


- Search
- Overview
- All resources
 - Use with Foundry
 - Foundry
 - AI Hubs
 - Azure OpenAI
 - AI Search
 - More services
 - Bot services
 - Computer vision
 - Custom vision
 - Content safety
 - Document intelligence
 - Face API
 - Health Insights
 - Machine Learning
 - Immersive reader
 - Language service
 - Speech service
 - Translator
 - Classic AI services
 - Anomaly detector (classic)
 - Azure AI services multi-service account (classic)
 - Content moderator (classic)
 - Language understanding (classic)


Innovate Anywhere with AI

Build intelligent apps faster using prebuilt and custom models, scalable infrastructure, and tools optimized for developers and data teams.




Create a Foundry Resource
Start building a Foundry resource to start building, customizing and evaluating AI solutions using Foundry.

Create a resource



Manage AI Resources
View and manage all your AI services.

View all resources



Explore Best Practices and Guidance
Learn how to securely design, deploy, and manage AI solutions.

View documentation

Give feedback

Home > Microsoft Foundry | Foundry

Microsoft Foundry | Foundry
Default Directory

Search

Create

- Overview
- All resources
- Use with Foundry
 - Foundry
 - AI Hubs
 - Azure OpenAI
 - AI Search
- More services
 - Bot services
 - Computer vision
 - Custom vision
 - Content safety
 - Document intelligence

dev-1141-resource
Foundry

Search

Go to Foundry portal Delete

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Resource visualizer
- Resource Management
- Security
- Monitoring
- Automation
- Help

Essentials

Resource group (move)	: foundry_dev	API Kind	: AIServices
Subscription (move)	: Azure subscription 1	Location	: eastus2
Subscription ID	: b3b8d075-fcce-43b4-bf57-059c905917de	Status	: Succeeded
Endpoints	: Click here to view endpoints		
Tags (edit)	: Add tags		

Build apps and agents in Foundry

Unlock advanced AI capabilities, seamlessly integrate with your systems, and accelerate AI application development with Foundry.

Go to Foundry portal



Deployments

Deployed models [Models](#) [Preview](#) [AI Services](#) [Batch jobs](#)

Search

PTU Calculator

Deploy

Name	Model	Version	Deployment status	Deployment type	Created on ↓	Created by
<input checked="" type="radio"/> gpt-5.5	gpt-5.5	2026-04-24	Succeeded	Data Zone Standard	26. 5. 26. 오후 2:19:45	dev@volt-ai.io
<input type="radio"/> text-embedding-3-large	text-embedding-3-large	1	Succeeded	Data Zone Standard	26. 5. 26. 오후 2:19:03	dev@volt-ai.io
<input type="radio"/> gpt-5.4-pro	gpt-5.4-pro	2026-03-05	Succeeded	Global Standard	26. 5. 26. 오후 2:16:07	dev@volt-ai.io
<input type="radio"/> mistral-document-ai-2512	mistral-document-ai-2512	1	Succeeded	Global Standard	26. 5. 26. 오후 2:15:09	dev@volt-ai.io
<input type="radio"/> DeepSeek-V4-Pro	DeepSeek-V4-Pro	2026-04-23	Succeeded	Global Standard	26. 5. 26. 오후 2:14:13	dev@volt-ai.io

1-5 of 5

< Prev Next >

Search with AI (Ctrl + K)



← claude-opus-4-8

[Details](#) [Deployments](#) [Benchmarks](#) [Responsible AI](#) [License](#)

Claude Opus 4.8 is our most intelligent Opus model and the best generally available model for coding and agents, with deeper reasoning for enterprise workflows.

Anthropic Version: 1

Description

Claude Opus 4.8 is our most intelligent Opus model and the best generally available model for coding and agents, with deeper reasoning for enterprise workflows.

Technical specs

Please refer to the [Claude Opus 4.8 System Card](#).

Training cut-off date

January 2026

Input formats

Image & text input: With powerful vision capabilities, Claude Opus 4.8 can process images and return text outputs to analyze and understand charts, graphs, technical diagrams, reports, and other visual assets.

Text output: Claude Opus 4.8 can output text of a variety of types and formats, such as prose, lists, Markdown tables, JSON, HTML, code in various

Quick facts

Model provider

Type

Lifecycle

Input type

Output type

Context window

Token limits

Pricing

Model ID

azureml/registries/azureml-opic/models/claude-opus-ions/1

Deploy claude-opus-4-8

Deployment name *

claude-opus-4-8

Deployment type

Global Standard

Global Standard: Pay per API call with the highest rate limits. [Learn more about Global deployment types](#). Data might be processed globally, outside of the resource's Azure geography, but data storage remains in the AI resource's Azure geography. [Learn more about data residency](#).

> Model version settings

Tokens per Minute Rate Limit

250000

/2000000

Guardrails

To configure Foundry guardrails with claude-opus-4-8, please use Azure AI Content Safety API, as integrated guardrails are not yet available. [Learn more](#).

Deployments

Deployed models [Models](#) [Preview](#) [AI Services](#) [Batch jobs](#)

Search

PTU Calculator

Deploy

Name	Model	Version	Deployment status	Deployment type	Created on ↓	Created by
<input checked="" type="radio"/> claude-opus-4-8	claude-opus-4-8	1	Failed	Global Standard	26. 6. 4. 오전 8:05:40	dev@volt-ai.io
<input type="radio"/> gpt-5.5	gpt-5.5	2026-04-24	Succeeded	Data Zone Standard	26. 5. 26. 오후 2:19:45	dev@volt-ai.io

claude-opus-4-8

v1 Failed

Deployed 2026. 6. 4., 오전 8:05:40

Open in playground

Edit

Delete

Send a request to this model.


← claude-opus-4-8

Details Metrics


 Request quota  Edit  Delete

Endpoint

Target URI

https://dev-1141-resource.services.ai.azure.com/anthropic/v1/messages 

Key


..... 

Deployment info

Name

claude-opus-4-8

Provisioning state

 Failed