Microsoft

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

Use Azure OpenAI APIs in your applications

Agenda

- Use the APIs
- Integrate Azure OpenAI into your app

Learning Objectives

After completing this module, you will be able to:

1. Integrate Azure OpenAI into your application
2. Differentiate between different endpoints available to your application
3. Generate completions to prompts using the REST API and language specific SDKs

Integrating Azure OpenAI into your app

Applications submit prompts to deployed models. Responses are completions.

- **Completion** - model takes an input prompt, and generates one or more predicted completions
- **Embeddings** - model takes input and returns a vector representation of that input
- **ChatCompletion** - model takes input in the form of a chat conversation (where roles are specified with the message they send), and the next chat completion is generated

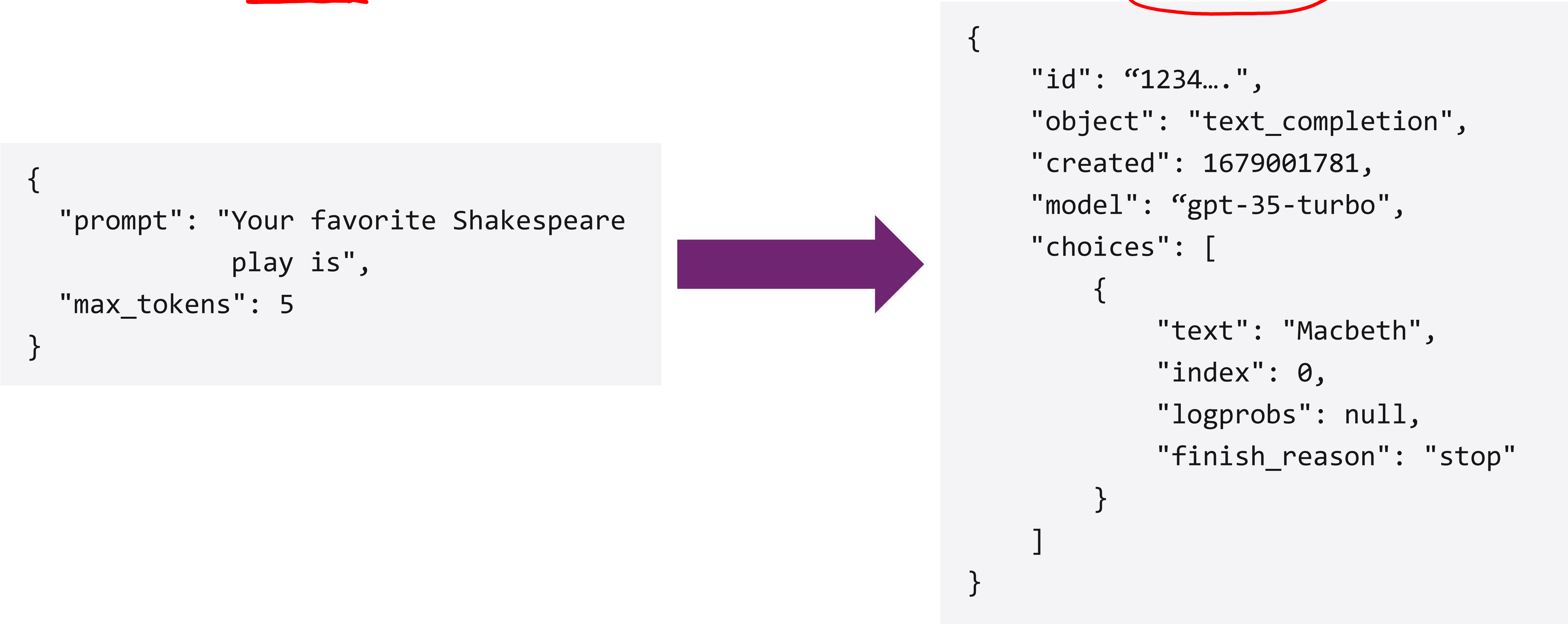
ChatCompletion will be the endpoint we focus on for this course

Use **Completion** and **Embeddings** with GPT-3 based models
Use **ChatCompletion** with GPT-35-Turbo and later models

Using the Azure OpenAI REST API

Completion Endpoint

https://*endpoint*.openai.azure.com/openai/deployments/*deployment*/**completions**

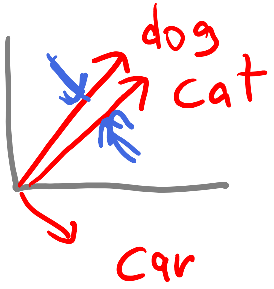


```
{
  "prompt": "Your favorite Shakespeare
             play is",
  "max_tokens": 5
}
```

```
{
    "id": “1234….",
    "object": "text_completion",
    "created": 1679001781,
    "model": “gpt-35-turbo",
    "choices": [
        {
            "text": "Macbeth",
            "index": 0,
            "logprobs": null,
            "finish_reason": "stop"
        }
    ]
}
```

Using the Azure OpenAI REST API

Embedding Endpoint



https://*endpoint*.openai.azure.com/openai/deployments/*deployment*/**embeddings**

```
{
  "input": "The food was delicious and
           the waiter was very
           friendly..."
}
```

```
{
  "object": "list",
  "data": [
    {
      "object": "embedding",
      "embedding": [
        0.0172990688066482523,
        ....
        0.0134544348834753042,
      ],
      "index": 0
    }
  ],
  "model": "text-embedding-ada:002"
}
```

?

Using the Azure OpenAI REST API

ChatCompletion Endpoint

https://*endpoint*.openai.azure.com/openai/deployments/*deployment*/**chat/completions**

```
{
  "messages":[
    {"role": "system",
     "content": "You are an assistant
      that teaches people about AI."},
    {"role": "user",
     "content": "Does Azure OpenAI
          support multiple languages?"},
    {"role": "assistant",
     "content": "Yes, Azure OpenAI
          supports several languages."},
    {"role": "user",
     "content": "Do other Cognitive
         Services support translation?"}
  ]
}
```

```
{
    "id": “unique_id", "object": "chat.completion",
    "created": 1679001781, "model": "gpt-35-turbo",
    "usage": { "prompt_tokens": 95,
               "completion_tokens": 84, "total_tokens": 179},
    "choices": [
          {"message":
                {  "role": "assistant",
                    "content": "Yes, other Azure Cognitive
                        Services also support translation…"},
            "finish_reason": "stop",
            "index": 0}
    ]
}
```

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.

Parameters such as *Max Response* and *Temperature* are defined in the chat options.

Both synchronous and asynchronous API versions are available.

Pseudo code structure:

```
<include library>

<create client>

<define chat messages and options>

<send request>

<extract response content>
```

Exercise: Integrate Azure OpenAI into your app

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

https://aka.ms/mslearn-azure-openai-api

Knowledge check

1 **Which REST endpoint should you use to interact with a GPT-4 model?**

□ Completion

□ Embeddings

☑ ChatCompletion

2 **When using the .NET SDK, which method should you use to call the ChatCompletion API?**

□ ChatMessage()

☑ GetChatCompletions()

□ GetCompletions()

3 **When using the Python SDK, which method should you use to call the ChatCompletion API?**

□ openai.ChatCompletion.get()

☑ openai.ChatCompletion.create()

□ openai.chat.complete()

Learning Recap

In this module, we:

Integrated Azure OpenAI into your application

Differentiated between different endpoints available to your application

Generated completions from the ChatCompletion endpoint

Resources

Use the Azure OpenAI service API

https://aka.ms/mslearn-build-openai

Microsoft

Thank you.