

Azure OpenAI Services

December 2025



What is OpenAI?

OpenAI is a leading research laboratory that develops safe and beneficial AI for personal and professional use. They've created some of the most advanced AI models, including GPT-5, Embeddings, DALL-E, and Whisper.

zappliance

[Introducing-OpenAI](#)



What is Azure OpenAI Service?

Azure OpenAI Service is a Microsoft Azure cloud service that provides enterprise access to OpenAI's advanced language and generative models, such as GPT models, through Azure infrastructure.

[Microsoft and OpenAI partnership unveils new AI opportunities](#)

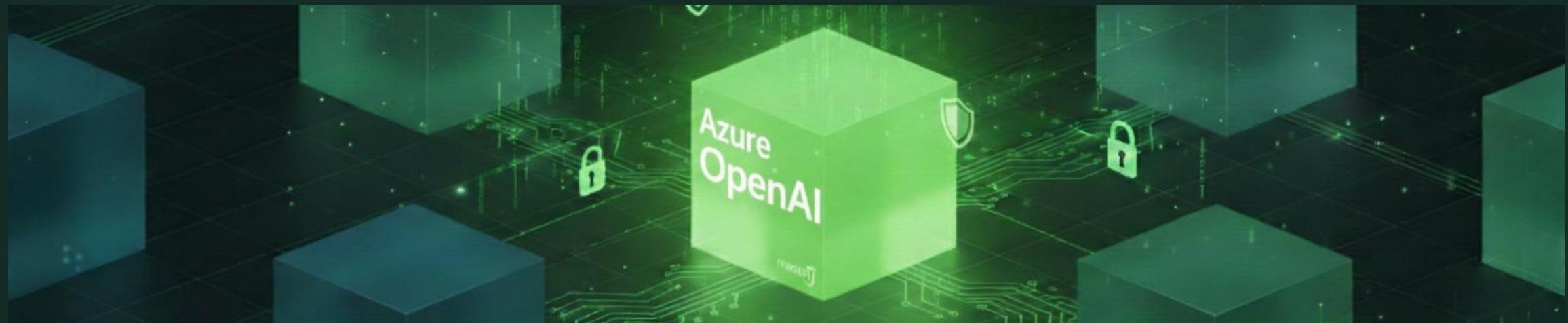
[What is Azure OpenAI Service?](#)



With Azure OpenAI, customers get the security capabilities of Microsoft Azure while running the same models as OpenAI.

Azure OpenAI offers private networking, regional availability, and responsible AI content filtering.

The Microsoft Responsible
AI Standard



Comparing Azure OpenAI and OpenAI

zappliance

Azure OpenAI offers superior security capabilities in comparison to OpenAI. Additionally, Azure OpenAI provides support for Azure Virtual Network (VNet) and private endpoints, which are not available in OpenAI. Another advantage of using Azure OpenAI is that it comes with Azure Cognitive Services, Service Level Agreement for Azure (this “SLA”), while OpenAI currently only offers a status page:

[SLA for Cognitive Services](#)

[Azure OpenAI vs OpenAI](#)

[Explore the benefits of Azure OpenAI Service with Microsoft Learn](#)



Data, privacy, and security for Azure OpenAI Service

zappliance

Azure OpenAI stores and processes data to provide the service and to monitor for uses that violate the applicable product terms.

Please also see the [Microsoft Products and Services Data Protection Addendum](#), which governs data processing by the Azure OpenAI Service except as otherwise provided in the applicable [Product Terms](#).





Important

The prompts (inputs) and completions (outputs):

- are NOT available to other customers.
- are NOT available to OpenAI.
- are NOT used to improve OpenAI models.
- are NOT used to improve any Microsoft or 3rd party products or services.
- are NOT used for automatically improving Azure OpenAI models for your use in your resource
(The models are stateless, unless you explicitly fine-tune models with your training data).

The Azure OpenAI Service is fully controlled by Microsoft; Microsoft hosts the OpenAI models in Microsoft's Azure environment and the Service does NOT interact with any services operated by OpenAI (e.g. ChatGPT, or the OpenAI API).

