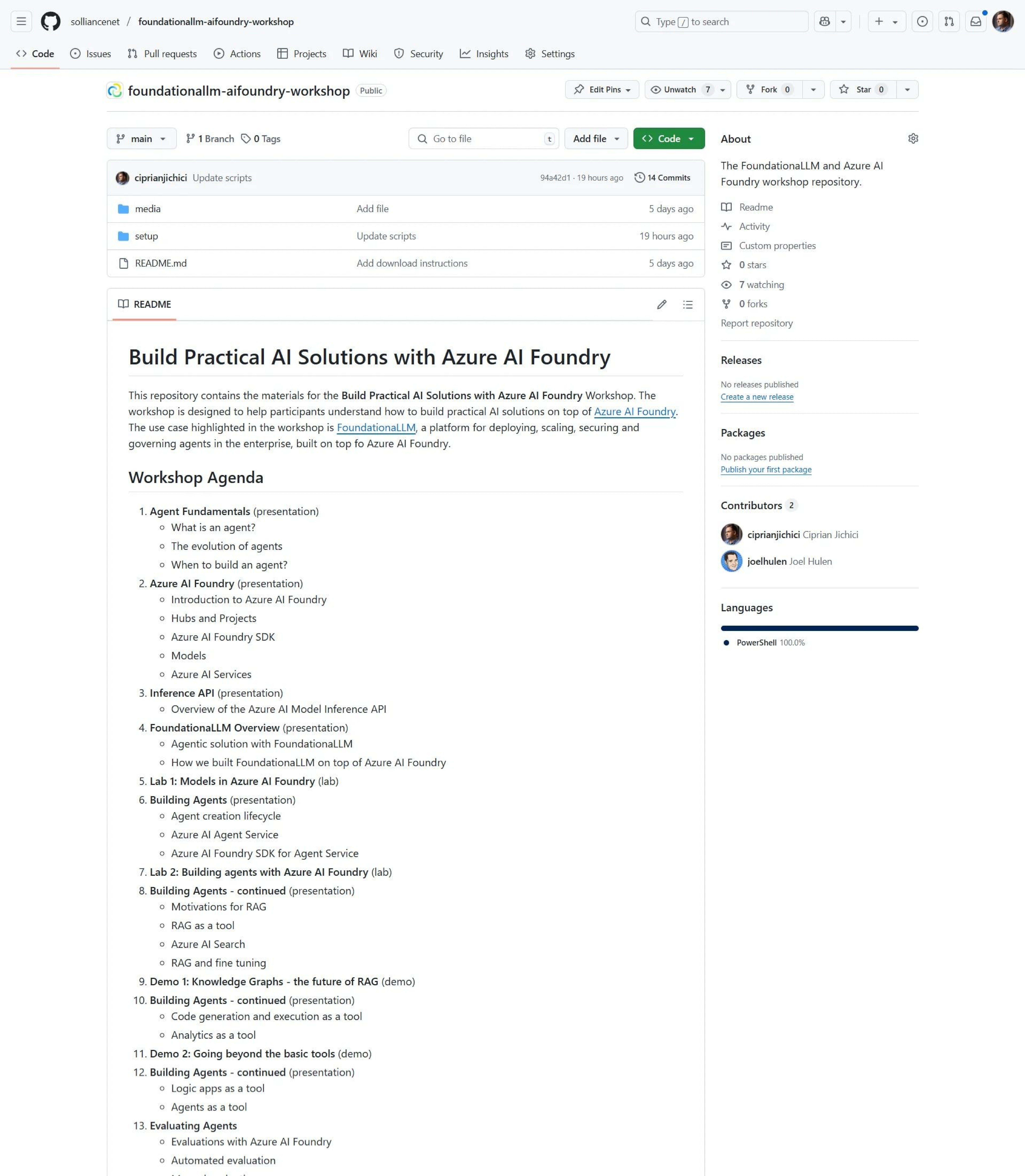


About

The FoundationaLLM and Azure AI Foundry workshop repository.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 2

ciprianjichici Ciprian Jichici
joelhulen Joel Hulen

Languages

PowerShell 100.0%

Build Practical AI Solutions with Azure AI Foundry

This repository contains the materials for the **Build Practical AI Solutions with Azure AI Foundry** Workshop. The workshop is designed to help participants understand how to build practical AI solutions on top of Azure AI Foundry. The use case highlighted in the workshop is FoundationaLLM, a platform for deploying, scaling, securing and governing agents in the enterprise, built on top fo Azure AI Foundry.

Workshop Agenda

1. **Agent Fundamentals** (presentation)
   - What is an agent?
   - The evolution of agents
   - When to build an agent?
2. **Azure AI Foundry** (presentation)
   - Introduction to Azure AI Foundry
   - Hubs and Projects
   - Azure AI Foundry SDK
   - Models
   - Azure AI Services
3. **Inference API** (presentation)
   - Overview of the Azure AI Model Inference API
4. **FoundationaLLM Overview** (presentation)
   - Agentic solution with FoundationaLLM
   - How we built FoundationaLLM on top of Azure AI Foundry
5. **Lab 1: Models in Azure AI Foundry** (lab)
6. **Building Agents** (presentation)
   - Agent creation lifecycle
   - Azure AI Agent Service
   - Azure AI Foundry SDK for Agent Service
7. **Lab 2: Building agents with Azure AI Foundry** (lab)
8. **Building Agents - continued** (presentation)
   - Motivations for RAG
   - RAG as a tool
   - Azure AI Search
   - RAG and fine tuning
9. **Demo 1: Knowledge Graphs - the future of RAG** (demo)
10. **Building Agents - continued** (presentation)
    - Code generation and execution as a tool
    - Analytics as a tool
11. **Demo 2: Going beyond the basic tools** (demo)
12. **Building Agents - continued** (presentation)
    - Logic apps as a tool
    - Agents as a tool
13. **Evaluating Agents**
    - Evaluations with Azure AI Foundry
    - Automated evaluation

- Manual evaluation
- Custom evaluators

14. **Responsible AI Tools & Practices**
    - Generative AI risks
    - Safguard with Trustworthy AI
    - Mitigation layers in Azure AI Foundry
15. **Use Cases and Open Q&A**

Labs and Demos

Lab 1: Models in Azure AI Foundry

- **Objective**: Understand the management of models in Azure AI Foundry, including Azure AI Model Inference API and evaluation.
- **Duration**: 30 minutes
- **Prerequisites**: Access to Azure AI Foundry.

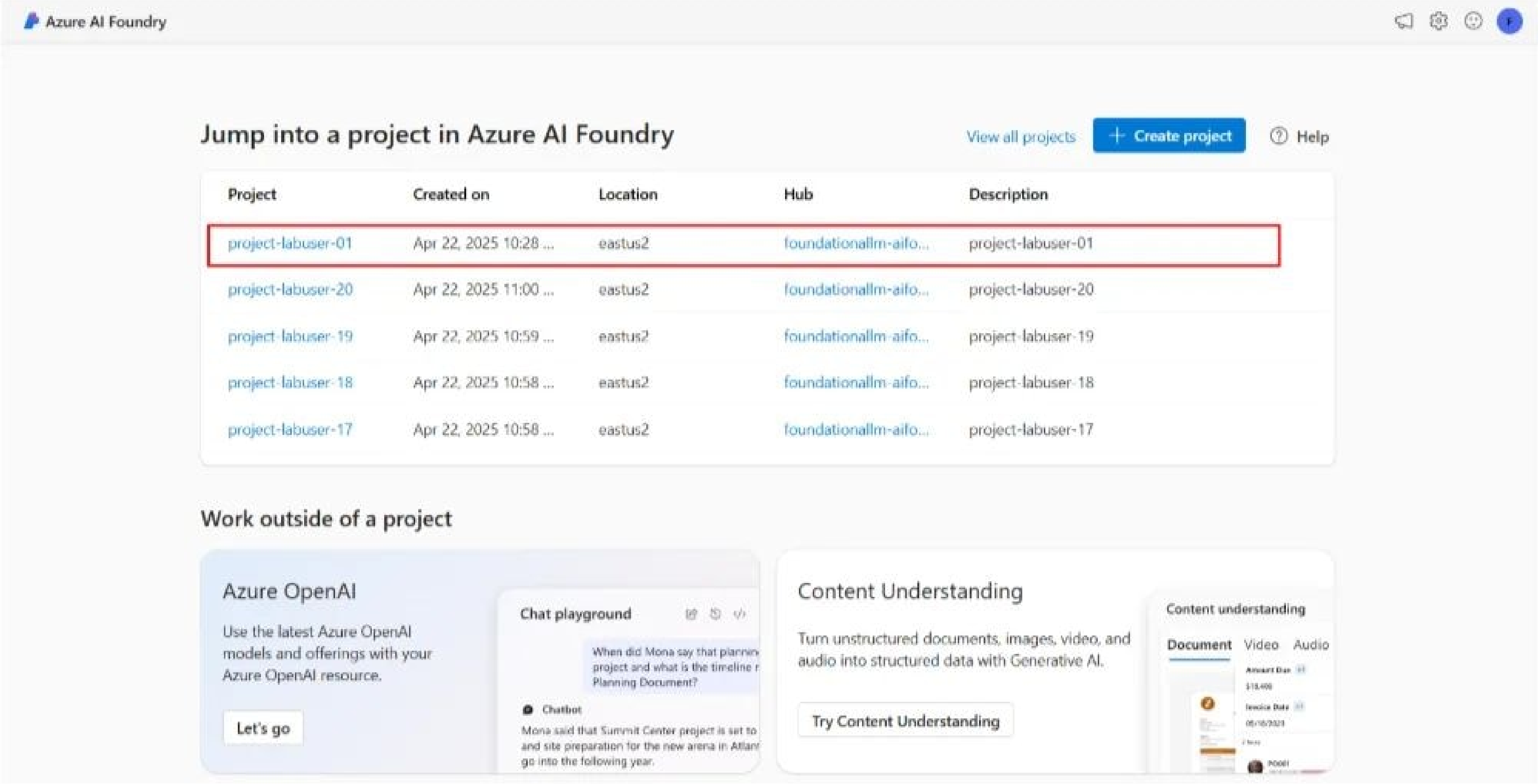
Exercise 1: A tour of the Azure AI Foundry Portal

In this exercise, you will tour the Azure AI Foundry Portal. You will be using your workshop account to access the portal.

> ⓘ **Note**
>
> Your workshop account should be `fllm-labuser-NN@foundationallm.ai` (where `NN` is your lab number assigned at the beginning of the workshop).

1. Login to the Azure AI Foundry Portal using your workshop account.

2. Navigate to the `project-labuser-NN` project that is displayed at the top of the screen.



Jump into a project in Azure AI Foundry

Project	Created on	Location	Hub	Description
project-labuser-01	Apr 22, 2025 10:28 ...	eastus2	foundationallm-aifo...	project-labuser-01
project-labuser-20	Apr 22, 2025 11:00 ...	eastus2	foundationallm-aifo...	project-labuser-20
project-labuser-19	Apr 22, 2025 10:59 ...	eastus2	foundationallm-aifo...	project-labuser-19
project-labuser-18	Apr 22, 2025 10:58 ...	eastus2	foundationallm-aifo...	project-labuser-18
project-labuser-17	Apr 22, 2025 10:58 ...	eastus2	foundationallm-aifo...	project-labuser-17

Work outside of a project

Azure OpenAI

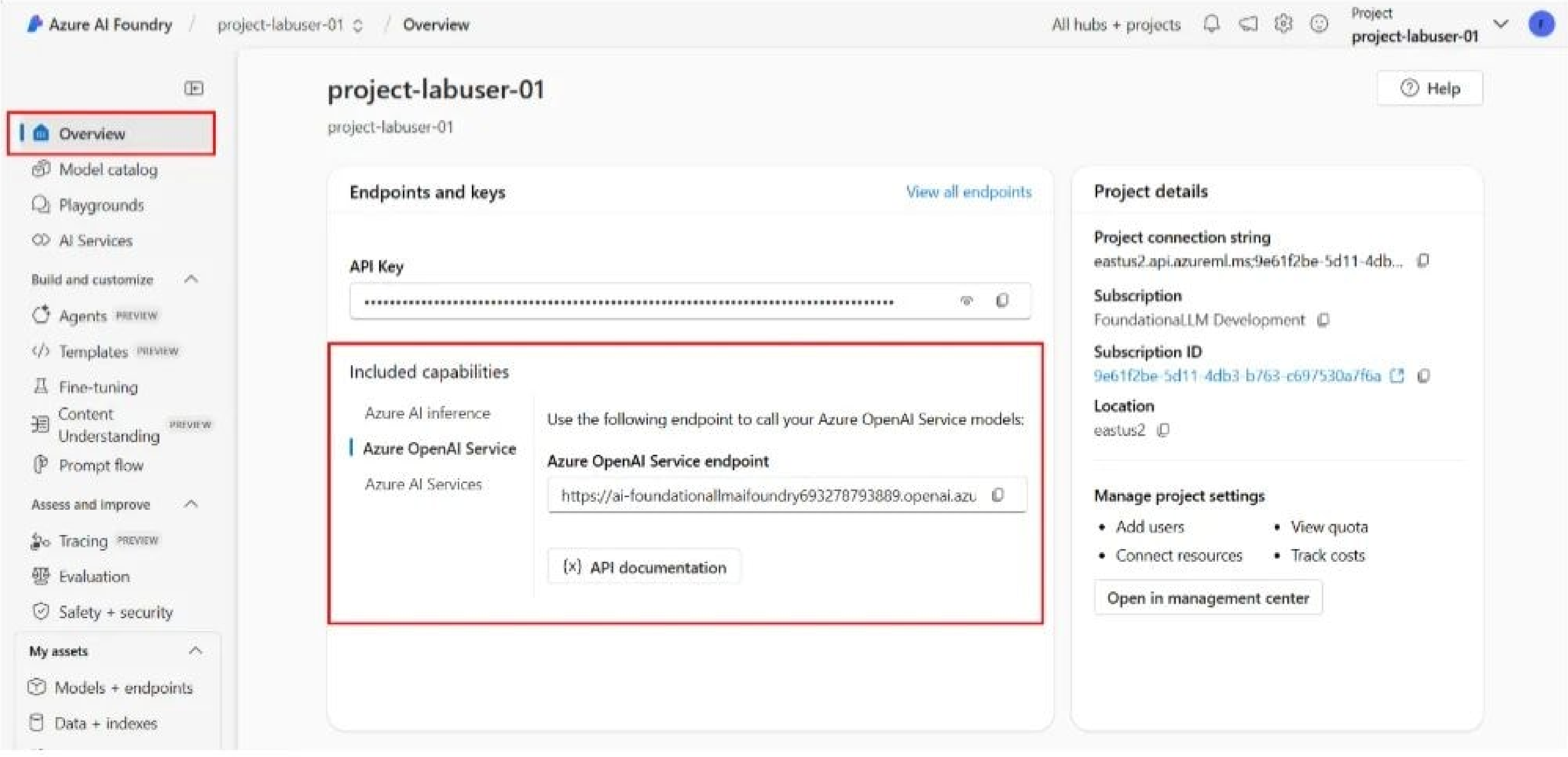
Use the latest Azure OpenAI models and offerings with your Azure OpenAI resource.

Let's go
Chat playground
When did Mona say that plannin project and what is the timeline r Planning Document?
Chatbot
Mona said that Summit Center project is set to and site preparation for the new arena in Atlan go into the following year.

Content Understanding

Turn unstructured documents, images, video, and audio into structured data with Generative AI.

3. Select the **Overview** section from the left navigation pane and explore the service endpoints listed under **Included capabilities**.



project-labuser-01

project-labuser-01

Endpoints and keys

API Key: ••

Included capabilities

Use the following endpoint to call your Azure OpenAI Service models:

Azure OpenAI Service endpoint: https://ai-foundationallmaifoundry693278793889.openai.azu

Project details

Project connection string: eastus2.api.azureml.ms;9e61f2be-5d11-4db...

