



Azure AI Foundry Part 1

CSCI E-94

Fundamentals of Cloud Computing - Azure

Joseph Ficara

Copyright © 2013-2025

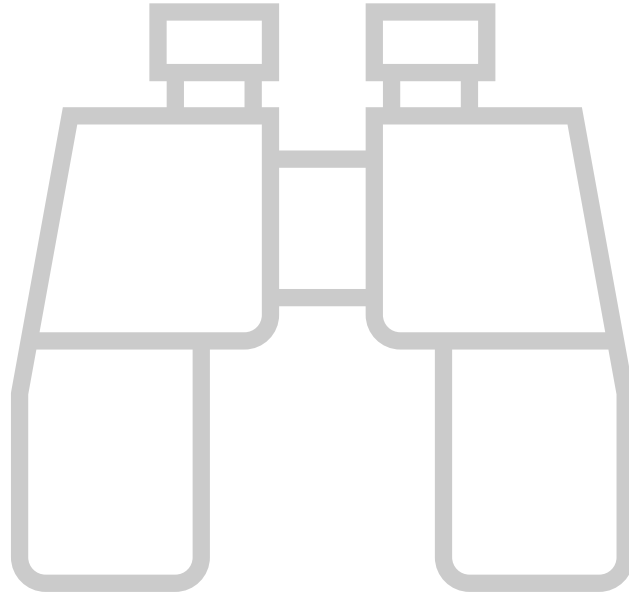


Agenda

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



Overview





Azure AI 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 with

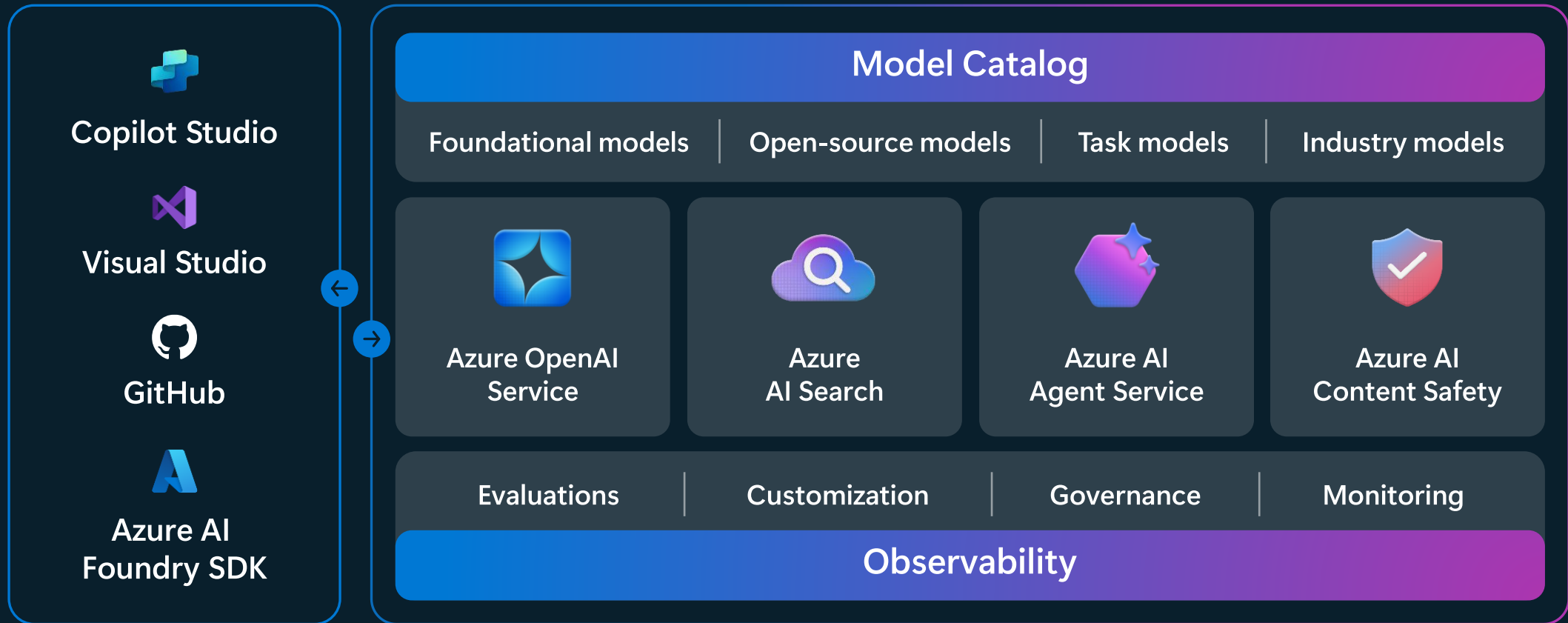


Azure AI Foundry - Overview

- 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



Azure AI Foundry Overview





Azure AI Foundry - Overview

What is it

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



Azure AI Foundry Overview

Open and Modular Platform

New Collaborations

accenture

gretel

KPMG

scale

STATSIG

Weights & Biases

Ingest data



Microsoft Fabric



Azure SQL Database



Cosmos DB



Azure AI Search



Azure Redis



SharePoint



Bing



OSS Databases

Provision cloud resources



Compute



Storage



Data



Identity



App Hosting



AI Services

Orchestrate with OSS frameworks



Semantic Kernel



LangChain



AutoGen



LlamaIndex

Leverage Azure App Services



Functions



Kubernetes



Container Apps



Logic Apps



Dynamic Sessions



PubSub

Monitor in production



Azure Monitor



Arize



Azure App Config



Weights & Biases



Microsoft Copilot Studio



Microsoft Power Platform



Azure AI Foundry portal



Azure Machine Learning



GitHub Enterprise



