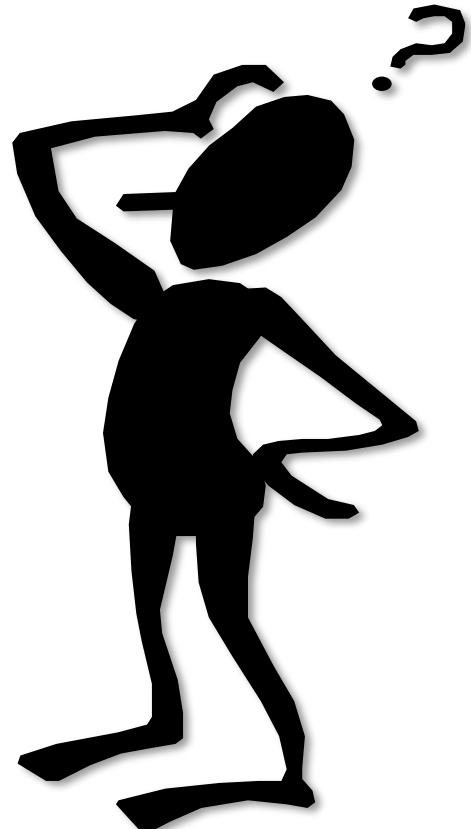
Demo

Create Microsoft Foundry, Project & Test deployment

Azure Portal



Configuring in Azure





Let's Code!

A large, faint watermark of a brain composed of binary code (0s and 1s) is centered on the slide.

```
10101101011010101101010110101011010101101010110111011010101011010101010101010  
10101101101101010101101010110101011010101011010101101010110101010101010101001  
01010101001010101010110100110101010101101010101010101010101010101101010101010  
10101100101010110101101010110101011010101010101010101010101010101010101010101  
011010101011010101011010001011010101010101010101010101010101010101010101011010  
10101010101010101010101010101010101010101010101010101010101010101010101010101  
01011010101010101010101010101010101010101010101010101010101010101010101011010  
10101010101010101011001101010101010101010101010101010101010101010101010101010  
100110010101010101010101010101010101010101010101010101010101010101010101010101  
00101010101010101010101010101010101010101010101010101010101010001010101010101  
01001010101010101010101010101010101010101010101010101010101010101010101010101  
01010101001100110010101010101010101010101010101010101010101010101010101010101  
0101011111010101010101010101010101010101010101010101010101010101010101010101  
01010100101010101010101010101010101010101010101010101010101010101010101010101  
01011010100001010101001010101010101010101010101010101010101010101010101010101  
01010101010010101010101010101010101010101010101010101010101010101010101010101  
10101010101010101010101010101010101010101010101010101010101010101010101010101  
10101010101010101010101010101010101010101010101010101010101010101010101010101  
0010101010101101010101010101011101010101010101010101010101010101010101010101  
100010101010101010101010101010101010101010101010101010101010101010101000101010101
```



Simple Chatbot

- Basic Requirements:
 - Allow a caller to pass in a prompt
 - Using a POST request
 - Use an **SLM** to process the content of the prompt
 - **SLM** -> Small Language Model
 - Return the result from the **SLM**
 - To the caller



Simple Chatbot

- Nuget packages needed

```
...<ItemGroup>
....<PackageReference Include="Microsoft.AspNetCore.OpenApi" Version="10.0.2" />
....<PackageReference Include="Swashbuckle.AspNetCore" Version="10.1.0" />
....<PackageReference Include="Azure.AI.OpenAI" Version="2.1.0" />
....<PackageReference Include="Microsoft.Extensions.AI" Version="10.2.0" />
....<PackageReference Include="Microsoft.Extensions.AI.OpenAI" Version="10.2.0-preview.1.26063.2" />
...</ItemGroup>
```



Simple Chatbot

- How it works
 - Reads settings
 - Uri to your Azure OpenAI Service in the connected resources
 - ApiKey to access the model
 - Connect to a deployed **gpt 5 mini SLM**
 - Using the settings defined by the **AISettings**
 - Passes the string provided in the body of the POST request
 - To the SLM
 - Returns result from the SLM
 - As the result of the POST request



Test in playground

- URL and API Key
 - Be sure to **choose OpenAI** to obtain the correct URL!

The screenshot shows the Azure OpenAI studio interface for a project named "01-lecture". The left sidebar has a yellow arrow pointing to the "Overview" tab. The main content area displays the "Endpoints and keys" section. An "API Key" field contains a redacted value with a copy icon. A blue arrow points to this field. Below it is a "Libraries" section with a "Microsoft Foundry" dropdown set to "Azure OpenAI", which is highlighted with a red box. Another red arrow points to this selection. To the right, an "Azure OpenAI endpoint" field contains the URL `https://ai-cscie94-eastus-foundry-demo-01.openai.azure.com`, also highlighted with a red box. A red arrow points to this URL. At the bottom, there is an "API documentation" link.