Subscription: FoundationalLM Development

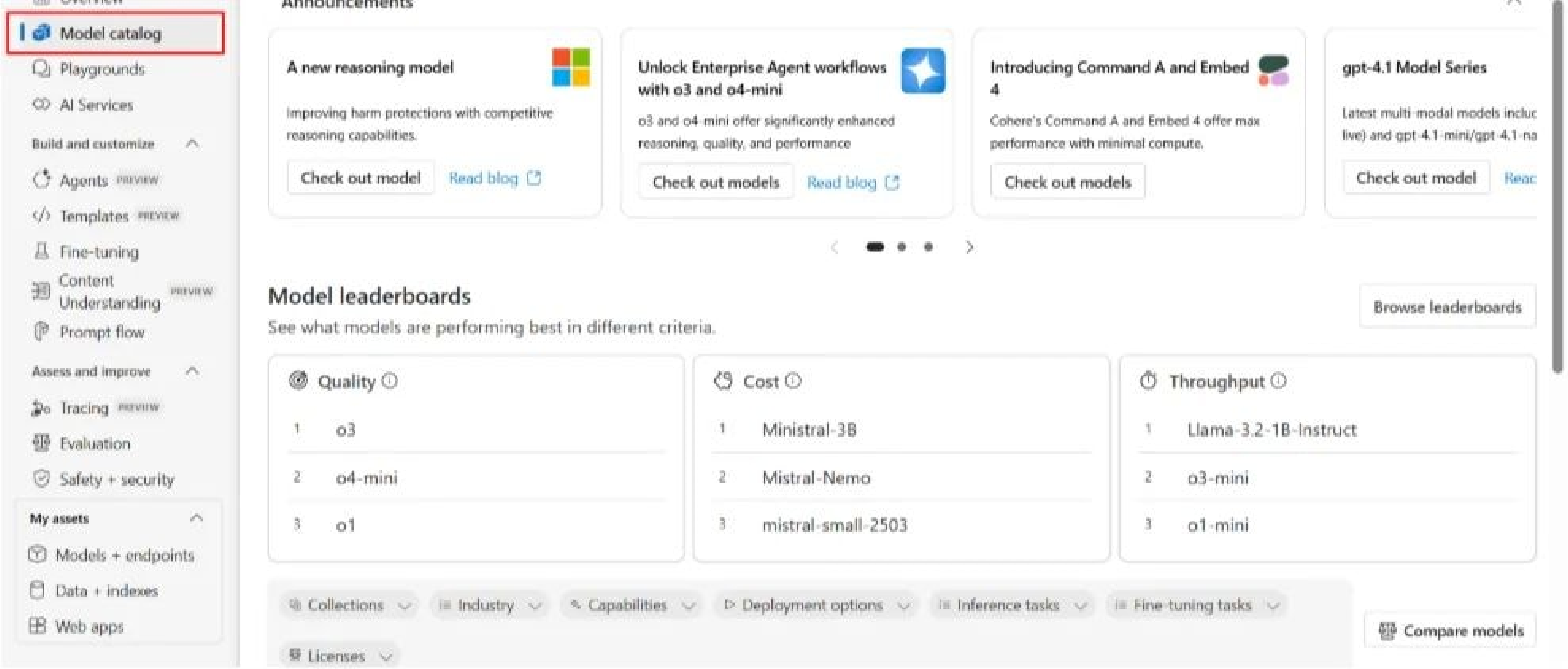
Subscription ID: 9e61f2be-5d11-4db3-b763-c697530a7f6a

Location: eastus2

Manage project settings
- Add users
- Connect resources
- View quota
- Track costs

4. Select the **Model catalog** section from the left navigation pane and explore the various models available within the project.

Find the right model to build your custom AI solution



A new reasoning model
Improving harm protections with competitive reasoning capabilities.

Unlock Enterprise Agent workflows with o3 and o4-mini
o3 and o4-mini offer significantly enhanced reasoning, quality, and performance

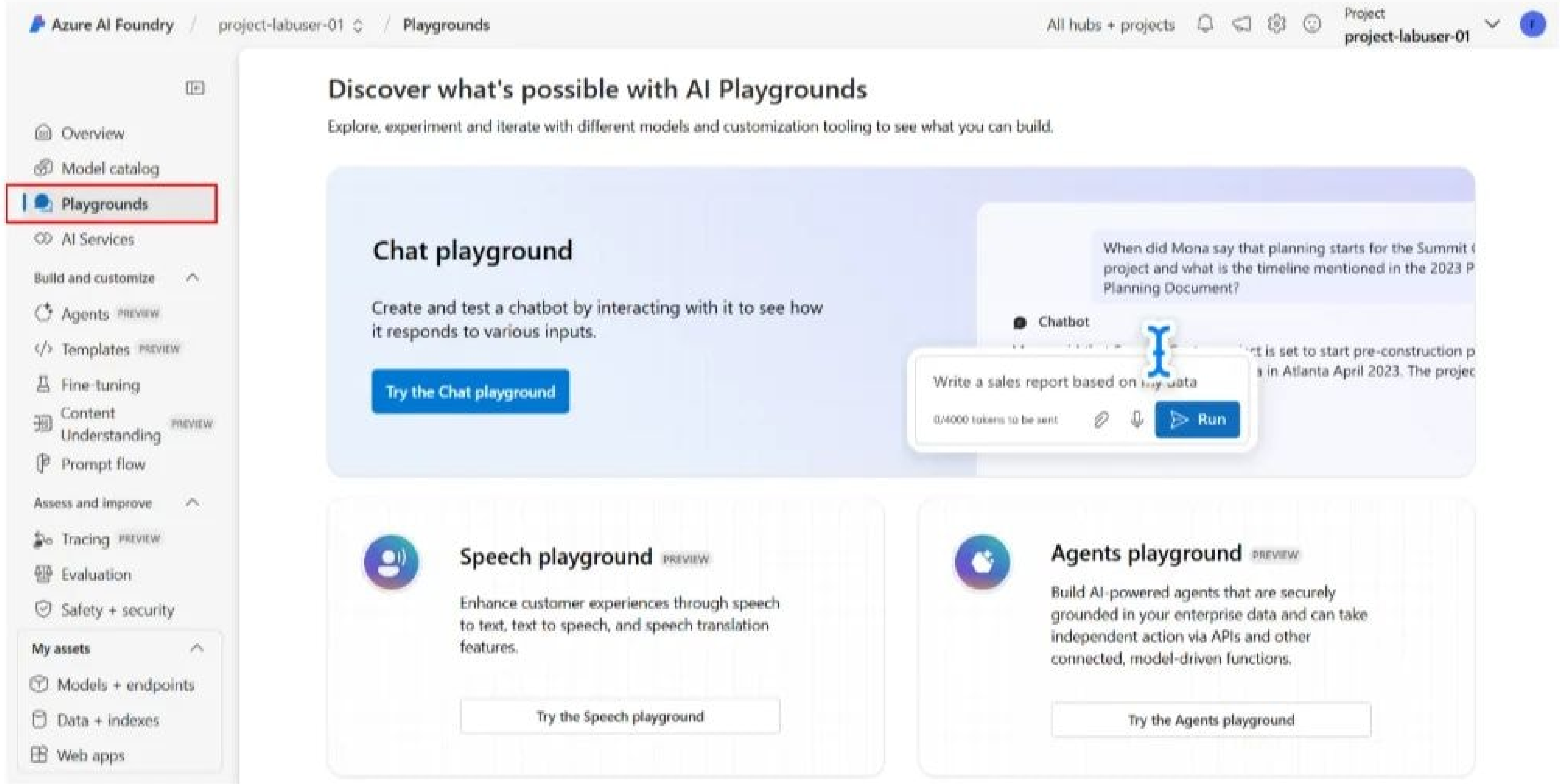
Introducing Command A and Embed 4
Cohere's Command A and Embed 4 offer max performance with minimal compute.

gpt-4.1 Model Series
Latest multi-modal models inclu live) and gpt-4.1-mini/gpt-4.1-na

Model leaderboards
See what models are performing best in different criteria.

	Quality		Cost		Throughput
1	o3	1	Ministral-3B	1	Llama-3.2-1B-Instruct
2	o4-mini	2	Mistral-Nemo	2	o3-mini
3	o1	3	mistral-small-2503	3	o1-mini

5. Select the **Playgrounds** section from the left navigation pane and explore the various playgrounds available within the project.



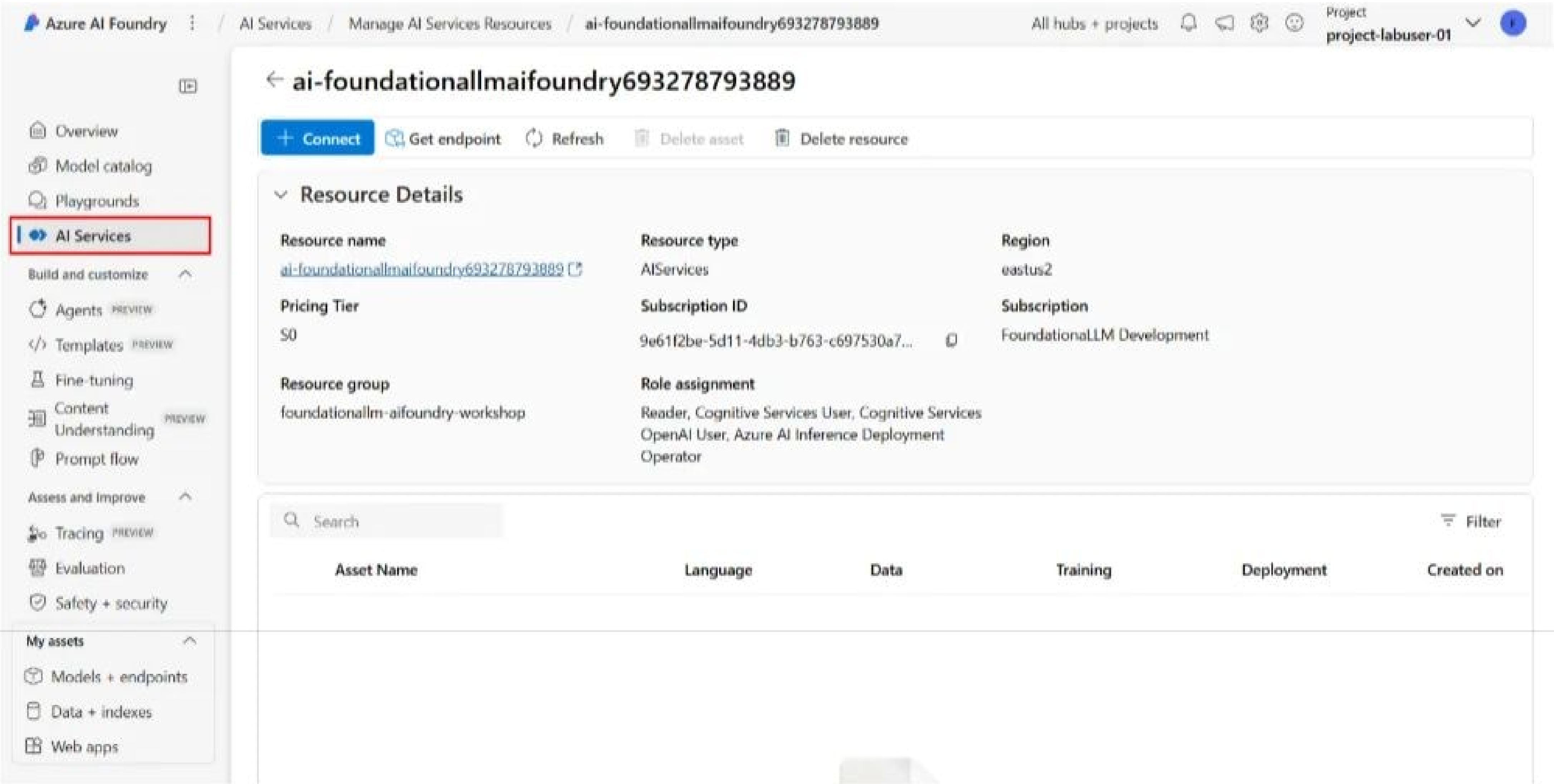
Discover what's possible with AI Playgrounds
Explore, experiment and iterate with different models and customization tooling to see what you can build.

Chat playground
Create and test a chatbot by interacting with it to see how it responds to various inputs.

Speech playground
Enhance customer experiences through speech to text, text to speech, and speech translation features.

Agents playground
Build AI-powered agents that are securely grounded in your enterprise data and can take independent action via APIs and other connected, model-driven functions.

6. Select the **AI Services** section from the left navigation pane. Select the **Manage AI Services** button from the top right corner of the screen. Note the AI Services instance named `ai-foundationallmaifoundry693278793889` that is connected to the project. Select its name to view the details of the connection.



ai-foundationallmaifoundry693278793889

Resource Details

Resource name
ai-foundationallmaifoundry693278793889

Resource type
AIServices

Region
eastus2

Pricing Tier
S0

Subscription ID
9e61f2be-5d11-4db3-b763-c697530a7...

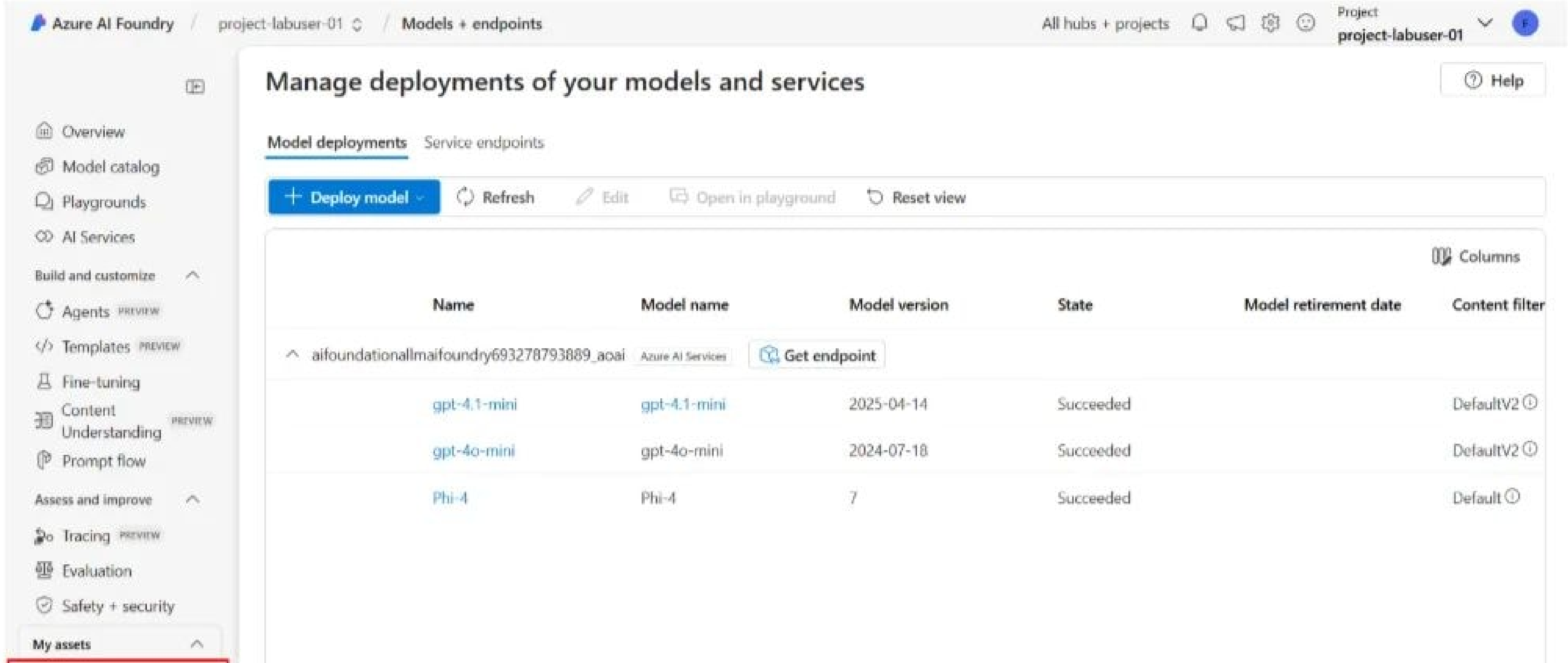
Subscription
FoundationaLLM Development

Resource group
foundationallm-aifoundry-workshop

Role assignment
Reader, Cognitive Services User, Cognitive Services OpenAI User, Azure AI Inference Deployment Operator

Asset Name	Language	Data	Training	Deployment	Created on

7. Select the **Models + endpoints** section from the left navigation pane and explore the various models and endpoints available within the project.



