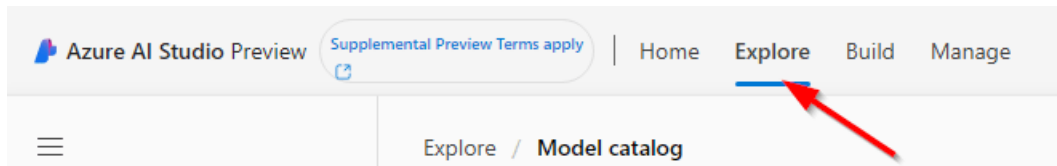


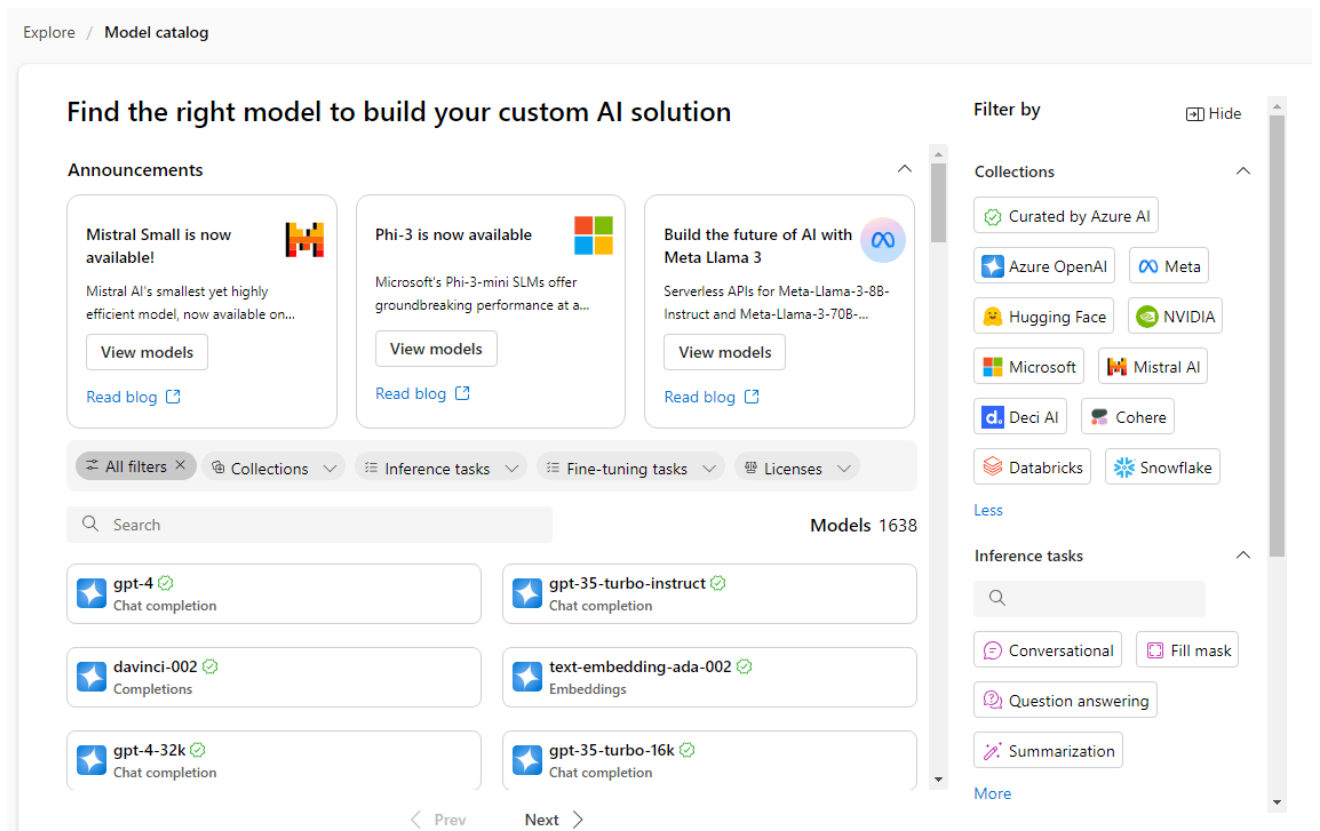
Exercise 3- AI Studio Explore

Model catalog

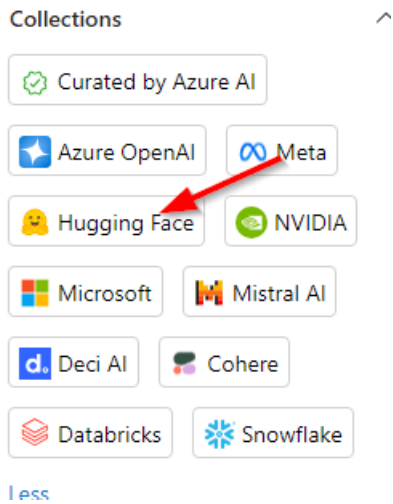
1. Click on the Explore tab in the navigation



2. The resulting model catalog. You will notice this is vastly different than the 4-5 models available in the current AI Studio.

