Microsoft

AI-050: Develop Generative AI Solutions with Azure OpenAI Service

Module 6

Use your own data with Azure OpenAI Service

Agenda

- How using your own data works
- Use the REST API
- Use language specific SDKs

Learning Objectives

After completing this module, you will be able to:

1. Describe the capabilities of Azure OpenAI on your data
2. Configure Azure OpenAI to use your own data
3. Use Azure OpenAI API to generate responses based on your own data

How Azure OpenAI can use your data

Set up your data source

- Use an existing data source, such as an Azure search resource
- Use the Azure OpenAI studio to create that data source, if you don't already have one
- When creating the data source, you can use data already in your account such as blob storage

Configure the studio or your app to connect to that data source

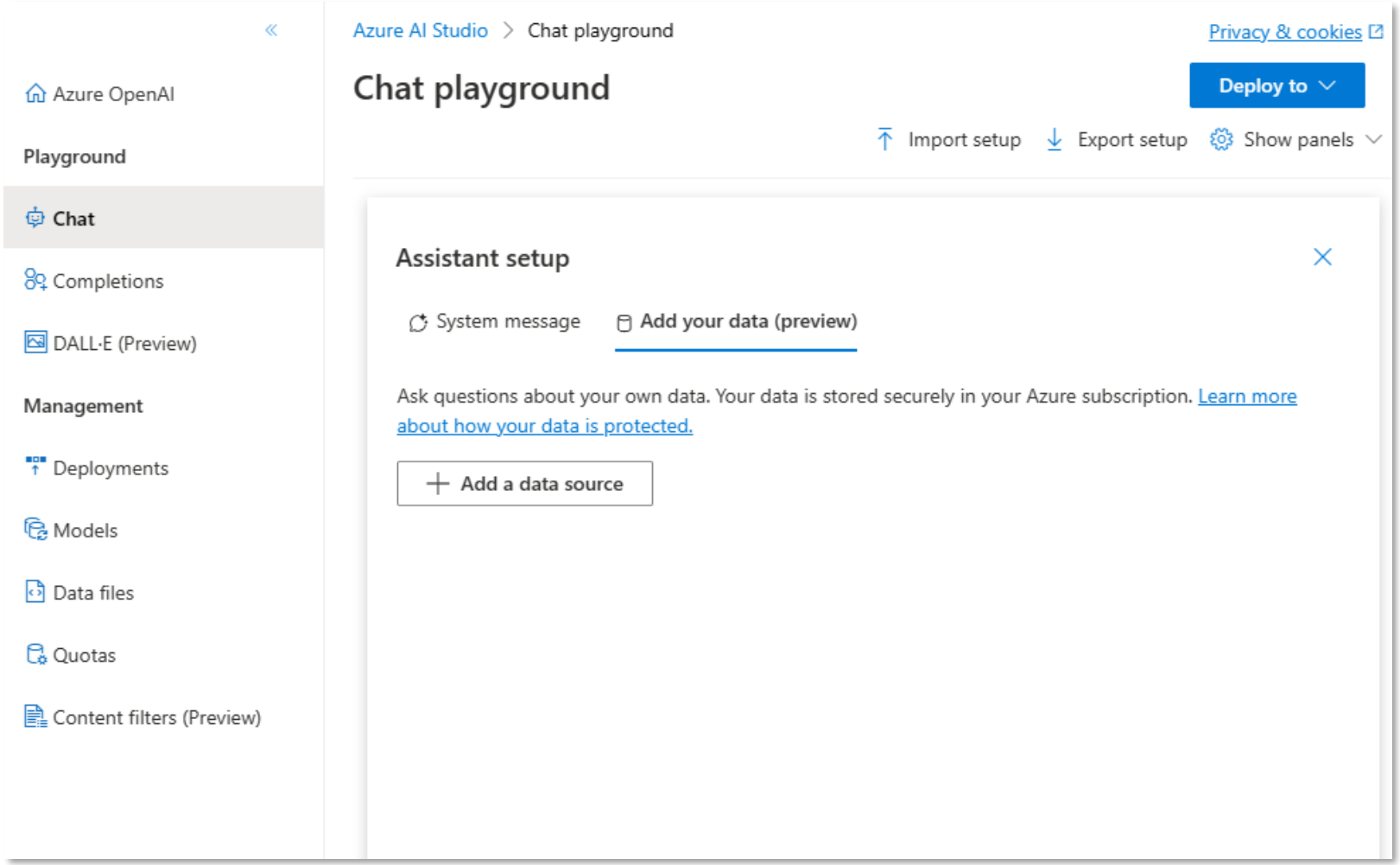
- In the studio, set up the connection by pointing it to the data source
- In your app, specify the data source in the prompt parameters
- Both configurations allow the search resource to augment the prompt

Use the Azure OpenAI model, which now uses your data for grounding

- Chat with the AI models like normal
- If the data source has relevant information about the prompt, it will use that data
- You can specify if the AI model is limited to just your data source

Connect to your data source

- Add your data source in the Chat playground, under Assistant setup
- Use an existing data source, or use that wizard to create a new one
- Once connected, a new chat session will start. Chat like normal, and see how the AI model references that data



Using the Azure OpenAI REST API

Using your own data

https://*endpoint*.openai.azure.com/openai/deployments/*deployment*/**chat/completions**?api-version=*version*

- With each call, you need to specify the data source values, along with the messages array and any other parameters
- Authentication in the data source definition is for your search resource, not your Azure OpenAI resource

```
{
    "data_sources": [
        {
            "type": "azure_search",
            "parameters": {
                "endpoint": "<your_search_endpoint>",
                “index_name": "<your_search_index>“,
                "authentication": {
                    "type": "system_assigned_managed_identity"
                }
            }
        }
    ],
    "messages":[
                ...
    ]
}
```

Using the Azure OpenAI SDKs

Language specific SDKs are available for use in your applications, in both C# and Python.

Code structure follows a similar pattern for both languages.

Current supported data sources are:

- Azure AI Search
- Azure Cosmos DB for MangoDB vCore
- Plus others in preview, soon to be released GA

Pseudo code structure:

```
<include library>

<create client>

<define chat messages and options>

<define data source object to include with
request>

<send request>

<extract response content>
```

Exercise: Implement Retrieval Augmented Generation (RAG) with Azure OpenAI Service

Use the hosted lab environment if provided, or view the lab instructions at the link below:

https://aka.ms/mslearn-openai-own-data

Knowledge check

1 **What does Azure OpenAI on your data enable developers to do?**

- ❑ Create their own AI chat models
- ❑ Access Azure OpenAI without an approved subscription
- ☑ Use supported AI chat models that can reference specific sources of data

2 **What is the recommended way to add data when using Azure OpenAI on your data?**

- ❑ Using any data source option available for Azure OpenAI on your data
- ☑ Using Azure AI Studio to create the search resource and index
- ❑ Connecting to files in a storage account without using Azure AI Studio

3 **What are some recommended prompt engineering techniques when using Azure OpenAI on your own data?**

- ☑ Break down the task and use chain of thought prompting
- ❑ Include as much conversation history as possible in your call
- ❑ Use a single long prompt to provide all necessary information

Learning Recap

In this module, we:

Described the capabilities of Azure OpenAI on your data

Configured Azure OpenAI to use your own data

Used Azure OpenAI API to generate responses based on your own data

Resources

Implement Retrieval Augmented Generation (RAG) with Azure OpenAI Service

https://aka.ms/mslearn-openai-own-data-module

Microsoft

Thank you.