Manage deployments of your models and services

Name	Model name	Model version	State	Model retirement date	Content filter
aifoundationallmaifoundry693278793889_aoai					
gpt-4.1-mini	gpt-4.1-mini	2025-04-14	Succeeded		DefaultV2
gpt-4o-mini	gpt-4o-mini	2024-07-18	Succeeded		DefaultV2
Phi-4	Phi-4	7	Succeeded		Default

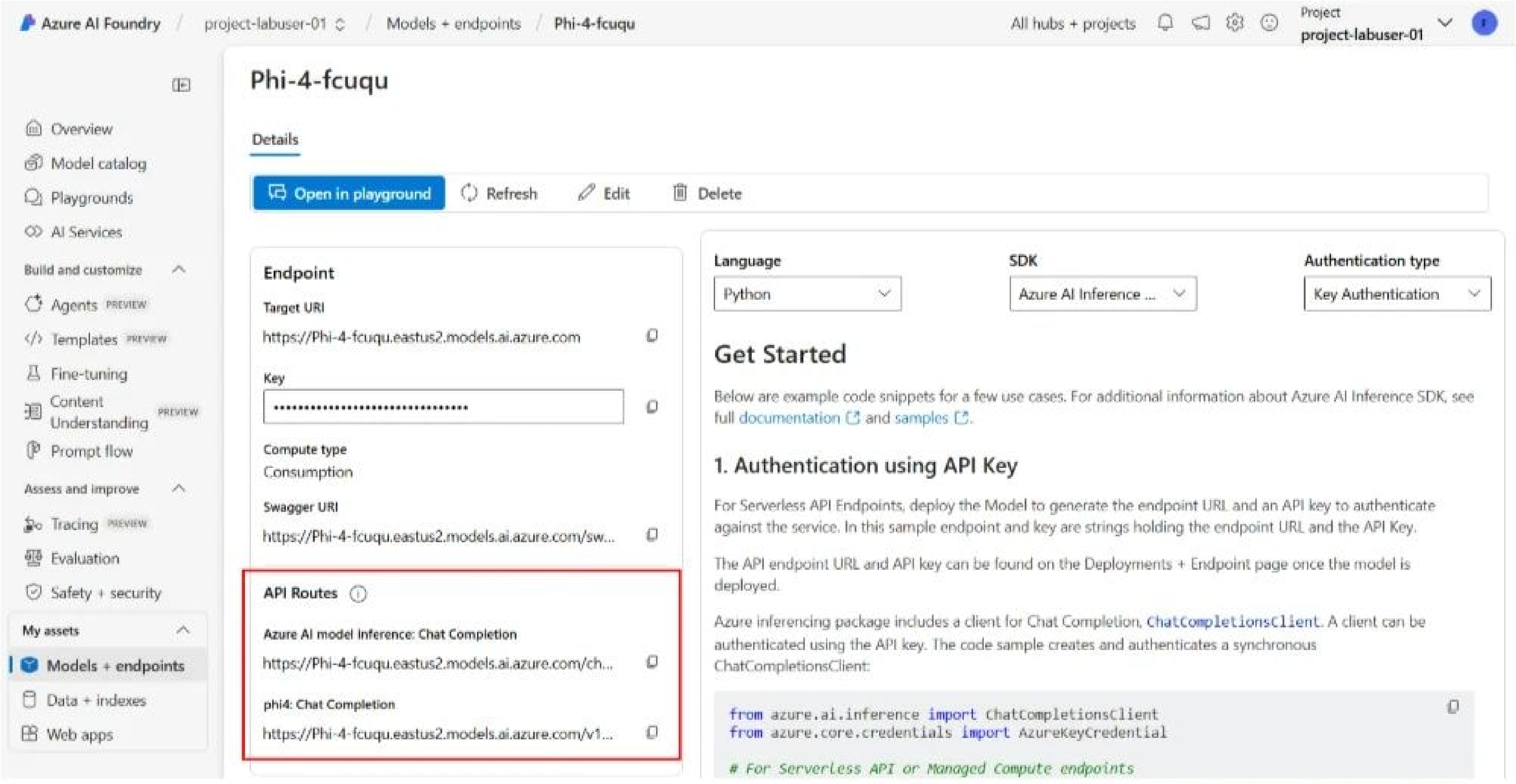
8. Spend a few more minutes exploring the rest of the sections in the left navigation pane.

Exercise 2: Using the Azure AI Model Inference API

In this exercise, you will deply a model and interact with it using the Azure AI Model Inference API. You will be using your workshop account to access the portal.

1. Login to the Azure AI Foundry Portal using your workshop account.

2. Select the **Models + endpoints** section from the left navigation pane and select the `Phi-4` model deployment. Explore the endpoint details and note how the URL `https://ai-foundationallmaifoundry693278793889.services.ai.azure.com/models/chat/completions?api-version=2024-05-01-preview` is agnostic to the model being used. This is because the Azure AI Model Inference API is designed to be a generic API that can be used with any model deployed in Azure AI Foundry.



Endpoint

Target URI
https://Phi-4-fcuqu.eastus2.models.ai.azure.com

Key
••••••••••••••••••••••••••••••••

Compute type
Consumption

Swagger URI
https://Phi-4-fcuqu.eastus2.models.ai.azure.com/sw...

API Routes

Azure AI model inference: Chat Completion
https://Phi-4-fcuqu.eastus2.models.ai.azure.com/ch...

phi4: Chat Completion
https://Phi-4-fcuqu.eastus2.models.ai.azure.com/v1...

Language: Python
SDK: Azure AI Inference ...
Authentication type: Key Authentication

Get Started

Below are example code snippets for a few use cases. For additional information about Azure AI Inference SDK, see full documentation and samples.

1. Authentication using API Key

For Serverless API Endpoints, deploy the Model to generate the endpoint URL and an API key to authenticate against the service. In this sample endpoint and key are strings holding the endpoint URL and the API Key.

The API endpoint URL and API key can be found on the Deployments + Endpoint page once the model is deployed.

Azure inferencing package includes a client for Chat Completion, ChatCompletionsClient. A client can be authenticated using the API key. The code sample creates and authenticates a synchronous ChatCompletionsClient:

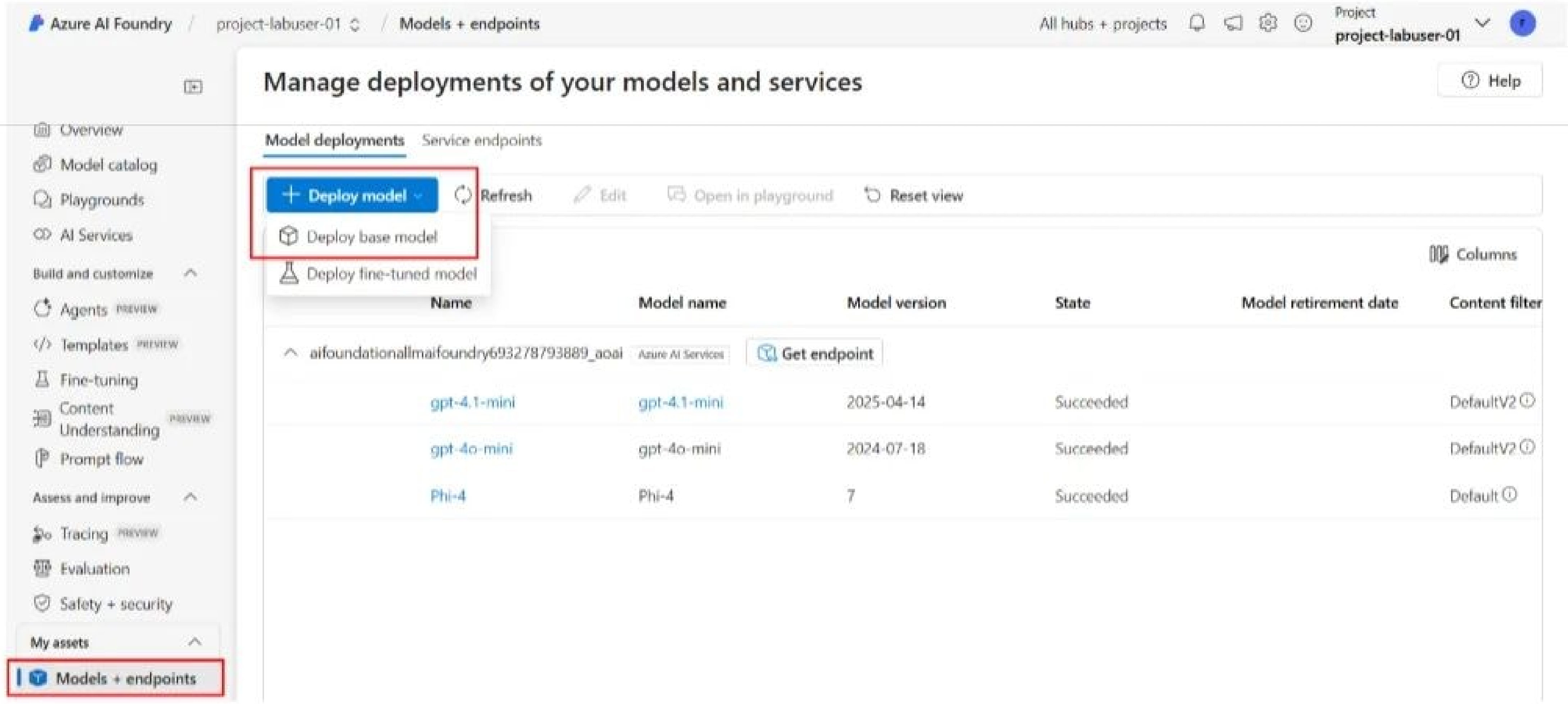
```
from azure.ai.inference import ChatCompletionsClient
from azure.core.credentials import AzureKeyCredential

# For Serverless API or Managed Compute endpoints
```

3. Explore the details on how to use the API with various programming languages. Note the code examples in various languages that demonstrate how to interact with the model using the Azure AI Model Inference API.

4. Select the **Open in playground** option to open the model in the playground. This will allow you to interact with the model and see how it responds to different questions.

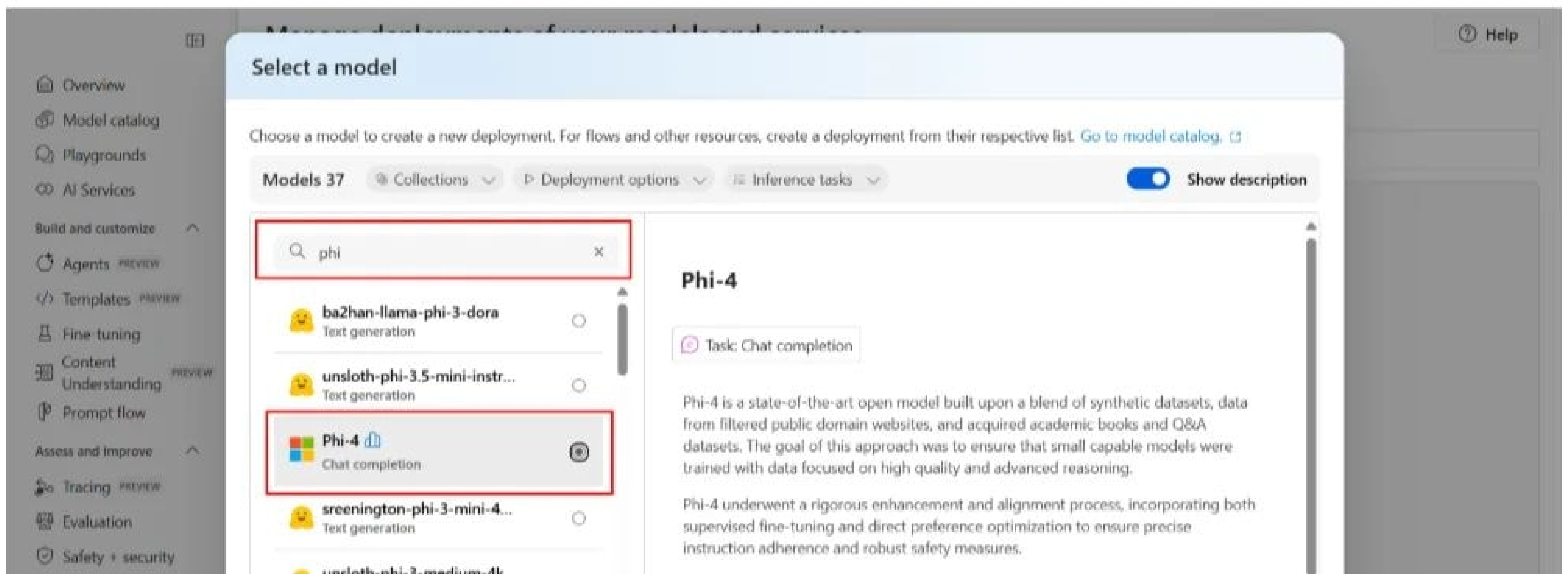
5. We will now deploy a new model. Return to the **Models + endpoints** section and select the **Deploy base model** option from the **+ Deploy model** dropdown.



Manage deployments of your models and services

Name	Model name	Model version	State	Model retirement date	Content filter
aifoundationallmaifoundry693278793889_aoai Azure AI Services Get endpoint					
gpt-4.1-mini	gpt-4.1-mini	2025-04-14	Succeeded		DefaultV2
gpt-4o-mini	gpt-4o-mini	2024-07-18	Succeeded		DefaultV2
Phi-4	Phi-4	7	Succeeded		Default

6. Enter the value `phi` in the search box, select the **Phi-4** model from the list of models, and select the **Confirm** button.



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.

phi

ba2han-llama-phi-3-dora
Text generation ○

unsloth-phi-3.5-mini-instr...
Text generation ○

Phi-4
Chat completion ●

sreenington-phi-3-mini-4...
Text generation ○

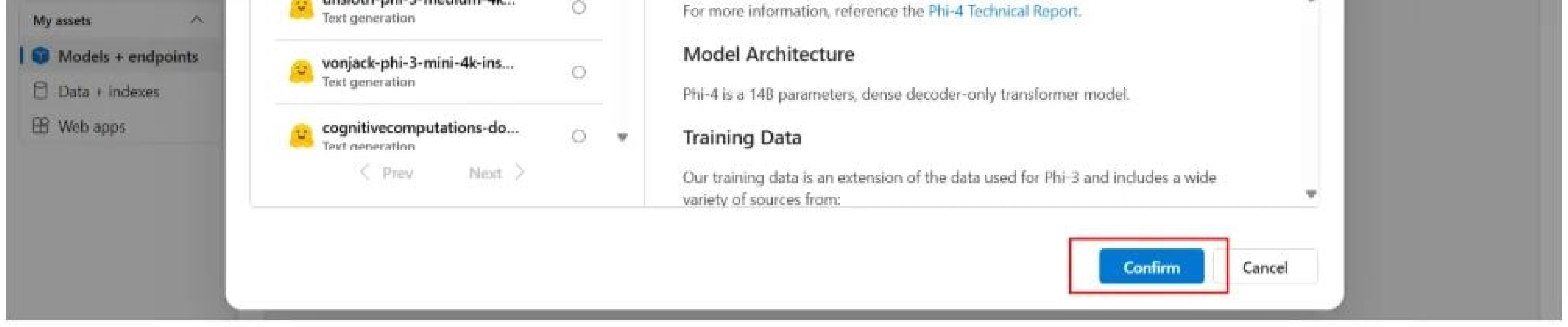
unsloth-phi-3-medium-4k

Phi-4

Task: Chat completion

Phi-4 is a state-of-the-art open model built upon a blend of synthetic datasets, data from filtered public domain websites, and acquired academic books and Q&A datasets. The goal of this approach was to ensure that small capable models were trained with data focused on high quality and advanced reasoning.

Phi-4 underwent a rigorous enhancement and alignment process, incorporating both supervised fine-tuning and direct preference optimization to ensure precise instruction adherence and robust safety measures.



For more information, reference the Phi-4 Technical Report.

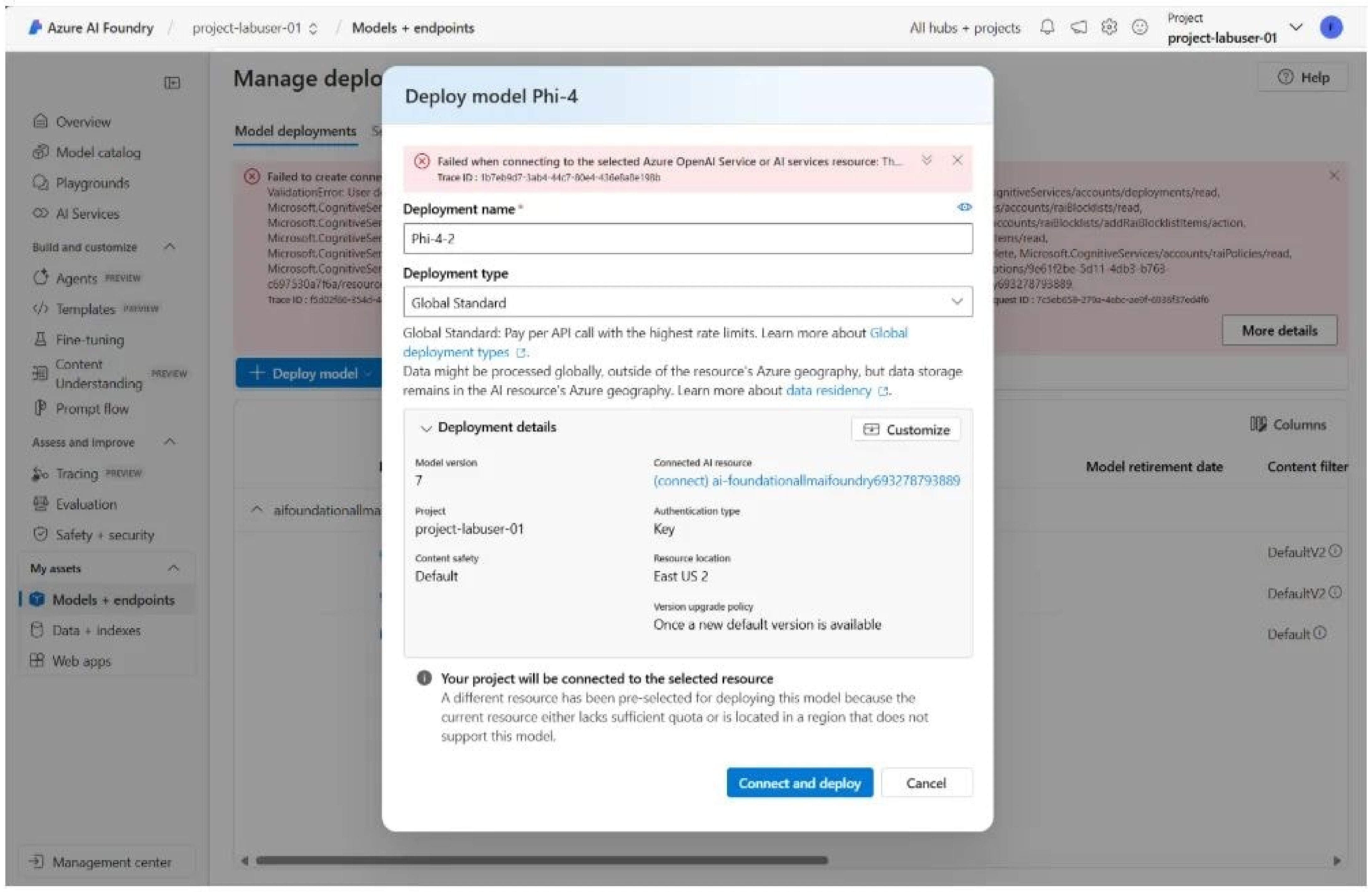
Model Architecture

Phi-4 is a 14B parameters, dense decoder-only transformer model.