[More details about Data, privacy, and security...](#)

Create and deploy an Azure OpenAI Service resource

Please note that for the sake of simplicity, this document provides step-by-step instructions to create an Azure OpenAI Service resource and deploy a model via Azure Foundry. You can choose either resource creation via Azure Command-Line-Interface (CLI) or PowerShell. Please see [here](#) for more details.

zappliance

Create customized, AI-powered search with an Azure free account

Get started with 12 months of free services including Azure AI Services

Start free

Pay as you go

1

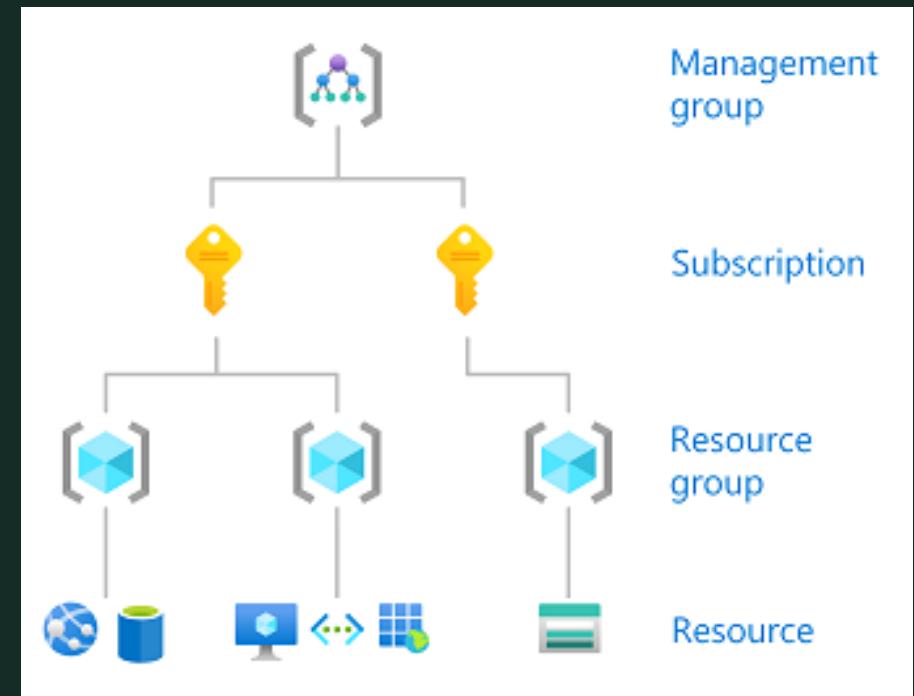
An Azure subscription

- [Create one for free](#)
- Start free. Get \$200 credit to use in 30 days. While you have your credit, get free amounts of popular services and 55+ other services.
 - Consider upgrading to the paid tier from the outset if you intend to utilize this service for production purposes.
- After your credit, move to pay as you go to keep getting popular services and 55+ other services. Only pay if you use more than the free monthly amounts.
- After 12 months, you'll continue getting 55+ services free always—and still only pay for what you use beyond the free monthly amounts.

2

Access permissions to create Azure OpenAI resources and to deploy models.

Please see [here](#) for more details.

