✓ 100 XP

What is Azure OpenAI?

3 minutes

Azure OpenAI Service is Microsoft's cloud solution for deploying, customizing, and hosting large language models. It brings together the best of OpenAI's cutting edge models and APIs with the security and scalability of the Azure cloud platform. Microsoft's partnership with OpenAI enables Azure OpenAI users to access the latest language model innovations.

Azure OpenAI supports many models that can serve different needs. These models include:

- **GPT-4 models** are the latest generation of *generative pretrained* (GPT) models that can generate natural language and code completions based on natural language prompts.
- **GPT 3.5 models** can generate natural language and code completions based on natural language prompts. In particular, **GPT-35-turbo** models are optimized for chat-based interactions and work well in most generative AI scenarios.
- **Embeddings models** convert text into numeric vectors, and are useful in language analytics scenarios such as comparing text sources for similarities.
- **DALL-E models** are used to generate images based on natural language prompts. Currently, DALL-E models are in preview. DALL-E models aren't listed in the Azure OpenAI Studio interface and don't need to be explicitly deployed.

Models differ by speed, cost, and how well they complete specific tasks. You can learn more about the differences and latest models offered in the Azure OpenAI Service documentation.

In many cases, models can be used as-is. For example, in Azure OpenAI Service, you can deploy a GPT-4 model and immediately start using it from an application. However, you can also use an existing model as a foundational model - a starting point for further training with your own data. This approach is called fine-tuning, and it enables you to train a custom model that builds on the pre-trained model, but which is tuned to data that is relevant for your particular scenario. For example, a legal firm might fine-tune a model with the text from existing contracts and other proprietary legal documents to train a model that is optimized for generating contractual content.

Azure AI Studio

Developers can work with these models in Azure AI Studio, a web-based environment where AI professionals can deploy, test, and manage LLMs that support generative AI app development on Azure.
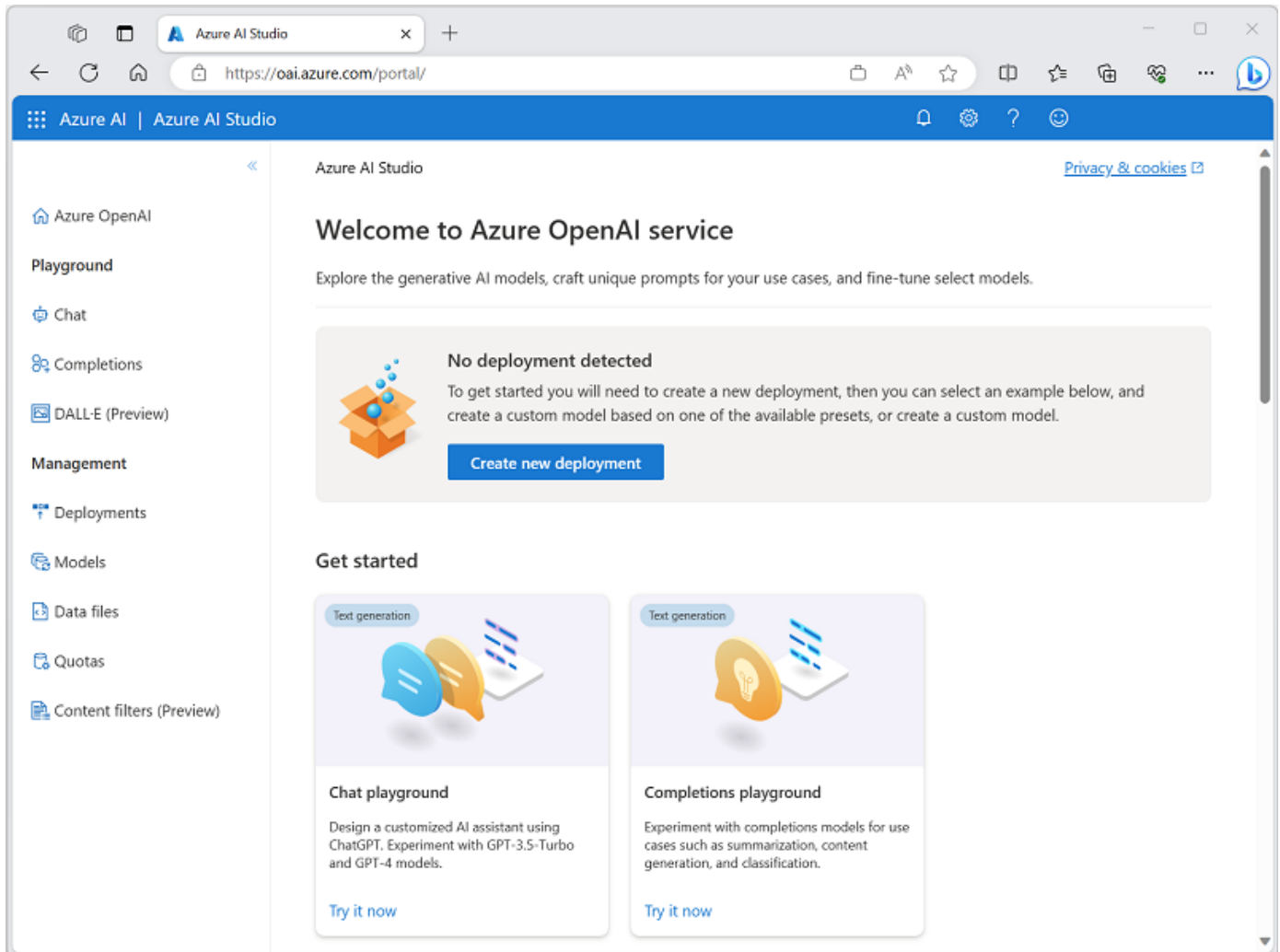
Within Azure AI Studio, you can deploy large language models, provide few-shot examples, and test them in Azure OpenAI Studio's Chat playground.

Azure OpenAI

Playground

🖧 Chat

🎛 Completions

Management

📌 Deployments

🗂 Models

📄 File Management

Chat playground (Preview)

⤒ Import setup   ⤓ Export setup   ⚙ Show panels ⌄

Assistant setup   ✕

Load example setup

Load example setup   ⌄

ⓘ Please click the 'Save changes' button below to apply your changes.

💾 Save changes

System message ⓘ

You are an AI assistant that helps people find information.

⌄ Few-shot examples ⓘ

User: 🗑

Assistant:

＋ Add few-shot example

Chat session

🖌 Clear chat   🔲 View code   ⬤ Show raw JSON

🤖   **Start chatting**

Test your assistant by sending queries below. Then adjust your assistant setup to improve the assistant's responses.

User message

Type user query here. (Shift + Enter for new line)   ▷

Parameters   ✕

Deployments

⌄

Max response ⓘ

800

Temperature ⓘ   0.7

Top P ⓘ   0.95

Stop sequence ⓘ

Stop sequences

Frequency penalty ⓘ

0

Presence penalty ⓘ

0

Next unit: What are copilots?

[Continue ›]