Training Data

Our training data is an extension of the data used for Phi-3 and includes a wide variety of sources from:

7. Select the **Agree and Proceed** button to accept the terms and conditions.
8. Note the details of the deployment in the **Deploy model Phi-4** popup window. Select the **Connect and deploy** button to proceed with the deployment. After a short while, the deployment will fail with an error message stating that you do not have the required permissions to deploy the model. This is expected as the workshop account does not have the required permissions to deploy the model to the connected AI Services instance.



Deploy model Phi-4

Failed when connecting to the selected Azure OpenAI Service or AI services resource: Th...
Trace ID : 1b7eb9d7-3ab4-44c7-80e4-436e8a8e198b

Deployment name *
Phi-4-2

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

Model version
7

Connected AI resource
(connect) ai-foundationallmaifoundry693278793889

Project
project-labuser-01

Authentication type
Key

Content safety
Default

Resource location
East US 2

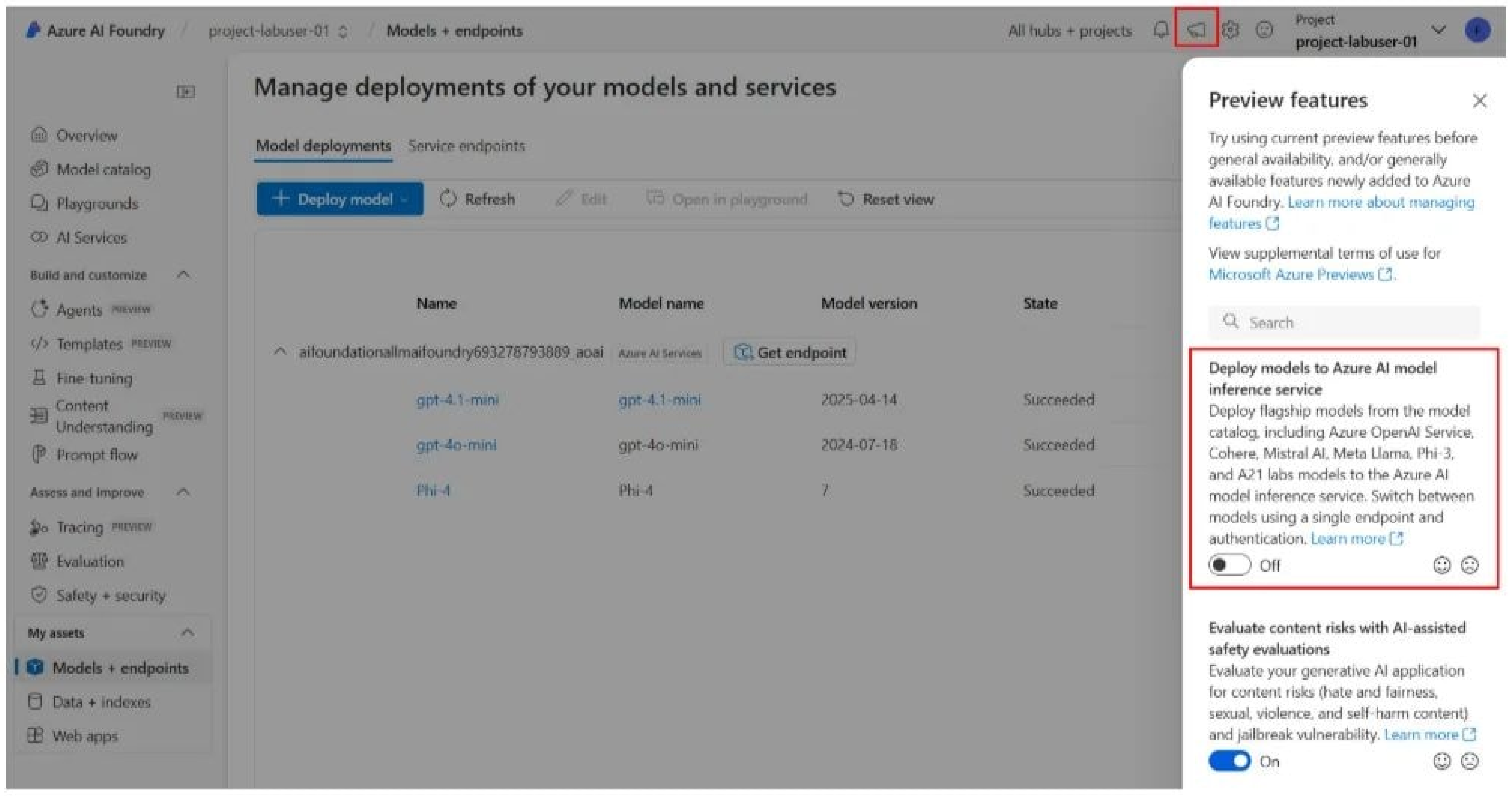
Version upgrade policy
Once a new default version is available

Your project will be connected to the selected resource
A different resource has been pre-selected for deploying this model because the current resource either lacks sufficient quota or is located in a region that does not support this model.

Connect and deploy | Cancel

> **Note**
>
> In real world scenarios, you will need to separate permissions between different teams from your organization, or even between members of the same team. To provide a simple example of how this works, your lab user was not granted the required permissions to deploy models at the hub level (this is intentional). The user's permissions are limited to the project level. Next, we will deploy the model at the project level.

9. By default, your project is configured to attempt to deploy models to the inference service of the AI Services instance that is connected to the project (the one that your user does not have permissions to deploy to). To change this, select the **Preview features** option from the top right corner of the screen and then make sure to turn off the **Deploy models to the Azure AI model inference service** option. This will allow you to deploy models to the project level.



Manage deployments of your models and services

Name	Model name	Model version	State
aifoundationallmaifoundry693278793889_aoai	Azure AI Services		
gpt-4.1-mini	gpt-4.1-mini	2025-04-14	Succeeded
gpt-4o-mini	gpt-4o-mini	2024-07-18	Succeeded
Phi-4	Phi-4	7	Succeeded

Preview features

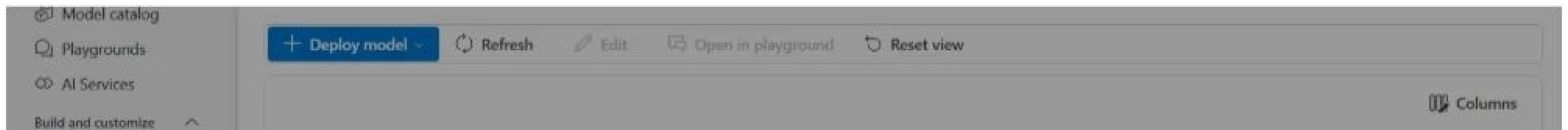
Try using current preview features before general availability, and/or generally available features newly added to Azure AI Foundry. Learn more about managing features

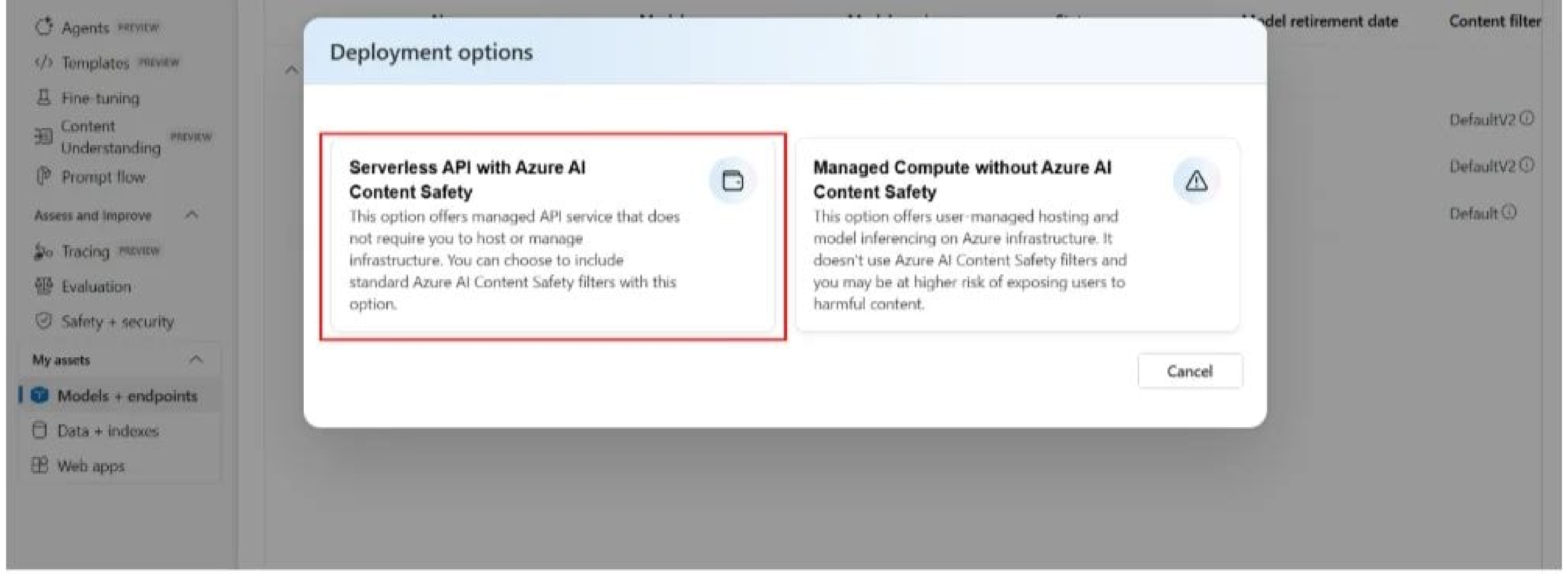
View supplemental terms of use for Microsoft Azure Previews.

Deploy models to Azure AI model inference service
Deploy flagship models from the model catalog, including Azure OpenAI Service, Cohere, Mistral AI, Meta Llama, Phi-3, and A21 labs models to the Azure AI model inference service. Switch between models using a single endpoint and authentication. Learn more
Off

Evaluate content risks with AI-assisted safety evaluations
Evaluate your generative AI application for content risks (hate and fairness, sexual, violence, and self-harm content) and jailbreak vulnerability. Learn more
On

10. Perform again the actions from steps 5 and 6. Note how the option has changed now. Select **Serverless API with Azure AI Content Safety** option to deploy the model. This option is available because the project is configured to deploy models to the project level.





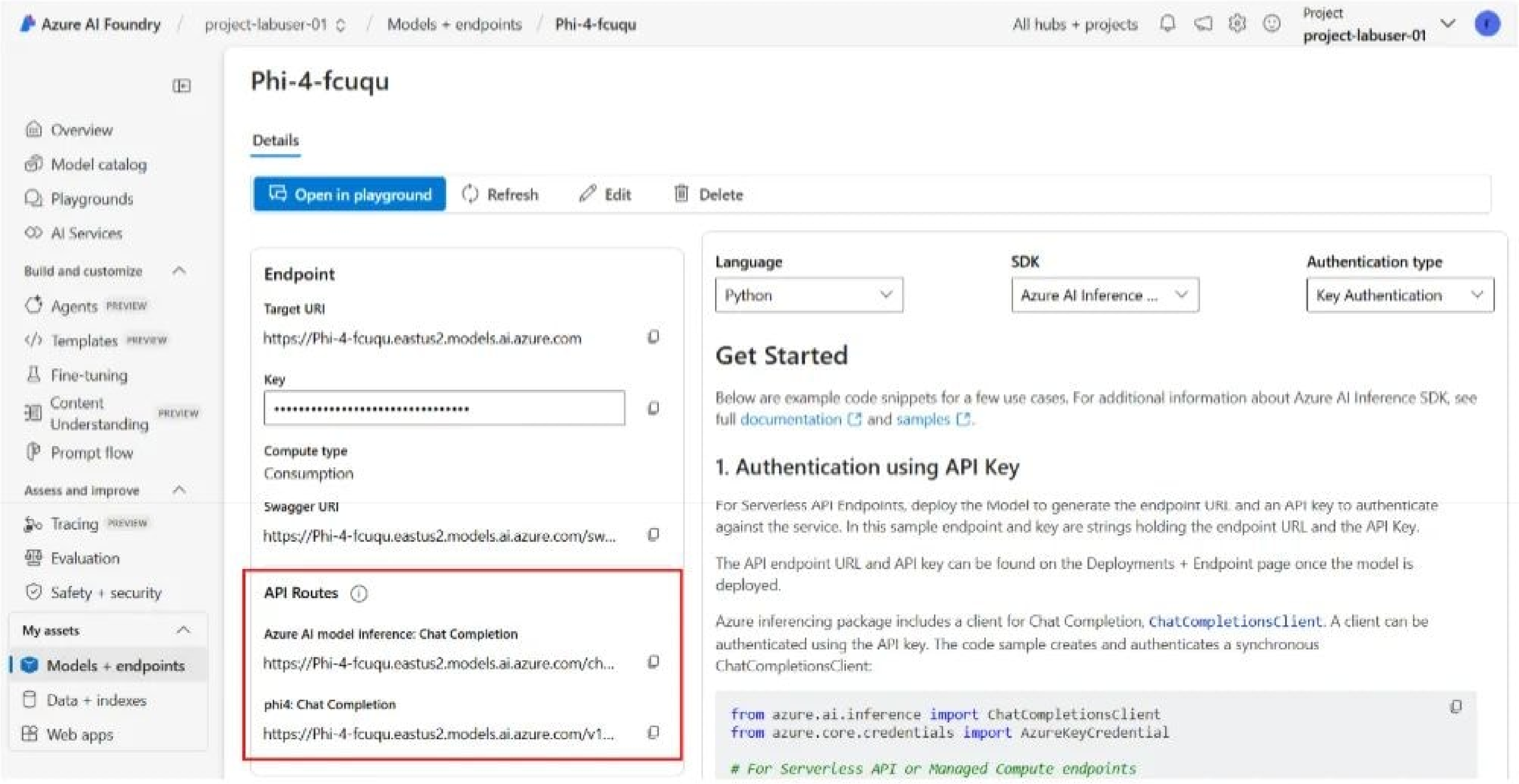
Deployment options

Serverless API with Azure AI Content Safety
This option offers managed API service that does not require you to host or manage infrastructure. You can choose to include standard Azure AI Content Safety filters with this option.

Managed Compute without Azure AI Content Safety
This option offers user-managed hosting and model inferencing on Azure infrastructure. It doesn't use Azure AI Content Safety filters and you may be at higher risk of exposing users to harmful content.

11. Select the **Deploy** button to deploy the model.

12. View the deployment status of the model in the **Models + endpoints** section. The deployment may take a few minutes to complete. Note the code examples in various languages that demonstrate how to interact with the model using the Azure AI Model Inference API. Note the **API Routes** section and how this deployment option also provides an Azure AI model inference API endpoint (you should see a URL that is similar to `https://Phi-4-fcuqu.eastus2.models.ai.azure.com/chat/completions`).