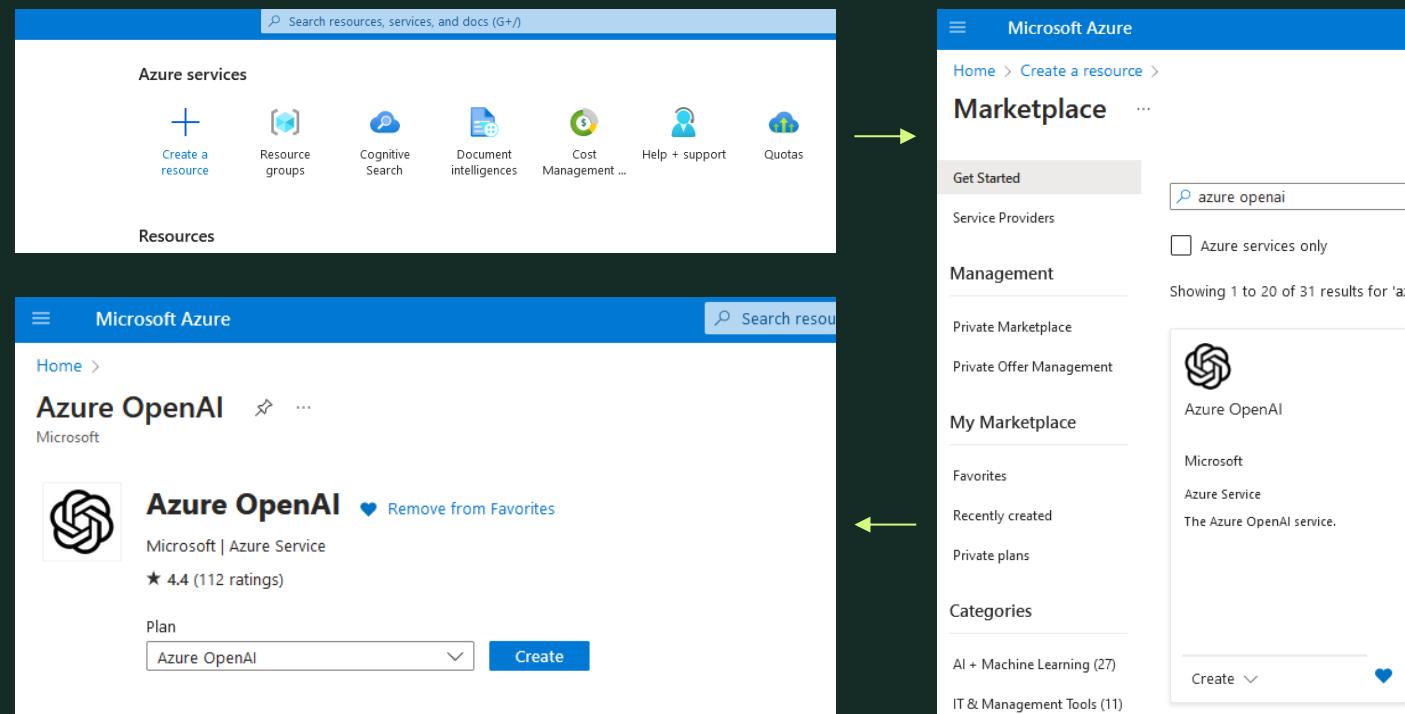


3

Create a resource

Resource is a manageable item that is available through Azure. Virtual machines, storage accounts, web apps, databases, and virtual networks are examples of resources. Resource groups, subscriptions, management groups, and tags are also examples of resources.