Visual Studio Code



Azure DevOps



Microsoft Azure Portal



Azure AI Foundry - Overview

- Diverse AI models
 - Over 1800 models
 - OpenAI
 - Hugging Face
 - Other open source and public models



Azure AI Foundry Overview

Support for diverse selection of models

Now 1800 + 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

 NTT DATA

New



New

Industry
AI Models



Azure AI Foundry - Overview

What is it

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



Azure AI Foundry Overview

Azure AI Content Safety



**Harm
categories**



**Groundedness
detection**



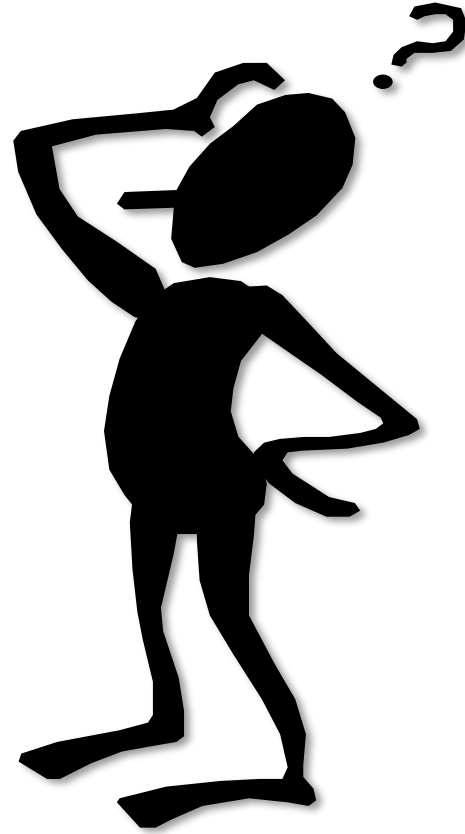
**Prompt
shields**



**Protected
material
detection**



Azure AI Foundry - Overview





Configuring in Azure

- What do you need?
 - Subscription
 - Hub Resource Group
 - Region
 - Azure AI Hub Name
 - Friendly Name
 - Default project resource group
 - Azure AI Services connection

Microsoft Azure Search resources, services, and docs (G+/) Copilot

Home > Azure AI Foundry >

Azure AI hub

Create an Azure AI hub resource

Basics Storage Networking Encryption Identity Tags Review + create

Organization

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources. An AI hub is a collaboration environment for a team to share project work, model endpoints, compute, (data) connections, security settings, govern usage.