Phi-4-fcuqu

Endpoint

Target URI
https://Phi-4-fcuqu.eastus2.models.ai.azure.com

Key
••••••••••••••••••••••••••••••••

Compute type
Consumption

Swagger URI
https://Phi-4-fcuqu.eastus2.models.ai.azure.com/sw...

API Routes

Azure AI model inference: Chat Completion
https://Phi-4-fcuqu.eastus2.models.ai.azure.com/ch...

phi4: Chat Completion
https://Phi-4-fcuqu.eastus2.models.ai.azure.com/v1...

Language: Python
SDK: Azure AI Inference ...
Authentication type: Key Authentication

Get Started

Below are example code snippets for a few use cases. For additional information about Azure AI Inference SDK, see full documentation and samples.

1. Authentication using API Key

For Serverless API Endpoints, deploy the Model to generate the endpoint URL and an API key to authenticate against the service. In this sample endpoint and key are strings holding the endpoint URL and the API Key.

The API endpoint URL and API key can be found on the Deployments + Endpoint page once the model is deployed.

Azure inferencing package includes a client for Chat Completion, `ChatCompletionsClient`. A client can be authenticated using the API key. The code sample creates and authenticates a synchronous ChatCompletionsClient:

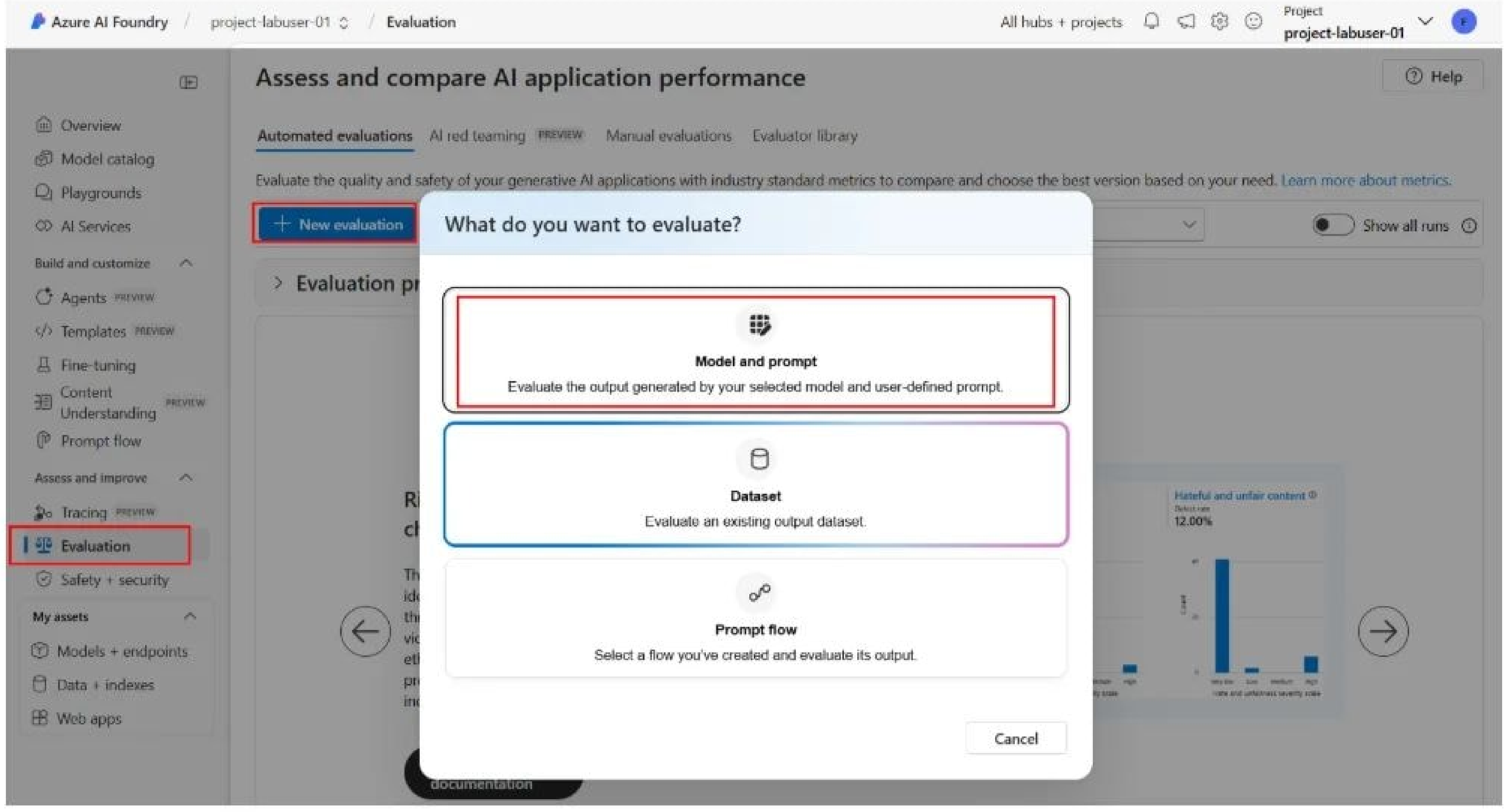
```
from azure.ai.inference import ChatCompletionsClient
from azure.core.credentials import AzureKeyCredential

# For Serverless API or Managed Compute endpoints
```

Exercise 3: Evaluate model performance

In this exercise, you will evaluate the performance of the model using the builtin capabilities for Azure AI Foundry. You will be using your workshop account to access the portal.

1. Select the **Evaluation** section from the left navigation pane and select the select the **+ New evaluation** button. Select the **Model and prompt** option.



Assess and compare AI application performance

Evaluate the quality and safety of your generative AI applications with industry standard metrics to compare and choose the best version based on your need. Learn more about metrics.

What do you want to evaluate?

Model and prompt
Evaluate the output generated by your selected model and user-defined prompt.

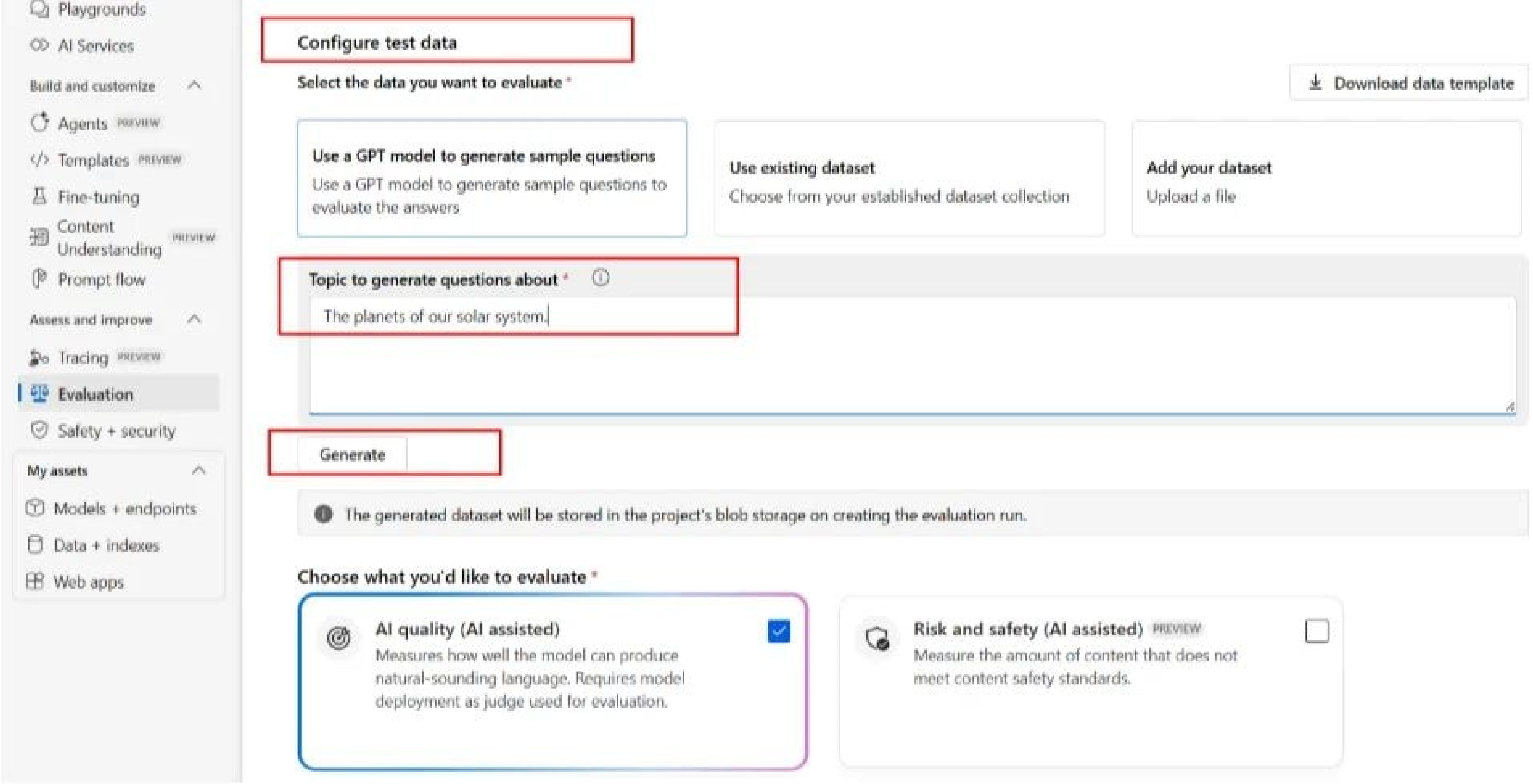
Dataset
Evaluate an existing output dataset.

Prompt flow
Select a flow you've created and evaluate its output.

2. Note the model that is selected for evaluation (`gpt-4.1-mini`). Navigate to the **Configure test data** section and enter the following text in the **Topic to generate questions about** field:

```
The planets of our solar system.
```

Select the **Generate** buton and note how the system generates a set of questions based on the topic.



Configure test data

Select the data you want to evaluate *

Download data template

Use a GPT model to generate sample questions
Use a GPT model to generate sample questions to evaluate the answers

Use existing dataset
Choose from your established dataset collection

Add your dataset
Upload a file

Topic to generate questions about *

The planets of our solar system.

Generate

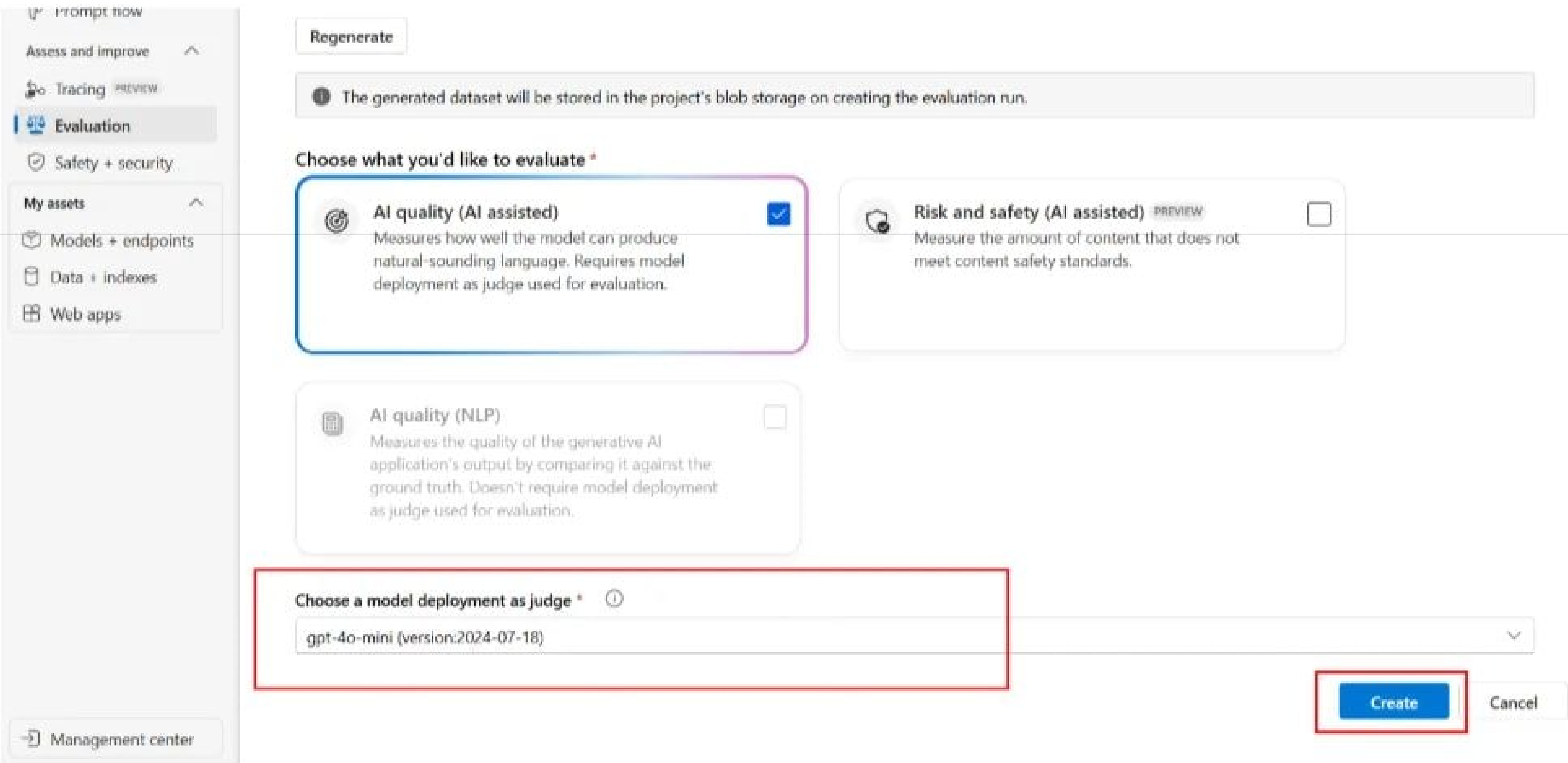
The generated dataset will be stored in the project's blob storage on creating the evaluation run.

Choose what you'd like to evaluate *

☑ AI quality (AI assisted)
Measures how well the model can produce natural-sounding language. Requires model deployment as judge used for evaluation.

☐ Risk and safety (AI assisted) PREVIEW
Measure the amount of content that does not meet content safety standards.

3. Under the **Choose a model deployment as judge** section, select the `gpt-4o-mini` model deployment and then select the **Create** button.



Regenerate

The generated dataset will be stored in the project's blob storage on creating the evaluation run.

Choose what you'd like to evaluate *

☑ AI quality (AI assisted)
Measures how well the model can produce natural-sounding language. Requires model deployment as judge used for evaluation.

☐ Risk and safety (AI assisted) PREVIEW
Measure the amount of content that does not meet content safety standards.

☐ AI quality (NLP)
Measures the quality of the generative AI application's output by comparing it against the ground truth. Doesn't require model deployment as judge used for evaluation.