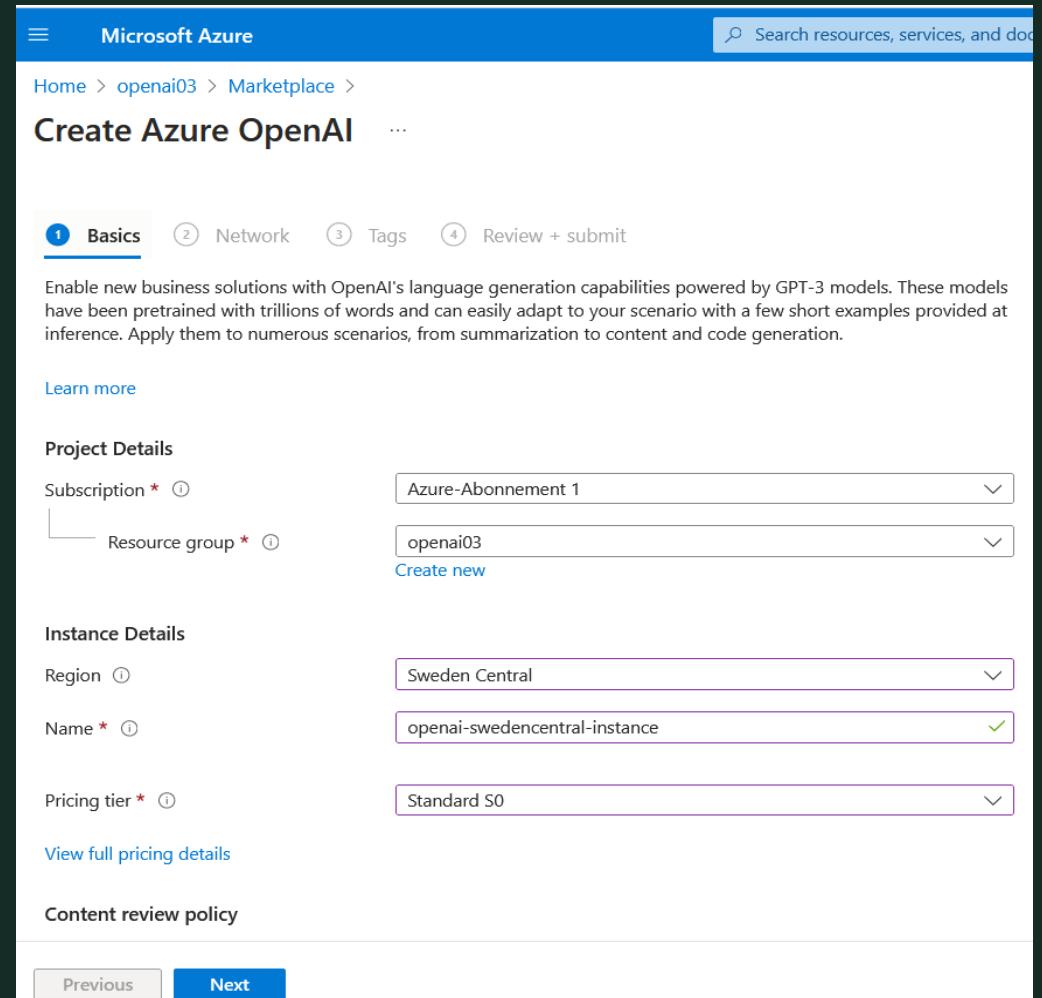
- Sign in with your Azure subscription in the Azure portal.
- Select **Create a resource** and search for the **Azure OpenAI Service**. When you locate the service, select **Create**.



4

On the Create Azure OpenAI page, provide the following information for the fields on the Basics tab:

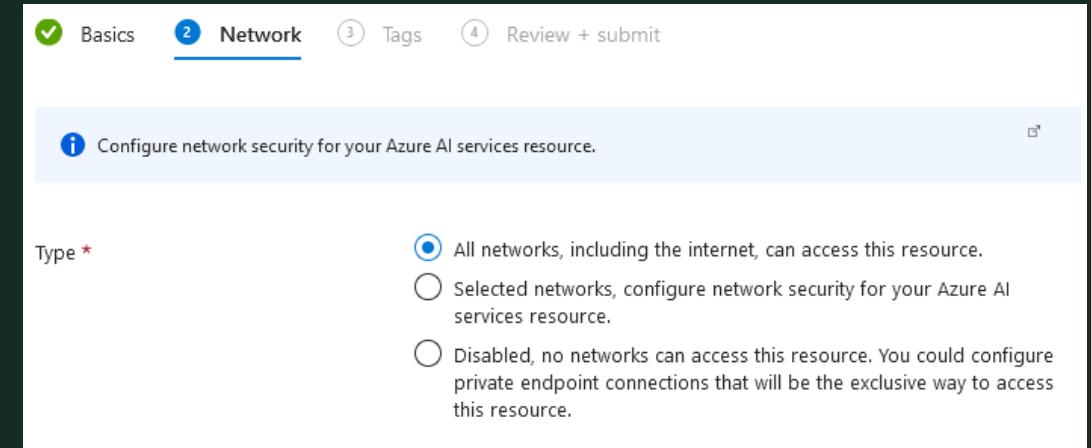
Field	Description
Subscription	The Azure subscription used in your Azure OpenAI Service onboarding application.
Resource group	The Azure resource group to contain your Azure OpenAI resource. You can create a new group or use a pre-existing group.
Region	The location of your instance. Different locations can introduce latency, but they don't affect the runtime availability of your resource.
Name	A descriptive name for your Azure OpenAI Service resource, such as MyOpenAIResource.
Pricing Tier	The pricing tier for the resource. Currently, only the Standard tier is available for the Azure OpenAI Service.