Subscription * ⓘ 01-Lab_Joe_Ficara Student

Resource group * ⓘ (New) rg_aihub-02
[Create new](#)

Region * ⓘ East US

Resource details

Name * ⓘ ai-cscie94-hub-demo-02 ✓

Friendly name ⓘ Ai cscie94 hub demo 02

Default project resource group ⓘ rg_aiprojectdemos-02

Azure AI services base models

Connect AI Services incl. OpenAI * ⓘ (new) ai-base-eastus-demo-02
[Create new](#)

[Review + create](#) < Previous Next : Storage



Configuring in Azure

- What do you need ...
 - Storage account
 - Credential Store
 - Microsoft-managed (preview)

Microsoft Azure Search resources, services, and docs (G+ /) Copilot

Home > Azure AI Foundry >

Azure AI hub

Create an Azure AI hub resource

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

Configure how your data is stored
As users work in Azure AI Foundry, uploaded data, stored credentials and generated artifacts like logs are stored.

Storage account * ⓘ (new) stazureaihub02 [Create new](#)

Credential store * ⓘ ☒ Microsoft-managed (preview) ⓘ [Learn more](#) ☐ Azure key vault

Key vault ⓘ [Create new](#)

Logs and docker images

Application insights ⓘ None [Create new](#)

Container registry ⓘ None [Create new](#)

[Review + create](#) < Previous Next : Networking



Configuring in Azure

- What do you need ...
 - Networking
 - Public

The screenshot shows the Microsoft Azure portal interface for configuring an Azure AI hub resource. The breadcrumb trail is 'Home > Azure AI Foundry > Azure AI hub'. The page title is 'Azure AI hub' with a subtitle 'Create an Azure AI hub resource'. The 'Networking' tab is selected, showing 'Network isolation' options. Three options are presented: 'Public' (selected with a red arrow), 'Private with Internet Outbound', and 'Private with Approved Outbound'. Each option lists its characteristics regarding workspace access, compute resource access, and outbound data movement. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next: Encryption'.

Microsoft Azure Search resources, services, and docs (G+/) Copilot

Home > Azure AI Foundry > Azure AI hub ...

Create an Azure AI hub resource

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

Network isolation

Projects associated to an Azure AI hub 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

Review + create < Previous Next: Encryption



Configuring in Azure

- What do you need ...
 - Encryption
 - Leave as is
 - Microsoft-managed keys

Microsoft Azure Search resources, services, and docs (G+/) Copilot

Home > Azure AI Foundry >

Azure AI hub

Create an Azure AI hub resource

Basics Storage Networking **Encryption** Identity 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 ☐

After workspace creation, you cannot change encryption key type between Microsoft-managed keys and Customer-managed keys.

[Review + create](#) [< Previous](#) [Next : Identity](#)



Configuring in Azure

- What do you need ...
 - Identity
 - Leave as is
 - System assigned identity
 - Credential-based access

The screenshot shows the 'Identity' tab of the 'Azure AI hub' configuration page in the Microsoft Azure portal. The page title is 'Azure AI hub' with a subtitle 'Create an Azure AI hub resource'. The navigation bar includes 'Basics', 'Storage', 'Networking', 'Encryption', 'Identity' (selected), 'Tags', and 'Review + create'. The 'Managed identity' section explains that it enables Azure resources to authenticate to cloud services without storing credentials. Under 'Identity type', 'System assigned identity' is selected with a red arrow pointing to it, and 'User assigned identity' is unselected. A warning message states: 'The managed user assigned identity option is only supported if an existing storage account, key vault, and container registry are used.' The 'Storage account access' section explains that Azure machine learning allows choosing between credential-based or identity-based access. Under 'Storage account access type', 'Credential-based access' is selected with a red arrow pointing to it, and 'Identity-based access' is unselected. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Tags'.



Configuring in Azure

■ Create it!

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

[Home](#) > [Azure AI Foundry](#) >

Azure AI hub

...