Choose a model deployment as judge *

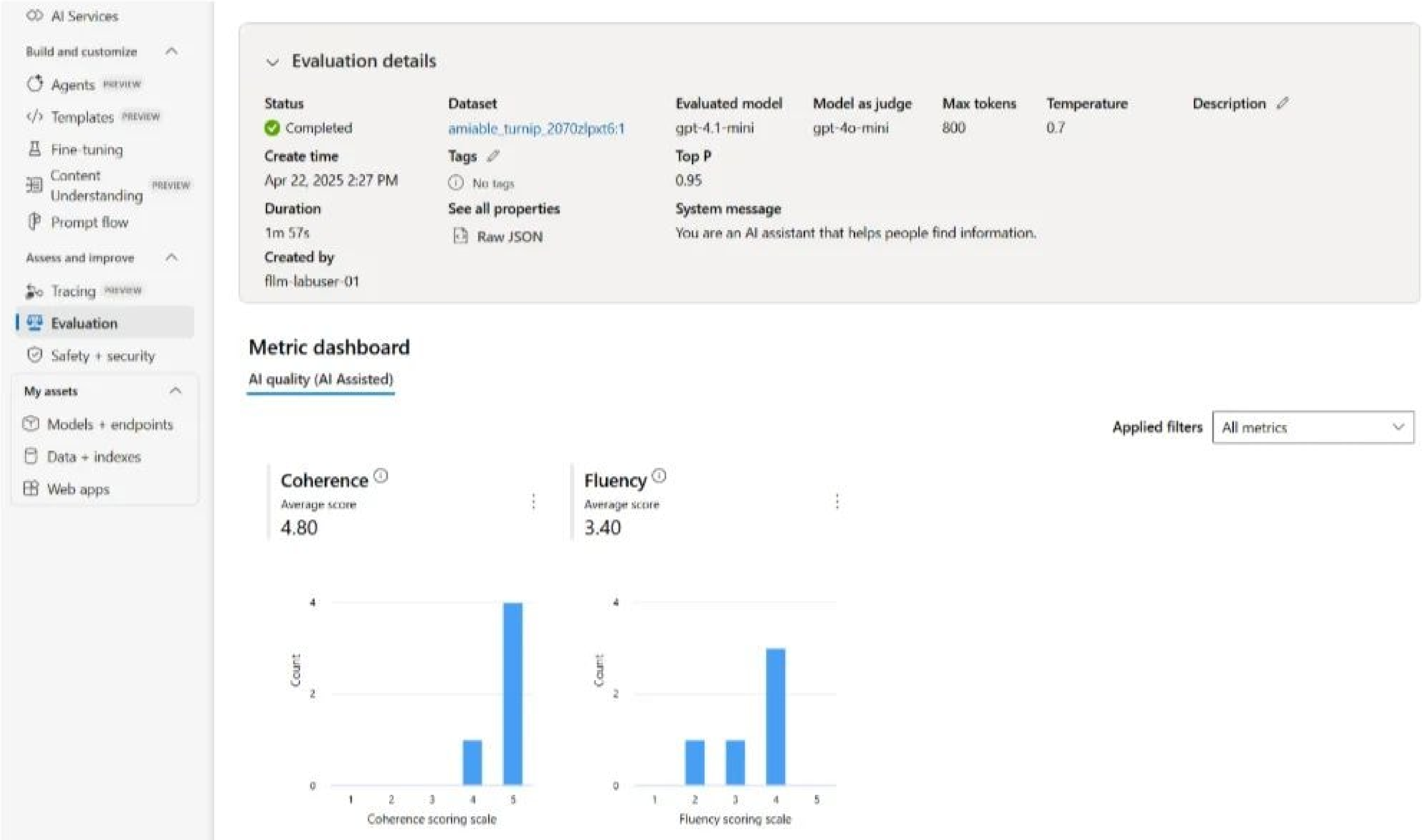
gpt-4o-mini (version:2024-07-18)

Create

Cancel

4. Wait until the evaluation is completed. This may take a few minutes. Once the evaluation is completed, review the report.

> [!NOTE] The evaluation might take several minutes to complete. If the evaluation is still running after a few minutes, you can skip this step and come back to it later (e.g. at the beginning of the next lab).



Evaluation details

Status	Dataset	Evaluated model	Model as judge	Max tokens	Temperature	Description
Completed	amiable_turnip_2070zlpxt6:1	gpt-4.1-mini	gpt-4o-mini	800	0.7	

Create time: Apr 22, 2025 2:27 PM
Tags: No tags
Top P: 0.95
Duration: 1m 57s
See all properties: Raw JSON
System message: You are an AI assistant that helps people find information.
Created by: fllm-labuser-01

Metric dashboard

AI quality (AI Assisted)

Applied filters: All metrics

Coherence
Average score
4.80

Fluency
Average score
3.40

You have now completed the lab. You have learned how to use the Azure AI Model Inference API to deploy and interact with models in Azure AI Foundry. You have also learned how to evaluate the performance of models using the builtin capabilities for Azure AI Foundry.

Lab 2: Building agents with Azure AI Foundry

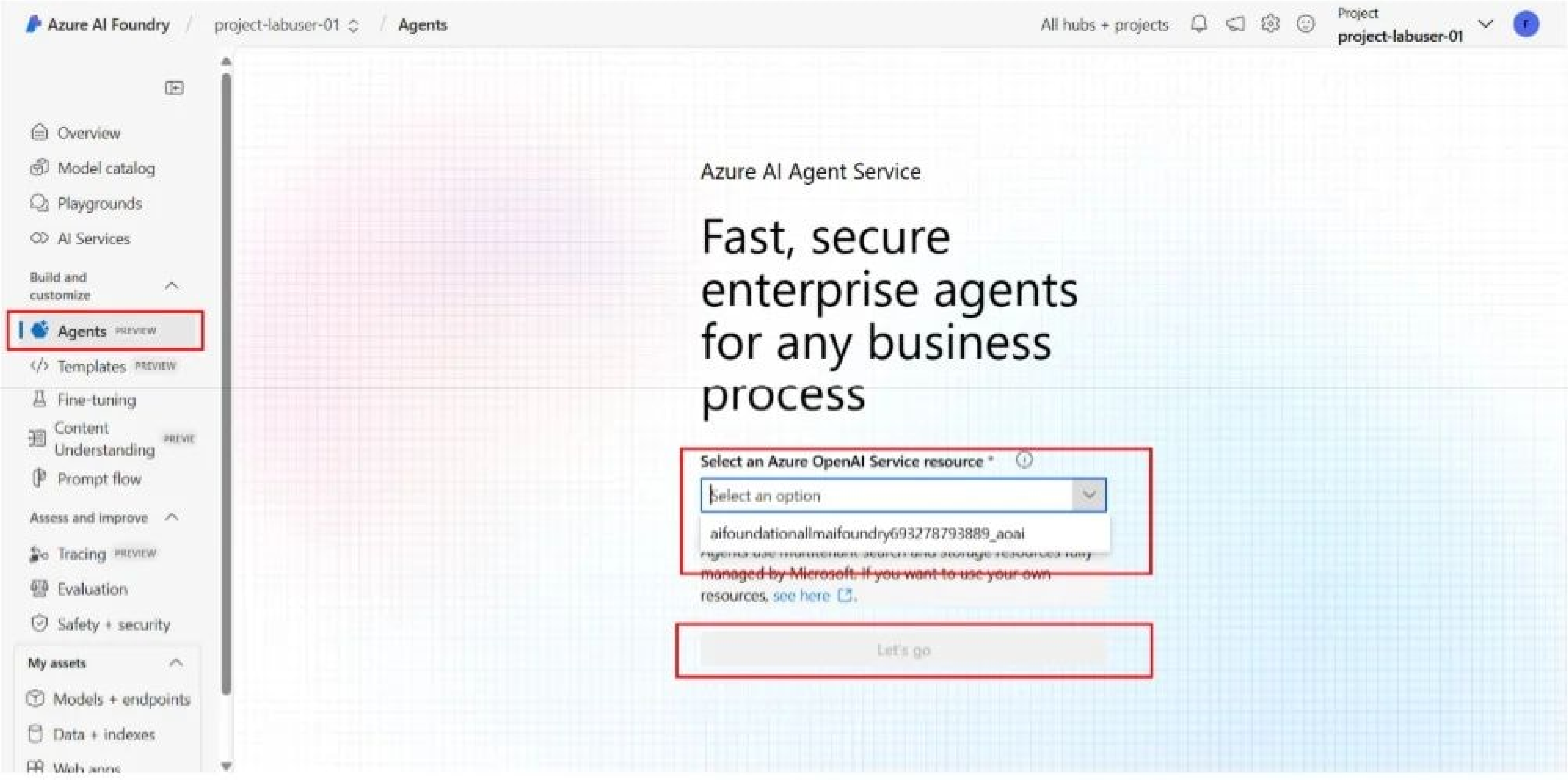
- **Objective**: Building agents atop Azure AI Foundry capabilities.
- **Duration**: 30 minutes
- **Prerequisites**: Access to Azure AI Foundry.

Exercise 1: Create a new agent in the Azure AI Foundry portal

Exercise 1: Create a new agent in the Azure AI Foundry portal

In this exercise, you will create a new agent in the Azure AI Foundry portal. You will be using your workshop account to access the portal.

1. Login to the [Azure AI Foundry Portal]() using your workshop account. Select the **Agents** section from the left navigation pane. If this is the first time you are accessing the portal, you will need to select an Azure Open AI Service resource. From the dropdown list select the `aifoundationallmaifoundry693278793889_aoai` option and the select the **Let's go** button. If you are not prompted to select an Azure Open AI Service resource, skip this step and select the **+ New agent** button.



Azure AI Agent Service

Fast, secure enterprise agents for any business process

Select an Azure OpenAI Service resource *

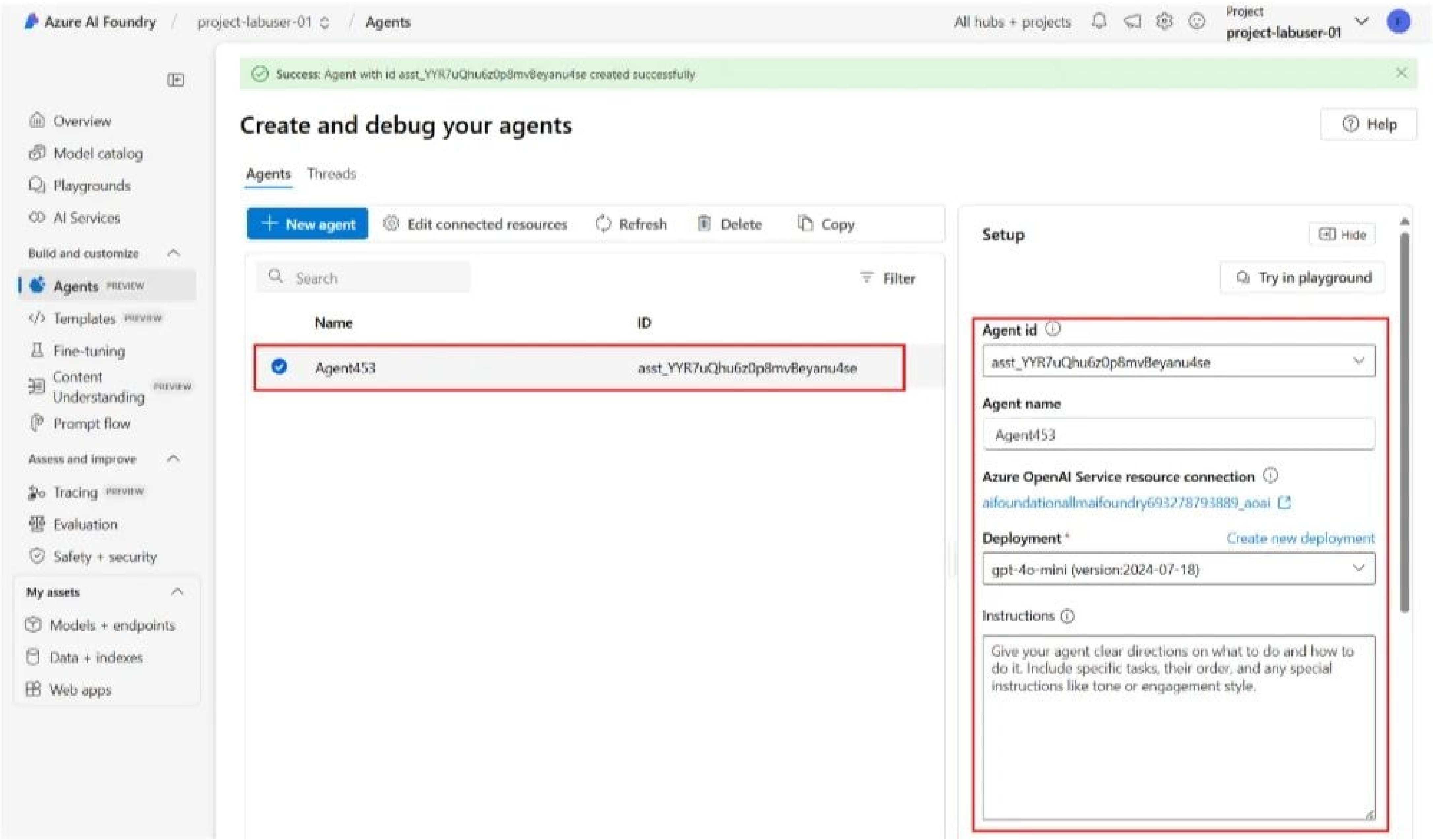
Select an option

aifoundationallmaifoundry693278793889_aoai

Agents use Microsoft search and storage resources fully managed by Microsoft. If you want to use your own resources, see here

Let's go

2. A new agent has been created for you. Note the properties of the agent. Note how the `gpt-4o-mini` model deployment has been automatically selected for the agent, as this is the only Azure OpenAI model deployment that is compatible with the agent.



Success: Agent with id asst_YYR7uQhu6z0p8mvBeyanu4se created successfully

Create and debug your agents

Name	ID
Agent453	asst_YYR7uQhu6z0p8mvBeyanu4se

Setup

Agent id: asst_YYR7uQhu6z0p8mvBeyanu4se

Agent name: Agent453

Azure OpenAI Service resource connection: aifoundationallmaifoundry693278793889_aoai

Deployment *: gpt-4o-mini (version:2024-07-18)

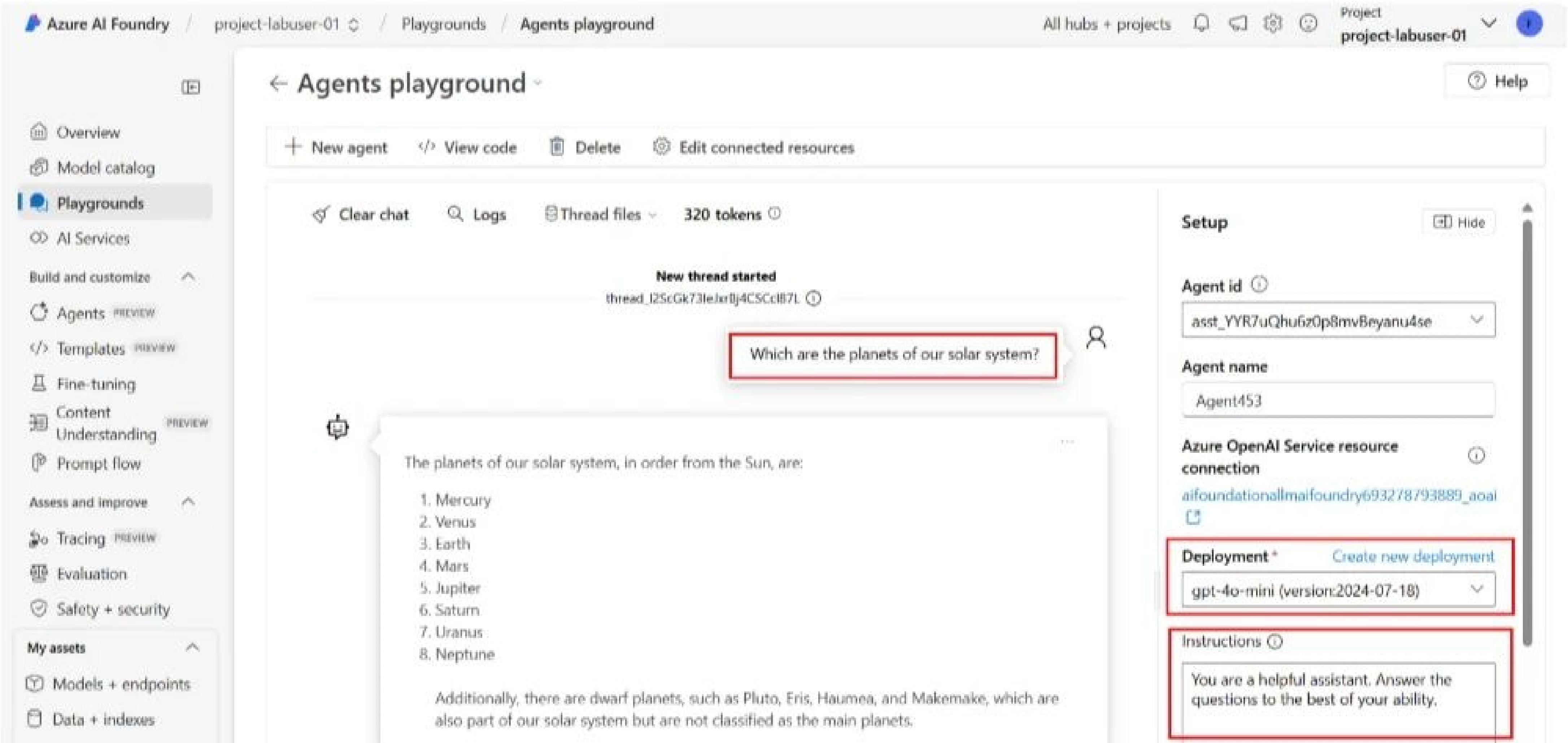
Create new deployment

Instructions: Give your agent clear directions on what to do and how to do it. Include specific tasks, their order, and any special instructions like tone or engagement style.

3. Under the **Instructions** section, enter the following instructions:

```
You are a helpful assistant. Answer the questions to the best of your ability.
```

4. Select the **Try in playground** option to test the agent in the playground. This will allow you to interact with the agent and see how it responds to different questions.



Agents playground

New thread started
thread_l25cGk73IeJxrBj4C5CcIB7L

Which are the planets of our solar system?

The planets of our solar system, in order from the Sun, are:

1. Mercury
2. Venus
3. Earth
4. Mars
5. Jupiter
6. Saturn
7. Uranus
8. Neptune

Additionally, there are dwarf planets, such as Pluto, Eris, Haumea, and Makemake, which are also part of our solar system but are not classified as the main planets.