5

Configure network security

Regarding your concerns and security, you can configure the network security in three types. More details about each type can be found [here](#).



6

Select **Next** and configure any **Tags** for your resource, as desired.

7

Select **Next** to move to the final stage in the process: **Review + submit**

8

Confirm your configuration settings, and select **Create**.

9

Deploy a model

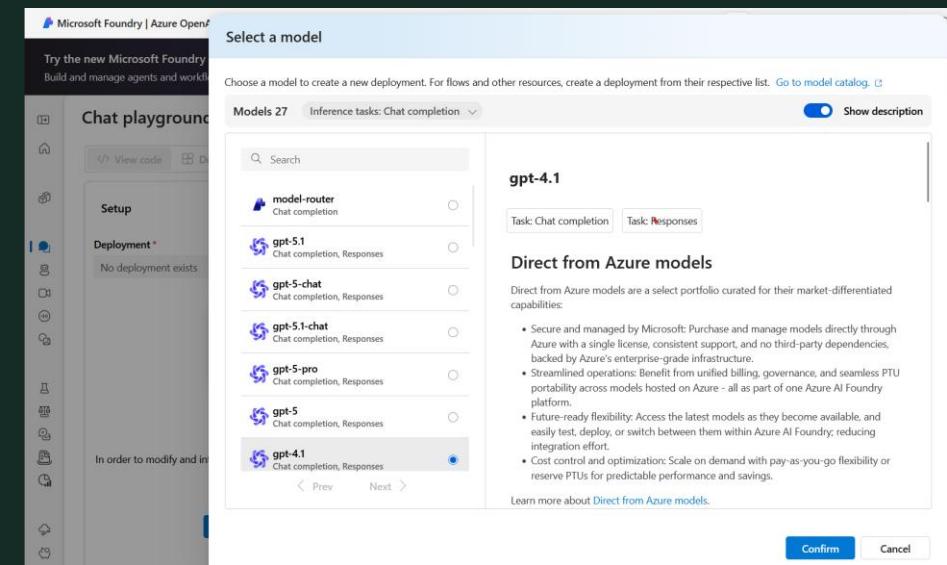
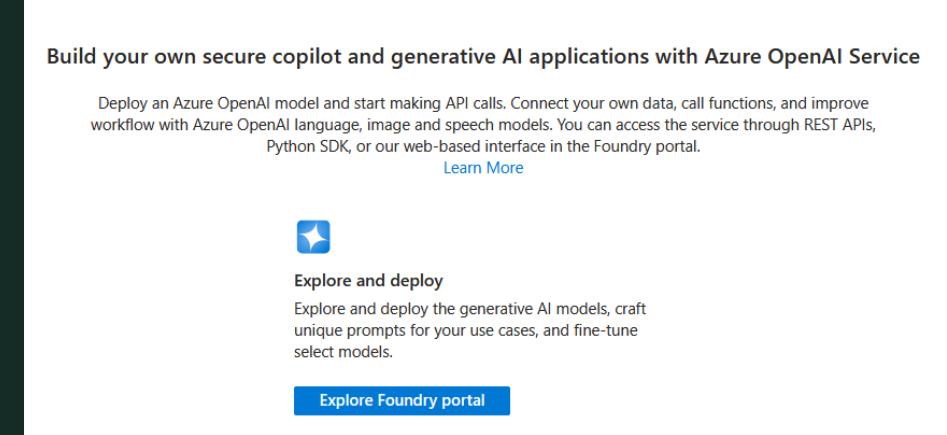
Before you can generate text or inference, you need to deploy a model. You can select from one of several available models in Azure Foundry Portal. To deploy a model, follow these steps:

9.1

Sign in Explore Foundry Portal

9.2

From Chat playground, create a new deployment with model GPT-4.1



9.3

Set the required field as shown in the screenshot:

- Deployment name: gpt-4.1
- Deployment type: Data Zone Standard
- Resource location: <your resource location in Europe, e.g. sweden central, west Europe, etc>

Field	Description
Select a model	Model availability varies by region. For a list of available models per region, see Model summary table and region availability .
Deployment name	Choose a name carefully. The deployment name is used in your code to call the model by using the client libraries and the REST APIs.
Advanced options (Optional)	You can set optional advanced settings, as needed for your resource. - For the Content Filter, assign a content filter to your deployment. - For the Tokens per Minute Rate Limit, adjust the Tokens per Minute (TPM) to set the effective rate limit for your deployment. You can modify this value at any time by using the Quotas menu.