×

Create an Azure AI hub resource

✓ Validation passed

Basics

Storage

Networking

Encryption

Identity

Tags

Review + create

Basics

Subscription

Resource group

Name

Default project resource group

01-Lab_Joe_Ficara Student

(New) rg_aihub-02

ai-cscie94-hub-demo-02

rg_aiprojectdemos-02

Resources

Region

AI Services

Storage account

Key vault

Application insights

Container registry

East US

(new) ai-base-eastus-demo-02

(new) stazureaihub02

Microsoft-managed

None

None

Networking

Connectivity method

Network isolation

Enable public access from all networks

Public

Encryption

Encryption type

Microsoft-managed keys

Identity

Identity type

Storage account access type

Shared key access

System assigned

Credential-based

Enabled

Create

< Previous

Next >

[Download a template for automation](#)



Configuring in Azure

■ Create the Azure AI Project

The image displays two screenshots from the Microsoft Azure portal, illustrating the process of creating an Azure AI project.

Left Screenshot: Overview of the Azure AI Hub

- The page title is "ai-cscie94-hub-demo-01" (Azure AI hub).
- The left sidebar shows the "Overview" tab selected, with a list of settings including Activity log, Access control (IAM), Tags, Diagnose and solve problems, Events, Settings (Projects, Networking, Encryption, Properties, Locks), Monitoring (Alerts, Metrics, Diagnostic settings, Logs), and Automation (CLI / PS, Tasks).
- The main content area shows the "Essentials" section with the following details:
 - Resource group: [rg_aihuh](#)
 - Location: East US
 - Subscription: [01-Lab_Joe_Ficara_Student](#)
 - Subscription ID: 4ab55d34-0213-4090-a521-48cf35794fff
 - Key Vault: Microsoft-managed
 - Project resource group...: [rg_aiprojectdemos](#)
 - Storage: [staicscie94hub](#)
 - Container Registry (edit): ...
 - Application Insights: [appi-ai-cscie94-hub-demo-01](#) (edit)
 - Provisioning State: Succeeded
- Below the essentials, there is a section titled "Govern the environment for your team in AI Foundry" with a "Launch Azure AI Foundry" button.

Right Screenshot: Azure AI project creation form

- The page title is "Azure AI project".
- The "Basics" tab is selected, showing the "Organization" section.
- The "Organization" section prompts to "Select a subscription and resource group to organize this and other resources, inherit access, and attribute cost.".
- The "Subscription" dropdown is set to "01-Lab_Joe_Ficara_Student".
- The "Resource group" dropdown is set to "rg_aiprojectdemos" (indicated by a red arrow).
- The "Resource details" section includes:
 - "Name" dropdown set to "cscie94-demo-02" (indicated by a red arrow).
 - "Friendly name" text box containing "Cscie94 demo 02" (indicated by a red arrow).
- The "Share security, connectivity, compute" section prompts to "Projects are grouped by a hub, which provides security configurations, pre-configured connectivity with other Azure resources, compute, storage, and quota.".
- The "Hub" dropdown is set to "ai-cscie94-hub-demo-01 (eastus)" (indicated by a red arrow).
- At the bottom, there is a "Review + create" button and navigation links for "< Previous" and "Next : Identity".



Configuring in Azure

■ Create the Azure AI Project

Microsoft Azure

Home > ai-cscie94-hub-demo-01 >

Azure AI project

Organize and track work, collaborate with others and upload data. Access your work in Azure AI Foundry or Azure Machine Learning Studio.

Basics **Identity** Tags Review + create

Managed identity

A managed identity enables Azure resources to authenticate to cloud services without storing credentials in code. Once enabled, all necessary permissions can be granted via Azure role-based access control. A workspace can be given either a system assigned identity or a user assigned identity.

Identity type

☒ System assigned identity

☐ User assigned identity

Storage account access

Azure machine learning allows you to choose between credential-based or identity-based access when connecting to the default storage account.

Storage account access type

☒ Credential-based access

☐ Identity-based access

[Review + create](#) [< Previous](#) [Next : Tags](#)

Microsoft Azure

Home > ai-cscie94-hub-demo-01 >

Azure AI project

Organize and track work, collaborate with others and upload data. Access your work in Azure AI Foundry or Azure Machine Learning Studio.