Setup

Agent id: asst_YYR7uQhu6z0p8mvBeyanu4se

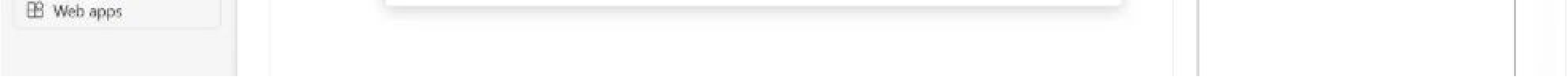
Agent name: Agent453

Azure OpenAI Service resource connection: aifoundationallmaifoundry693278793889_aoai

Deployment *: gpt-4o-mini (version:2024-07-18)

Create new deployment

Instructions: You are a helpful assistant. Answer the questions to the best of your ability.

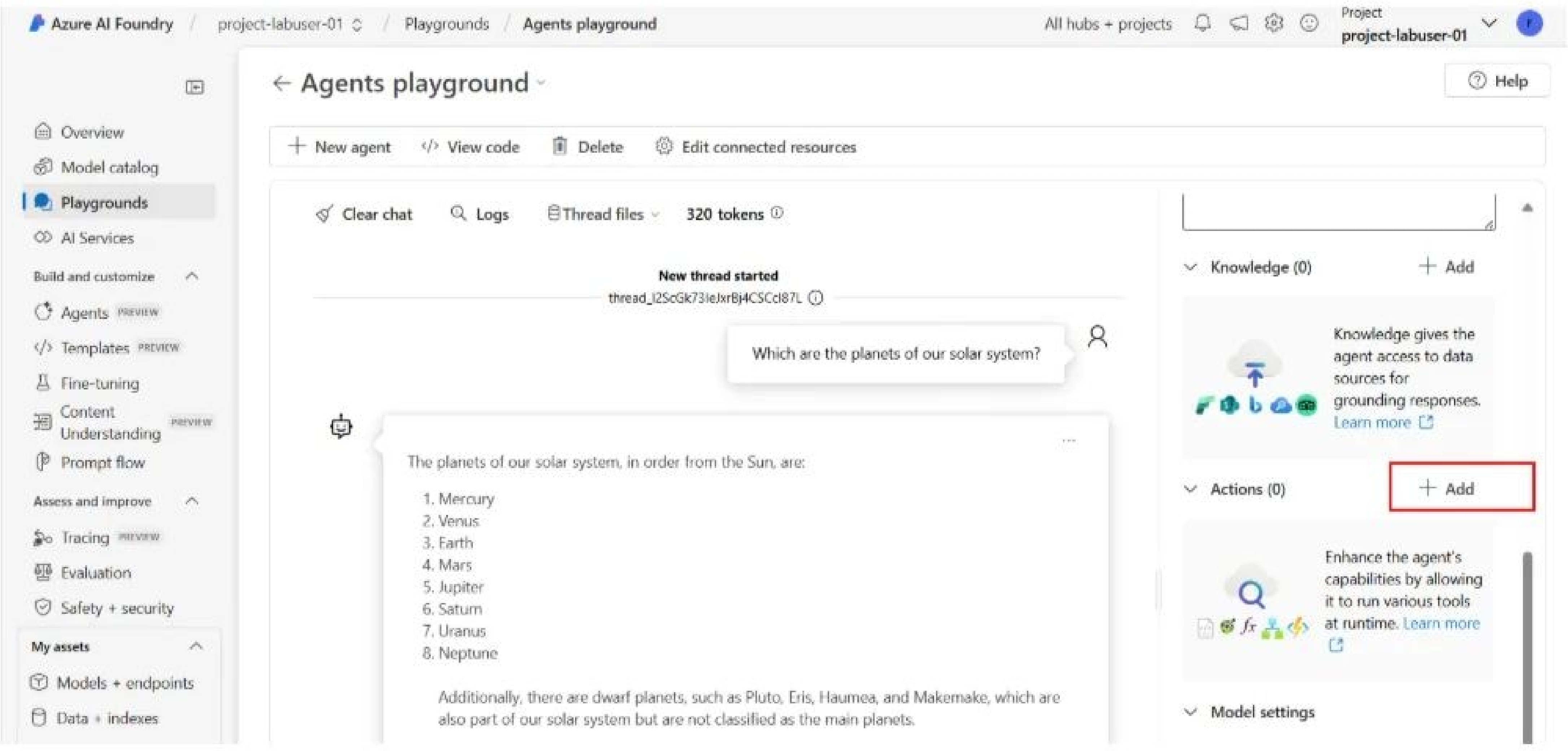


5. Change the following agent settings to see how the agent responds after each change:
   - Instructions
   - Mode settings - Temperature
   - Mode settings - Top P

Exercise 2: Add a code interpreter action to the agent

In this exercise, you will add a code interpreter action to the agent. The code interpreter action allows the agent to execute code and return the results.

1. Under the **Actions** section, select the **+ Add** option.



New thread started
thread_I2ScGk73IeJxrBj4CSCcl87L

Which are the planets of our solar system?

The planets of our solar system, in order from the Sun, are:

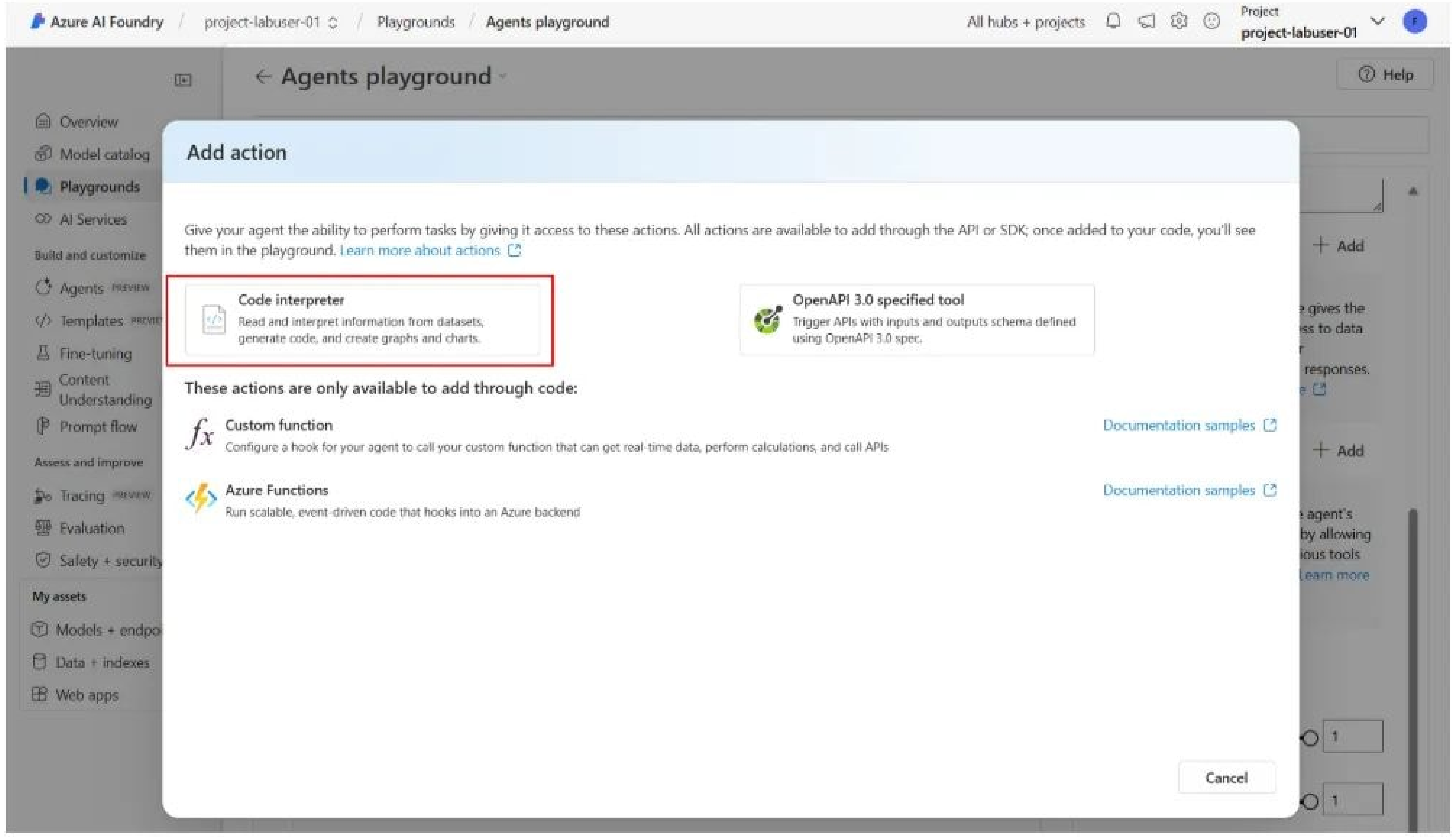
1. Mercury
2. Venus
3. Farth
4. Mars
5. Jupiter
6. Saturn
7. Uranus
8. Neptune

Additionally, there are dwarf planets, such as Pluto, Eris, Haumea, and Makemake, which are also part of our solar system but are not classified as the main planets.

Knowledge gives the agent access to data sources for grounding responses. Learn more

Enhance the agent's capabilities by allowing it to run various tools at runtime. Learn more

2. Select the **Code interpreter** action from the list of actions.



Add action

Give your agent the ability to perform tasks by giving it access to these actions. All actions are available to add through the API or SDK; once added to your code, you'll see them in the playground. Learn more about actions

Code interpreter
Read and interpret information from datasets, generate code, and create graphs and charts.

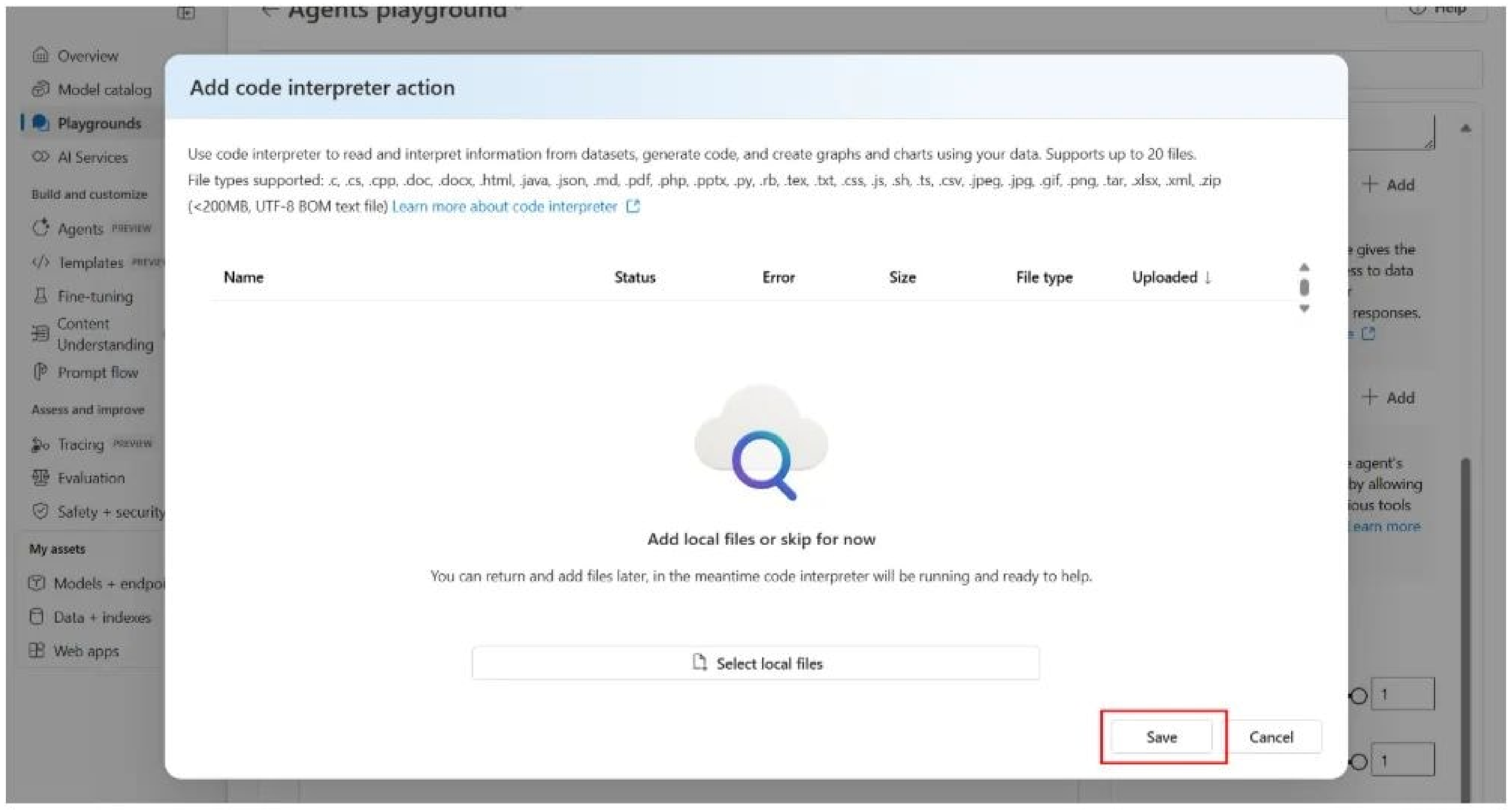
OpenAPI 3.0 specified tool
Trigger APIs with inputs and outputs schema defined using OpenAPI 3.0 spec.

These actions are only available to add through code:

Custom function — Documentation samples
Configure a hook for your agent to call your custom function that can get real-time data, perform calculations, and call APIs

Azure Functions — Documentation samples
Run scalable, event-driven code that hooks into an Azure backend

3. Select the **Save** button to add the action to the agent.



Add code interpreter action

Use code interpreter to read and interpret information from datasets, generate code, and create graphs and charts using your data. Supports up to 20 files.
File types supported: .c, .cs, .cpp, .doc, .docx, .html, .java, .json, .md, .pdf, .php, .pptx, .py, .rb, .tex, .txt, .css, .js, .sh, .ts, .csv, .jpeg, .jpg, .gif, .png, .tar, .xlsx, .xml, .zip (<200MB, UTF-8 BOM text file) Learn more about code interpreter

Name	Status	Error	Size	File type	Uploaded ↓

Add local files or skip for now

You can return and add files later, in the meantime code interpreter will be running and ready to help.

4. Navigate to the playground and ask the agent the following question:

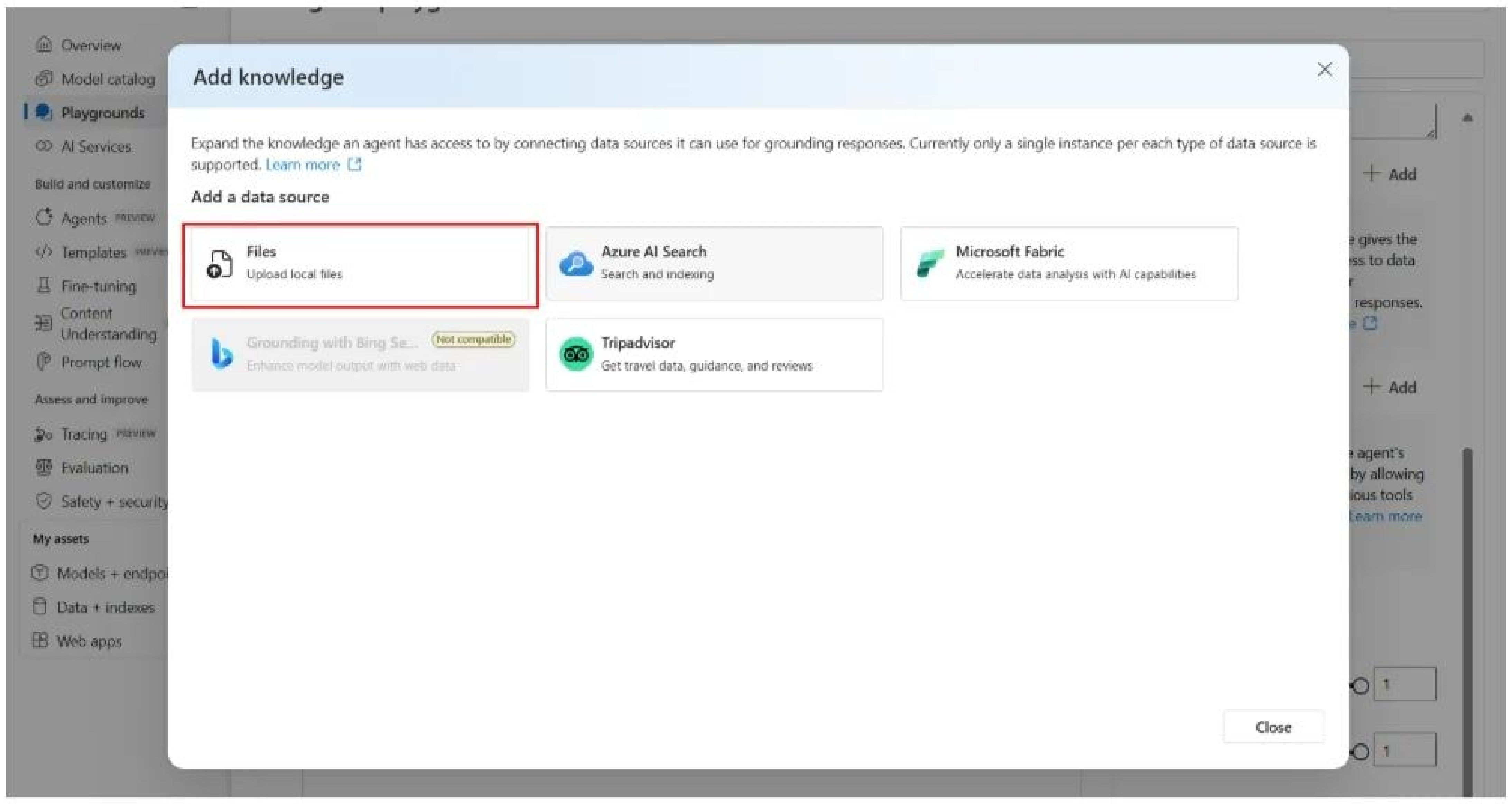
```
Create Python code to calculate the first 10 Fibonacci numbers. Calculate the numbers and make the
```

5. Note how the agent responds with the code and the download link. Download the code file and open it to validate the code.