Deploy gpt-4.1

Deployment name*

Deployment type

Data Zone Standard: Pay per API call with higher rate limits. Learn more about [Data Zone deployment types](#).

Data might be processed anywhere within the Microsoft specified data zone, outside of the AI resource's Azure geography, but data storage remains in the AI resource's Azure geography. Learn more about [data residency](#).

Model version upgrade policy

Model version

Resource location

300K tokens per minute quota available for your deployment

Tokens per Minute Rate Limit

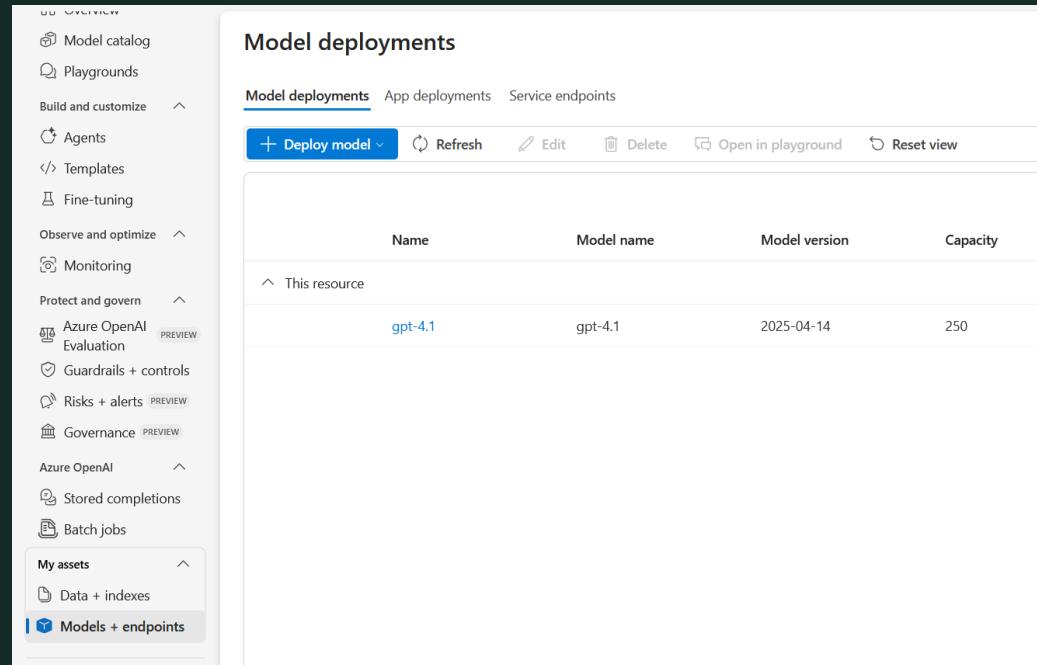
Corresponding requests per minute (RPM) = 300

zappliance | Create and deploy an Azure OpenAI Service resource

10

The deployments table shows a new entry that corresponds to your newly created model.

When the deployment completes, your model deployment status changes to **succeeded**.



The screenshot shows the 'Model deployments' section of the Azure OpenAI Service portal. The left sidebar includes sections for Model catalog, Playgrounds, Build and customize (Agents, Templates, Fine-tuning, Observe and optimize), Protect and govern (Azure OpenAI Evaluation, Guardrails + controls, Risks + alerts, Governance), and Azure OpenAI (Stored completions, Batch jobs). The 'My assets' section is expanded, showing 'Models + endpoints'. The main content area is titled 'Model deployments' and shows a table with the following data:

Name	Model name	Model version	Capacity
gpt-4.1	gpt-4.1	2025-04-14	250

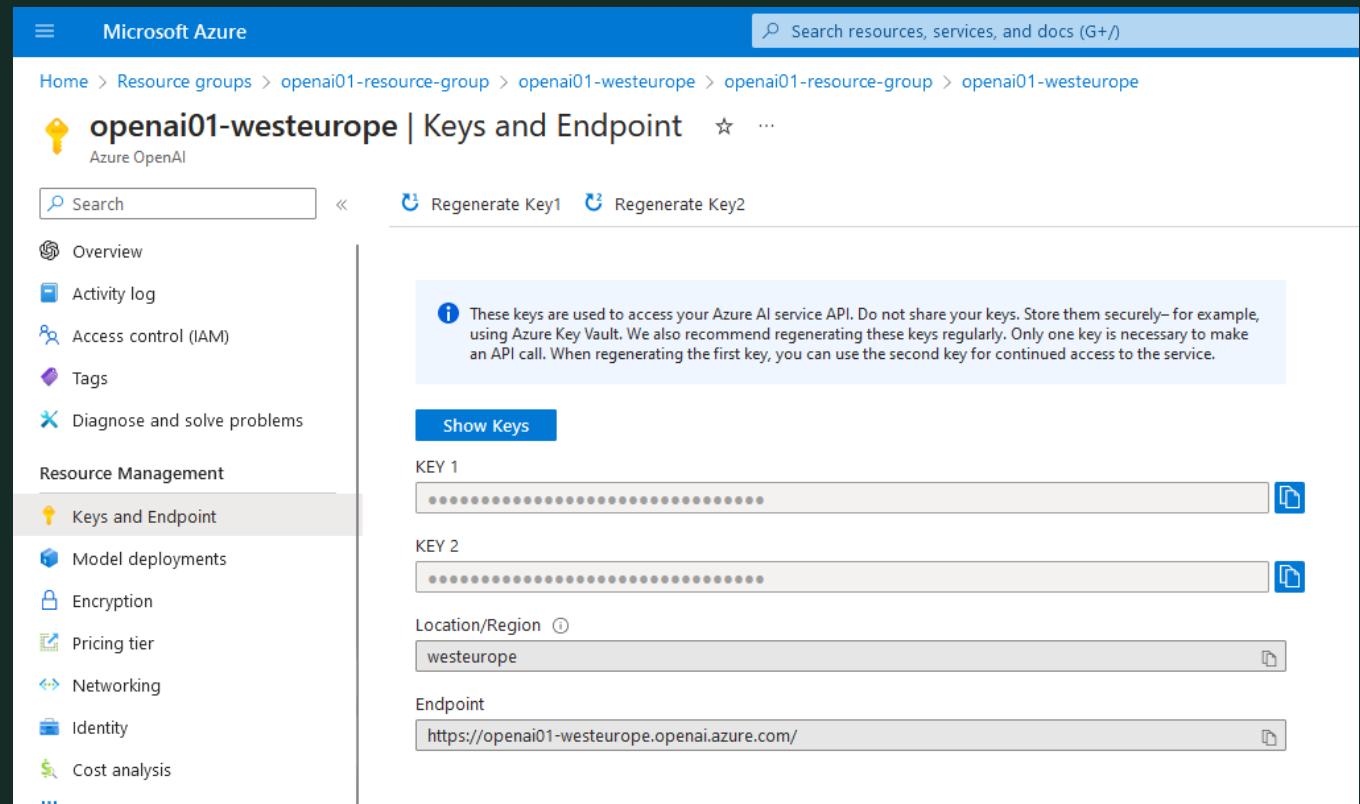
Retrieve key and endpoint

To successfully make a call against the Azure OpenAI service, you'll need the following:

Field	Description
Endpoint	This value can be found in the Keys & Endpoint section when examining your resource from the Azure portal. Alternatively, you can find the value in Azure OpenAI Studio > Playground > Code View. An example endpoint is: https://docs-test-001.openai.azure.com/ .
API-KEY	This value can be found in the Keys & Endpoint section when examining your resource from the Azure portal. You can use either KEY1 or KEY2.
Deployment - name	This value will correspond to the custom name you chose for your deployment when you deployed a model. This value can be found under Resource Management > Deployments in the Azure portal or alternatively under Management > Deployments in Azure OpenAI Studio.