✓ Validation passed

Basics **Identity** Tags **Review + create**

Basics

Subscription	01-Lab_Joe_Ficara Student
Resource group	rg_aiprojectdemos
Name	cscie94-demo-02
Friendly name	Cscie94 demo 02
Hub	ai-cscie94-hub-demo-01

Identity

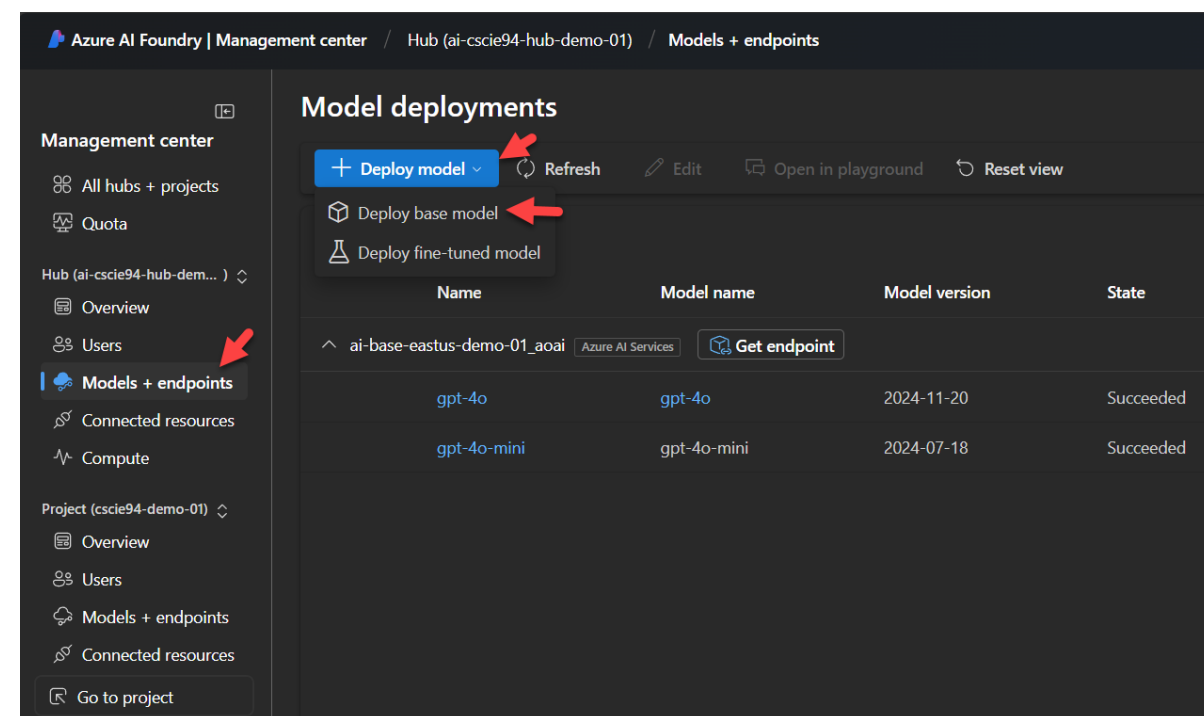
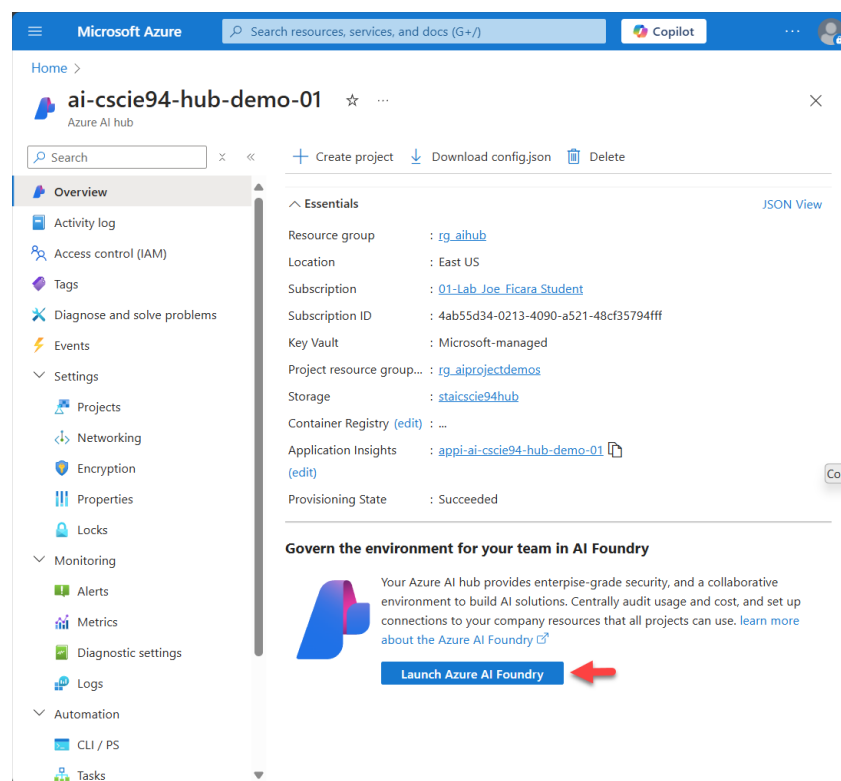
Identity type	System assigned
Storage account access type	Credential-based