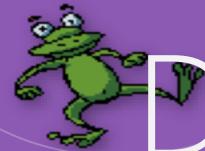
Simple Chatbot

- How do you do it?
 - Get an instance of **IChatClient**
 - Using dependency injection, in your controller's constructor
 - Create a **ChatOptions**
 - Include settings to direct the SLM
 - **Temperature (GPT 5 mini only accepts a Temperature of 1)**
 - The randomness of the SLM/LLM's responses
 - Lower values = more deterministic
 - Higher values = increase creativity
 - **MaxOutputTokens**
 - Maximum number of tokens the SLM/LLM can generate in a response



Simple Chatbot

- Submit the prompt
 - Using `IChatClient` instance & `GetResponseAsync` method
 - Provide
 - The `ChatOptions` instance
 - The `prompt` provided by the call to the REST API
- Return the result from the SLM
 - To the caller



Demo

Simple Chatbot

`SimpleWebAPIChatDemoSolution.sln`

`SimpleWebAPIChatDemo.csproj`



Structured Output JSON

- Why structured output?
 - Facilitates integration into your business logic
 - Makes it easier to deal with results
 - Strongly typed output
- How do you do it?
 - Get an instance of **IChatClient**
 - Using dependency injection
 - In your controller's constructor



Structured Output JSON

- Nuget packages needed

```
...<ItemGroup>
....<PackageReference Include="Microsoft.AspNetCore.OpenApi" Version="10.0.2" />
....<PackageReference Include="Swashbuckle.AspNetCore" Version="10.1.0" />
....<PackageReference Include="Azure.AI.OpenAI" Version="2.1.0" />
....<PackageReference Include="Microsoft.Extensions.AI" Version="10.2.0" />
....<PackageReference Include="Microsoft.Extensions.AI.OpenAI" Version="10.2.0-preview.1.26063.2" />
...</ItemGroup>
```



Structured Output JSON

How do you do it ...

- Create the shape you expect to be returned
 - A public C# Class with public properties
 - Used to define the JSON structure to be returned
- Generate the JSON Schema
 - Using Nuget package [NJsonSchema](#)
- Create a **ChatResponseFormatJson** instance
 - Supply the schema
 - That defines the desired output format



Structured Output JSON

How do you do it ...

- Create a **ChatOptions**
 - Specify the **ResponseFormat**
 - With the **ChatResponseFormatJson** instance
 - Include settings to direct the SLM
 - **Temperature**
 - The randomness of the SLM/LLM's responses
 - Lower values = more deterministic
 - Higher values = increase creativity
 - **MaxOutputTokens**
 - Maximum number of tokens the SLM/LLM can generate in a response
 - Etc...



Structured Output JSON

How do you do it ...

- Submit the prompt using **GetResponseAsync**
 - Include the **ChatOptions** instance
 - Include an enhanced user prompt
\$"Identify and return a JSON list of the most important 3 key phrases from the following text: {prompt}"
 - The prompt in the interpolated C# string is the input from the user



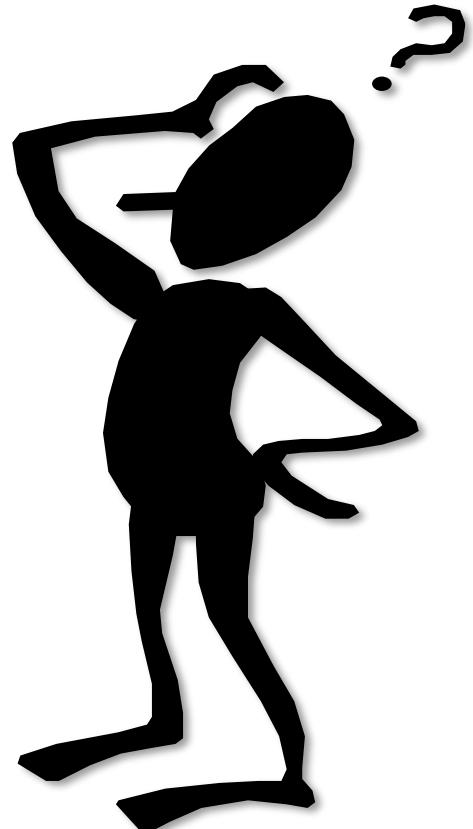
Demo

Structured Output JSON

`StructuredOutputWebAPISolution.sln`
`StructuredOutputWebAPI.csproj`



Questions





Further Reading

■ Microsoft Foundry

- [What is Microsoft Foundry? - Microsoft Foundry | Microsoft Learn](#)
- [Microsoft Foundry architecture - Microsoft Foundry | Microsoft Learn](#)
- [Microsoft Foundry frequently asked questions - Microsoft Foundry | Microsoft Learn](#)