11

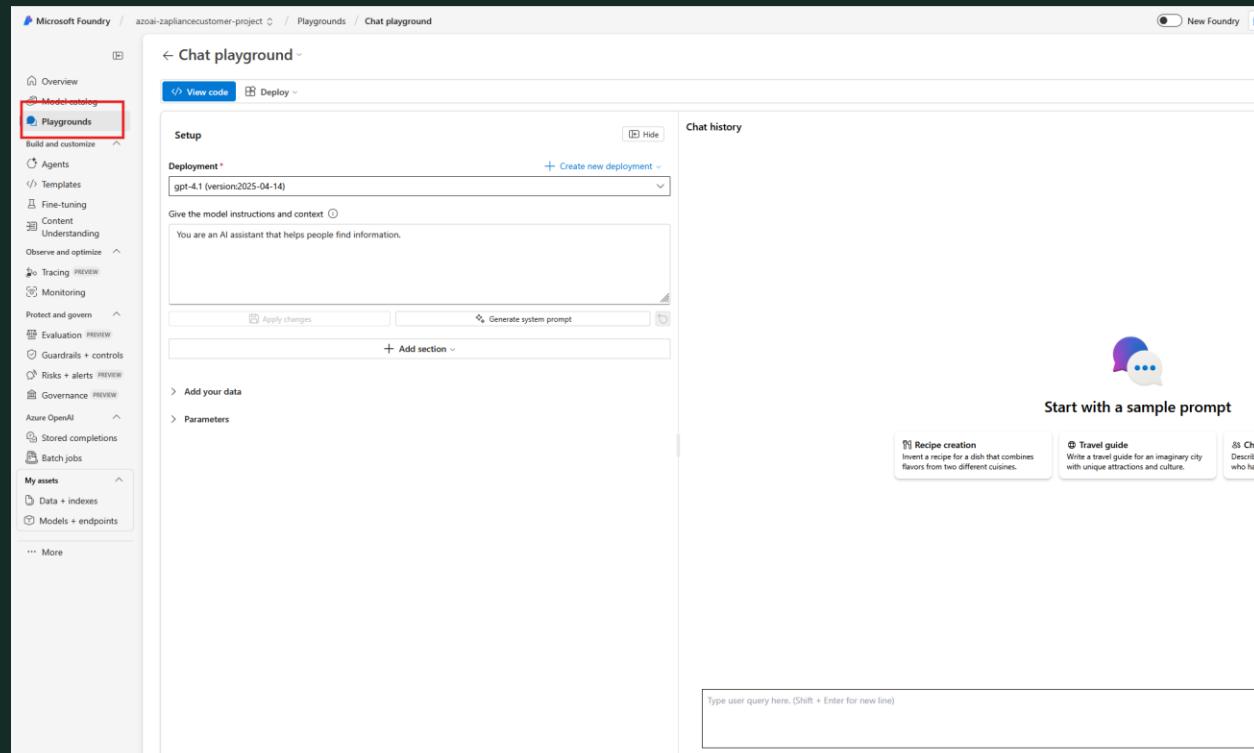
Go to your resource in the Azure portal. The Endpoint and Keys can be found in the Resource Management section. Copy your endpoint and access key as you'll need both for authenticating your API calls. You can use either **KEY1** or **KEY2**. Always having two keys allows you to securely rotate and regenerate keys without causing a service disruption.



Azure OpenAI Playground

zappliance

In the Playground, you can start testing your deployments with different prompts.



You will also notice, on the right-hand side of the **Playground**, a set of parameters to be tuned. Those are necessary in order to make the model more tailored to different use cases

Required API Configuration



We'll need to get the following information from your Azure OpenAI service:

```
# COMPANY_NAME

AZURE_OPENAI_ENDPOINT = <your azure openai endpoint>
AZURE_OPENAI_API_KEY = <your azure openai api key>
AZURE_OPENAI_API_TYPE = azure
AZURE_OPENAI_MODEL = <your openai model name>
AZURE_OPENAI_DEPLOYMENT_NAME = <your openai deployment name>
```

Thank you