[Create](#) [< Previous](#) [Next >](#) [Download a template for automation](#)



Configuring in Azure

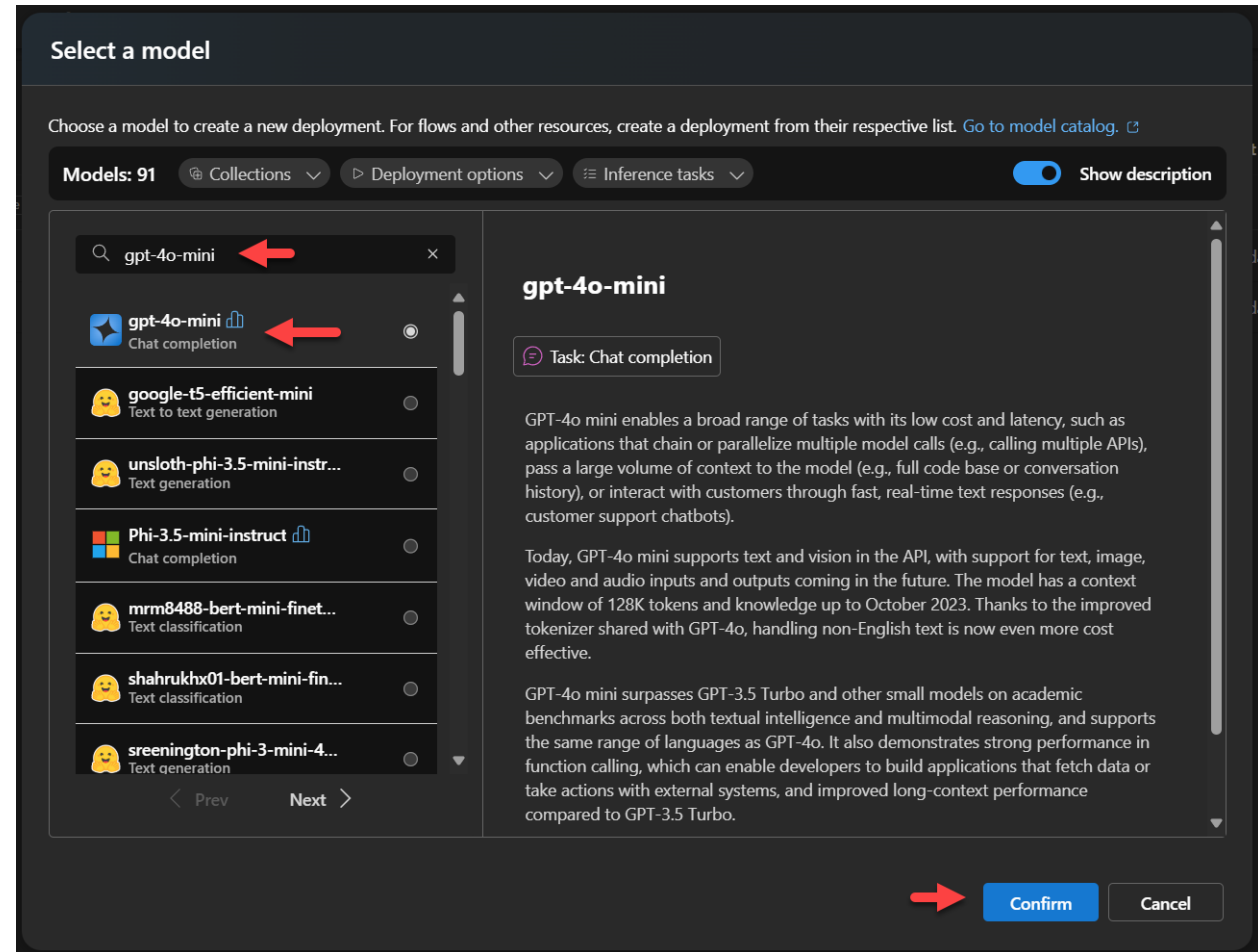
- Deploy a model
 - Choose Deploy base model