Exercise 3: Add a knowledge action to the agent

In this exercise, you will add a knowledge action to the agent. The knowledge action allows the agent to access external knowledge sources and use them to answer questions.

1. Under the **Knowledege** section, select the **+ Add** option. Select the **Files** options from the list of data sources.

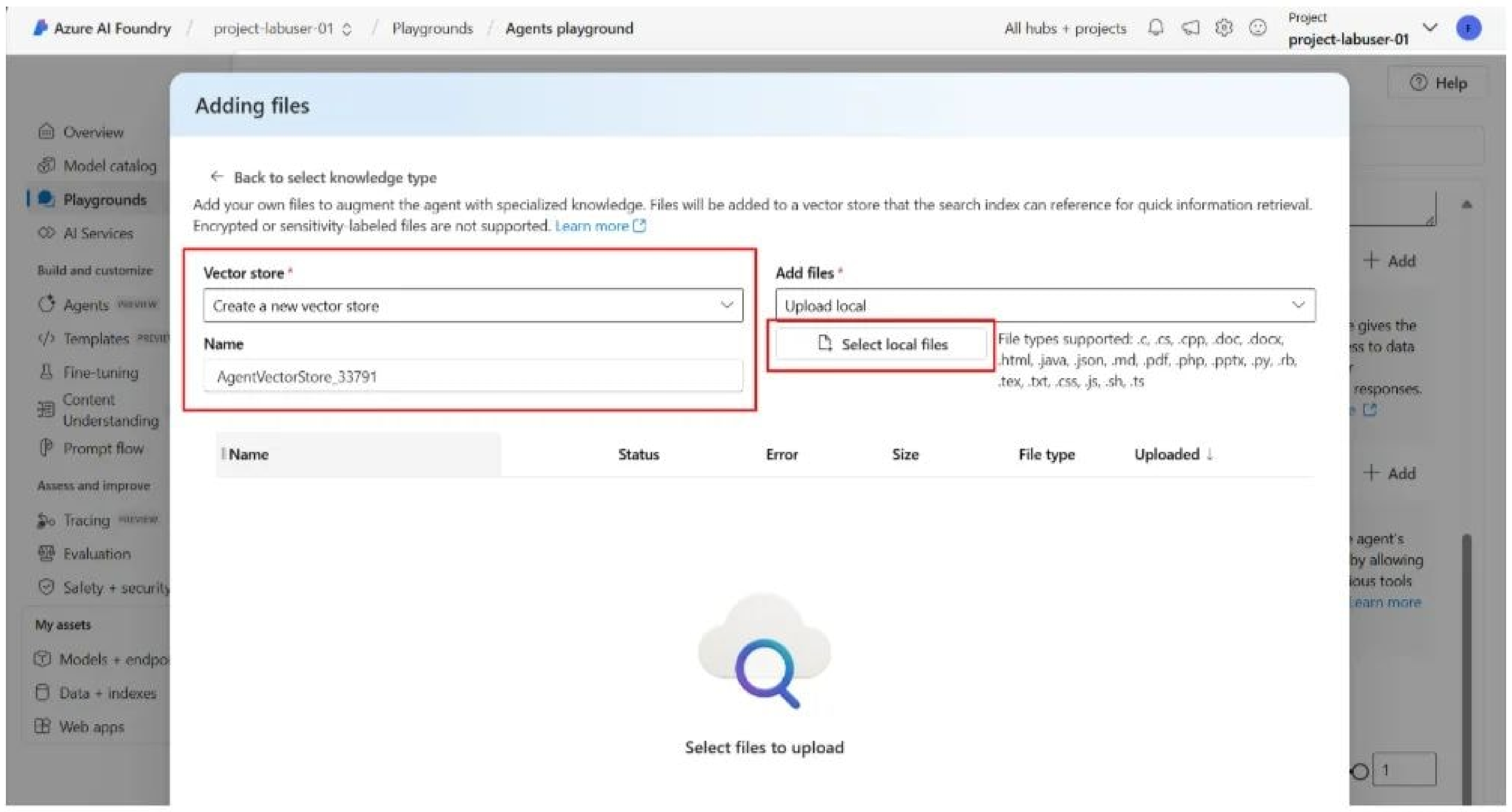


Add knowledge

Expand the knowledge an agent has access to by connecting data sources it can use for grounding responses. Currently only a single instance per each type of data source is supported. Learn more

Add a data source

2. Do not change the default options related to the vector store. Select the **Select local files** to upload files.



Adding files

← Back to select knowledge type

Add your own files to augment the agent with specialized knowledge. Files will be added to a vector store that the search index can reference for quick information retrieval. Encrypted or sensitivity-labeled files are not supported. Learn more

Vector store *: Create a new vector store

Name: AgentVectorStore_33791

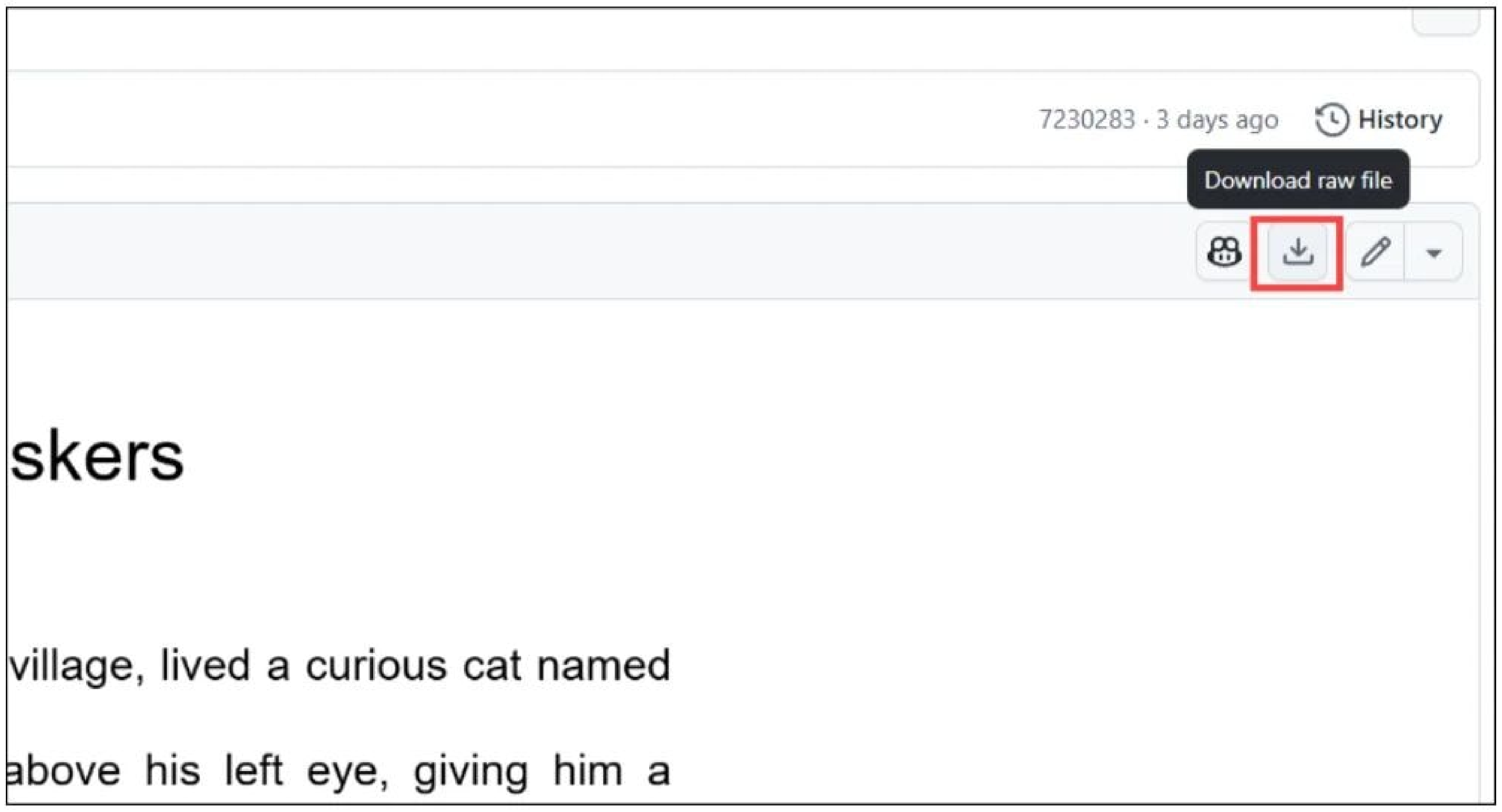
Add files *: Upload local

File types supported: .c, .cs, .cpp, .doc, .docx, .html, .java, .json, .md, .pdf, .php, .pptx, .py, .rb, .tex, .txt, .css, .js, .sh, .ts

Name	Status	Error	Size	File type	Uploaded ↓

Select files to upload

3. Download the The Adventure of Whiskers file. In GitHub select the **Download raw file** option to download the file.

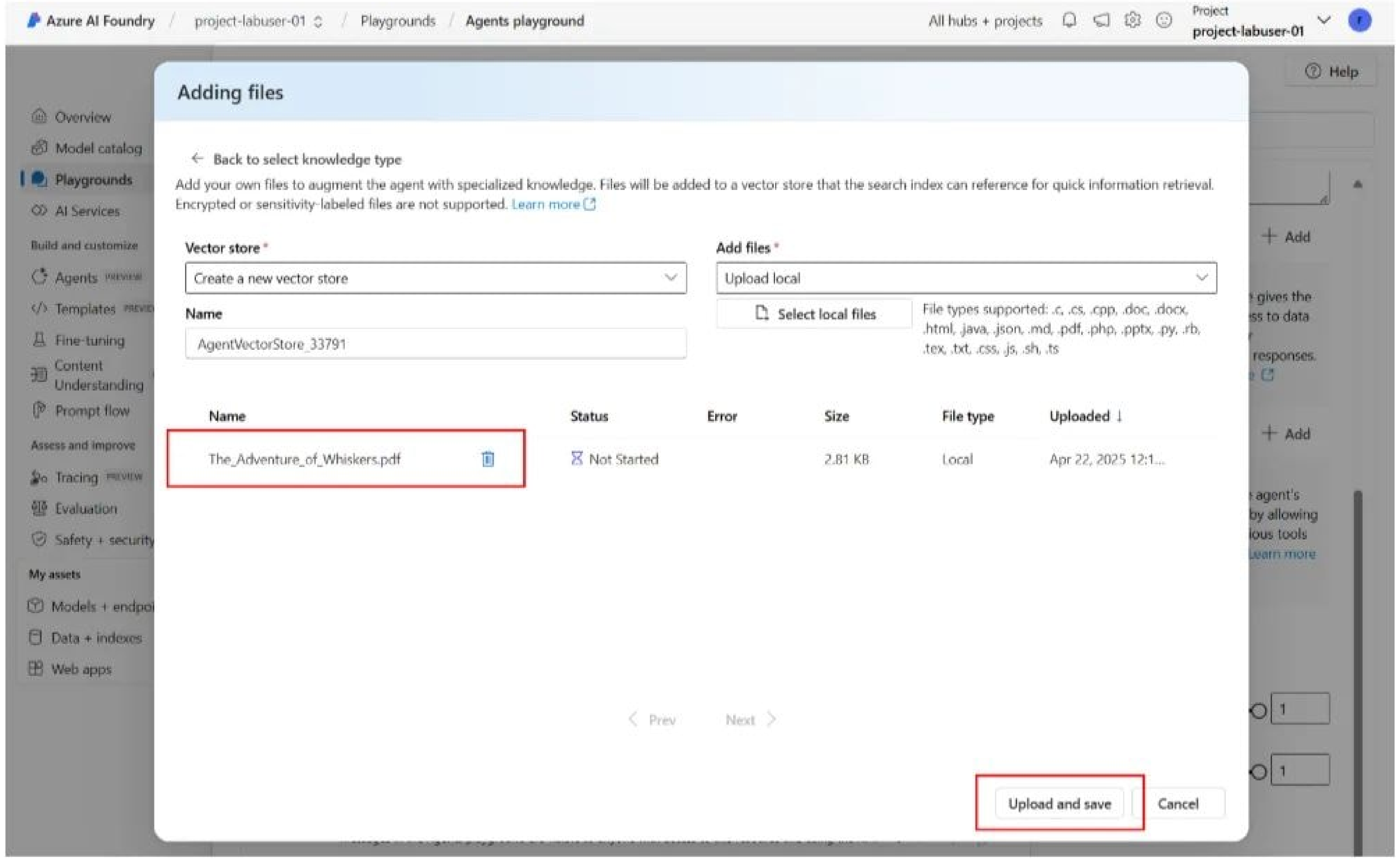


skers

village, lived a curious cat named

above his left eye, giving him a

4. Upload the The Adventure of Whiskers file and select **Upload and save** to save the action.



Adding files

← Back to select knowledge type

Add your own files to augment the agent with specialized knowledge. Files will be added to a vector store that the search index can reference for quick information retrieval. Encrypted or sensitivity-labeled files are not supported. Learn more

Vector store *: Create a new vector store

Name: AgentVectorStore_33791

Add files *: Upload local

Select local files

File types supported: .c, .cs, .cpp, .doc, .docx, .html, .java, .json, .md, .pdf, .php, .pptx, .py, .rb, .tex, .txt, .css, .js, .sh, .ts

Name	Status	Error	Size	File type	Uploaded ↓
The_Adventure_of_Whiskers.pdf	⌛ Not Started		2.81 KB	Local	Apr 22, 2025 12:1...

< Prev Next >

Upload and save Cancel

5. Navigate to the playground and ask the agent the following question:

```
Tell me about the adventure of Whiskers.
```

6. Note how the agent responds with the content.

Exercise 4: Use a combination of knwoledge and code interpreter actions

In this exercise, you will use a combination of knowledge and code interpreter actions to create a PDF file that contains a summarized version of the adventure of Whiskers.

1. Ask the agent the following question:

```
Create a PDF file that contains a summarized version of the adventure of Whiskers. Make the file av
```

You have now completed the lab. You have learned how to create an agent in Azure AI Foundry and how to add actions to the agent. You have also explored how to use the agent in the playground and how to interact with it using different actions.

Demo 1: Knowledge Graphs - the future of RAG

- **Objective**: Understand the limitations of basic RAG, understand the potential of Knowledge Graphs and how can they be applied to improve RAG performance.
- **Duration**: 30 minutes
- **Prerequisites**: None

In this interactive demo we will explore how to use knowledge graphs to enhance the performance of RAG (Retrieval-Augmented Generation) agents. We will demonstrate the FoundationalLLM implementation of knowledge graphs and how they can be used to improve the performance of RAG agents.

Demo 2: Going beyond the basic tools

- **Objective**: Understand which are the benefits and the challenges of implementing a practical, enterprise-grade AI solution on top of Azure AI Foundry.
- **Duration**: 30 minutes
- **Prerequisites**: None

In this interactive demo we will explore how to use the FoundationalLLM platform to build agents that go beyond the basic tools. We will demonstrate how to use the FoundationalLLM platform to build agents that can perform complex tasks and how to use the FoundationalLLM platform to enrich the performance of agents with capabilities such as prompt rewriting, semantic caching, analytics, and more.