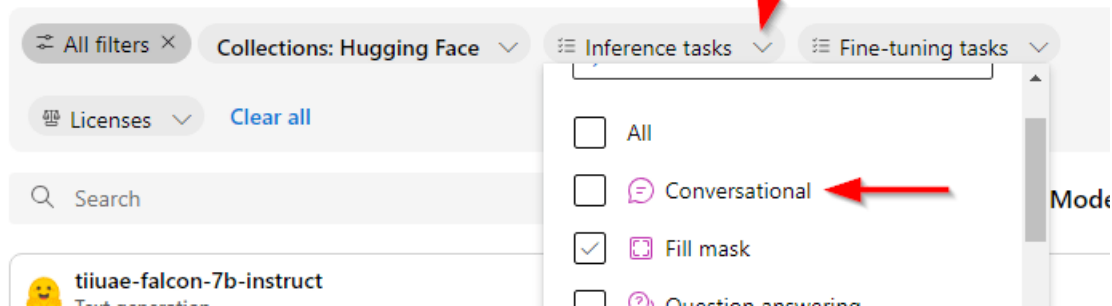


3. Click on the Hugging Face icon

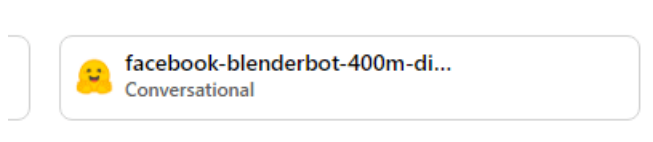


- Hugging Face has over 600,000 models available <https://huggingface.co/>.
- After filtering to Hugging Face models click on the **Inference tasks** dropdown and choose **Conversational**

Find the right model to build your custom AI solution



- Look for the **facebook-blenderbot-400m...** and select it

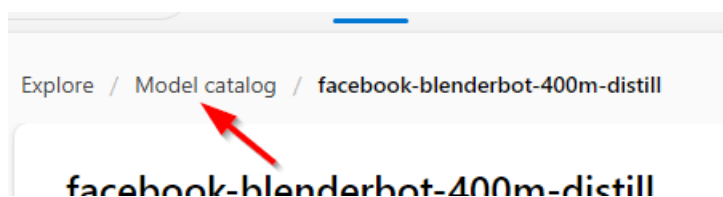


- It should throw an error at the top

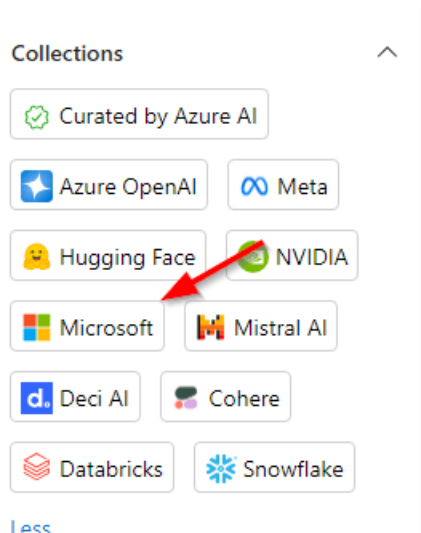
facebook-blenderbot-400m-distill

Failed to parse artifact data:

- Click the breadcrumb that says **Model catalog**



9. This time click on the **Microsoft** collection



















10. Select the **microsoft-phi-2** model and review the details about the model

Find the right model to build your custom AI solution

Filters: All filters x Collections: Microsoft v Inference tasks v Fine-tuning tasks v Licenses v

[Clear all](#)

Search Models 8

 Phi-3-mini-4k-instruct  Chat completion	 Phi-3-mini-128k-instruct  Chat completion
 microsoft-phi-2  Text generation	 microsoft-Orca-2-7b  Text generation
 microsoft-Orca-2-13b  Text generation	 microsoft-swinv2-base-patch4-...  Image classification
 microsoft-beit-base-patch16-2...  Image classification	 microsoft-phi-1-5  Text generation

1. On the **Details** page review the **Description** section. the scroll down to the **Sample** section

1. At the top of the page click on the **Deploy** button

microsoft-phi-2

Task: Text generation Fine-tuning task: text-generation Languages: en License: mit

Refresh Fine-tune Deploy View license

- In the **Deploy model** popup make sure the project previously created is selected. And then click the **Shared quota** notification

Deploy model

For the selected model, the scoring script and environment are auto generated for you.
[Learn More](#)

Project [Create a new project](#)

democopilot-c-9304

You have no dedicated quota. A temporary 168-hour endpoint will be created for you. Alternatively, you can [request for quota for persistent endpoints](#). [Learn more about shared quota](#).

☐ I want to use shared quota and I acknowledge that this endpoint will be deleted in 168 hours

Deploy Cancel

- The popup will update to display the **Virtual machine** options

☒ I want to use shared quota and I acknowledge that this endpoint will be deleted in 168 hours

Virtual machine * *i*

Standard_NC12s_v3 12 Cores, 224 GB (RAM), 672 GB (Disk), \$6.12/hr

Instance count * *i*

1

- In the bottom section you can adjust the **Endpoint name** and the **Deployment name**

Endpoint

☒ New

Endpoint name * ⓘ



democopilot-c-ai-nfmxn

ⓘ An endpoint URL will be generated after creating an endpoint.

<https://democopilot-c-ai-nfmxn.eastus.inference.ml.azure.com/score>

[Learn how to consume](#)

Deployment name * ⓘ



msft-phi-2-17

Inferencing data collection ⓘ

☐ Disabled

Deploy

Cancel

5. Click **Cancel** in the popup before deploying