Configuring in Azure

- Deploy a model
 - Pick **gpt-4o-mini**







Configuring in Azure


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


Deploy model gpt-4o-mini


Deployment name * 


gpt-4o-mini-2 

Deployment type


Global Standard 

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	2024-07-18	Connected AI resource	ai-base-eastus-demo-01_aoi
AI hub	ai-cscie94-hub-demo-01	Authentication type	Key
Capacity	10K tokens per minute (TPM)	Resource location	East US
Content safety	DefaultV2		

 **Deploy** **Cancel**



Test in playground

- You can now experiment in the playground

Management center

- All hubs + projects
- Quota
- Hub (ai-cscie94-hub-dem...)
- Overview
- Users
- Models + endpoints
- Connected resources
- Compute
- Project (cscie94-demo-01)
- Overview
- Users
- Models + endpoints**
- Connected resources
- Go to project

Manage deployments of your models and services

Model deployments Service endpoints

+ Deploy model Refresh Edit Open in playground Reset view

Name	Model name	Model version	State
ai-base-eastus-demo-01_aoi	Azure AI Services	Get endpoint	
gpt-4o	gpt-4o	2024-11-20	Succeeded
gpt-4o-mini	gpt-4o-mini	2024-07-18	Succeeded



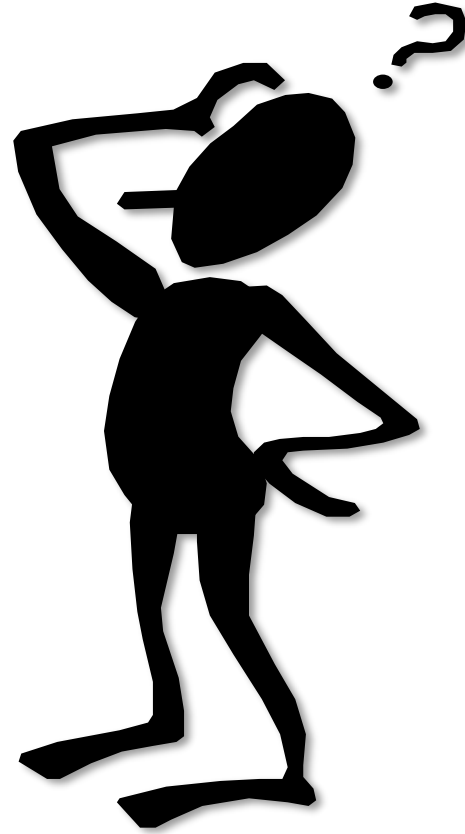
Demo

Create Azure AI Foundry, Project & Test deployment

Azure Portal



Configuring in Azure





Let's Code!

1010110101101010110101101010101011010110101010101101110110101010101101010101010
10101101101101010101011001
010101010010101010101011010011010101010101011010101010101010101010101010101010
1010110010101010110101101
01101010101010101010101010100010
101
010110
10
10011001
001
01001
01010101010010110011001
01010101111101
010101001
0101101010000101010101001
0101010101010101001
101
101
001
10001



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="9.0.1"·/>
····<PackageReference·Include="Swashbuckle.AspNetCore"·Version="7.2.0"·/>
····<PackageReference·Include="Azure.AI.OpenAI"·Version="2.1.0"·/>
····<PackageReference·Include="Microsoft.Extensions.AI"·Version="9.1.0-preview.1.25064.3"·/>
····<PackageReference·Include="Microsoft.Extensions.AI.Ollama"·Version="9.1.0-preview.1.25064.3"·/>
····<PackageReference·Include="Microsoft.Extensions.AI.OpenAI"·Version="9.1.0-preview.1.25064.3"·/>
····<PackageReference·Include="Microsoft.Extensions.Hosting"·Version="9.0.1"·/>
··</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 4o 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



Simple Chatbot

■ Uri and API Key

Azure AI Foundry | Management center / Project (cscie94-demo-01) / Connected resources

Manage connected resources in this project

+ New connection Refresh Delete Edit Reset view

Search Filter Columns

Name	Type	Target	Key	Authentication type	Access	Owner	Added by
ai-base-eastus-demo-01_...	Azure OpenAI Service	https://ai-base-eastus-demo-01.openai.azure.c...	API key	Shared to all projects	ai-cscie94-hub-demo-01	jficara2025@outlook.com
ai-base-eastus-demo-01	AI Services	https://ai-base-eastus-demo-01.cognitiveservi...	API key	Shared to all projects	ai-cscie94-hub-demo-01	jficara2025@outlook.com
cscie94-demo-01/worksp...	Azure Blob Storage	https://staicscie94hub.blob.core.windows.net/7...	--	Microsoft Entra ID	cscie94-demo-01	cscie94-demo-01	System
cscie94-demo-01/worksp...	Azure Blob Storage	https://staicscie94hub.blob.core.windows.net/7...	--	Microsoft Entra ID	cscie94-demo-01	cscie94-demo-01	System

CSCE-94 Joseph Ficara Copyright © 2015-2025 version 1.0.4



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**
 - 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...



Simple Chatbot

- Submit the prompt
 - Using **IChatClient** instance & **CompleteAsync** 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="9.0.1" />
• • • <PackageReference Include="Swashbuckle.AspNetCore" Version="7.2.0" />
• • • <PackageReference Include="Azure.AI.OpenAI" Version="2.1.0" />
• • • <PackageReference Include="Microsoft.Extensions.AI" Version="9.1.0-preview.1.25064.3" />
• • • <PackageReference Include="Microsoft.Extensions.AI.Ollama" Version="9.1.0-preview.1.25064.3" />
• • • <PackageReference Include="Microsoft.Extensions.AI.OpenAI" Version="9.1.0-preview.1.25064.3" />
• • • <PackageReference Include="Microsoft.Extensions.Hosting" Version="9.0.1" />
• • • <PackageReference Include="NJsonSchema" Version="11.1.0" />
• • </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 **CompleteAsync**
 - 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



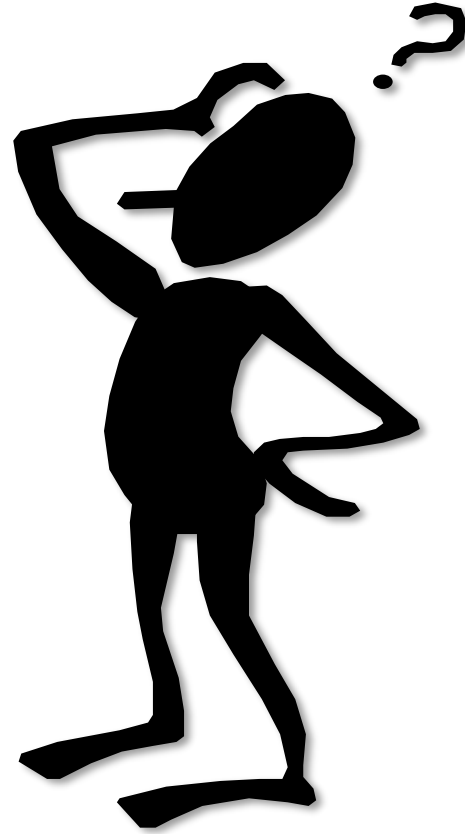
Demo

Structured Output JSON

StructuredOutputWebAPISolution.sln
StructuredOutputWebAPI.csproj



Questions





Further Reading

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