

# 目次

- 第1部: イントロ
- 第3部: 利用料金（の可能性）について
- 第3部: 利用申請
- 第4部: 利用料金発生時の流れ
- 第5部: 代表的サービスの使い方
  - 所有者の追加
  - OpenAIのサービスを利用しよう
  - 仮想マシンを作ろう
  - ノートブックを使おう

# OpenAIのサービスを利用しよう

このスライドは 2025年2月21日 UTokyo Azure 利用説明会のために作成したものです。

Azureの仕様は頻繁に変わりますので、以下で動作しない場合は [Azure OpenAI Serviceのドキュメント](#)を参照してください。

# 目的

- 学習済みモデルを利用する2つの選択肢
  - OpenAIのモデルをすぐに使いたい → Azure OpenAI Service
  - OpenAIを含め様々なモデルを使いたい → Azure AI Foundry
- Azure OpenAI Serviceのロードマップ
  1. リソースグループを作成
  2. モデルカタログからモデルを選択 (たとえば" gpt-4o "を選ぶ)
  3. チャットプレイグラウンドからチャットを実行
  4. 外部のプログラムからWeb APIを通じてサービスを使う
  5. その他の話題
    - 5.1 料金
    - 5.2 1分当たりのトークン制限
    - 5.3 データ、プライバシー、セキュリティ
- その他の学習済みモデルの利用 (Azure AI Foundryの利用)
  - DeepSeek R1

# 1. リソースグループを作成



① Azure Portalを開く  
<https://portal.azure.com/#home>

② “OpenAI”で検索

③ “Azure OpenAI”  
をクリック

The screenshot shows the Microsoft Azure Portal interface. At the top, the search bar contains the text "openai". Below the search bar, a dropdown menu displays search results. The first result is "Azure OpenAI", which is highlighted. Other results include "Azure AI services", "Azure AI Foundry", and "Azure Machine Learning OpenAI". On the left side of the page, there is a "リソース" (Resources) section with a list of recent resources, including "ml-workspace", "ml\_test", "llmjp", and "gpt-4o-test-1". At the bottom of the page, there are navigation icons for "サブスクリプション" (Subscription), "リソース グループ" (Resource Group), "すべてのリソース" (All Resources), and "ダッシュボード" (Dashboard).

# Azure AI services | Azure OpenAI

検索 + 作成 削除されたリソースの管理 ビューの管理 更新 CSV にエクスポート クエリを開く タグの割り当て 削除

- 概要
- All Azure AI services
- Azure AI services
  - Azure AI services
  - Azure OpenAI**
  - AI Search
  - Computer Vision
  - Face API
  - Custom Vision
  - 音声サービス
  - 言語サービス
  - 翻訳
  - Document intelligence
  - ボット サービス
  - Anomaly Detector
  - Content Moderator
  - Health Insights

任意のフィールドのフィルター... サブスクリプション 次の値と等しい すべて 種類 次の値と等しい すべて フィルターの追加 増やす (2)

0 件中 0 ~ 0 件のレコードを表示しています。 グループ化なし リストビュー

名前 ↑↓	サブタイプ ↑↓	場所 ↑↓	カスタム ドメイン名 ↑↓	価格レベル ↑↓	状態 ↑↓
-------	----------	-------	---------------	----------	-------



## 表示する Azure OpenAI がありません

さまざまな自然言語タスクを実行します。

[Azure OpenAI の作成](#)

[詳細情報](#)

[フィードバックの送信](#)

# Azure OpenAI の作成

**⚠** このステップを変更すると、後で選択する内容がリセットされる可能性があります  
[詳細情報](#)

## プロジェクトの詳細

サブスクリプション \* ⓘ

リソース グループ \* ⓘ  [新規作成](#)

## インスタンスの詳細

リージョン ⓘ

名前 \* ⓘ

価格レベル \* ⓘ

[価格の詳細を表示](#)

## コンテンツ レビュー ポリシー

Azure OpenAI Service の有害な使用を検出して軽減するために、Microsoft は、お客様が完了 API およびイメージ生成 API に送信したコンテンツと、それが返送するコンテンツをログに記録します。サービスのフィルターによってコンテンツにフラグが設定されている場合は、Microsoft の常勤従業員によってレビューされる可能性があります。

モデル稼働させるリソースだけでなく、パブリックIPやディスクなど、いろいろなリソースがこのモデルに紐づきます。**課金を完全に止めるなら、リソースグループごと削除するのが簡単です。**逆に言えば、不要になったら丸ごと消していい単位でリソースグループを分けておくことをお勧めします。

そのサービスのリソースがあるリージョンのみ  
選べる  
(たとえばJapan Westは選べない)

# Product Availability by Region

With 60+ announced regions, more than any other cloud provider, Azure makes it easy to choose the datacenter closest to your users.

© Microsoft 2024. All rights reserved. Release dates, features, and requirements are subject to change prior to final announcement. MICROSOFT MAKES NO WARRANTIES, EXPRESSED, IMPLIED, OR STATUTORY, AS TO THE INFORMATION ON THIS PAGE.

\* Reserved Access Regions: Certain regions are access restricted to support specific customer scenarios, for example in the United States, the United Kingdom, and the European Union.

Azure OpenAI Serviceに限らず、どのリージョンにどのようなリソースがあるかはこのページで確認できる。  
SKUとはStock Keeping Unitの略。

Search Products  Select a Geography

Geographies		Japan		Korea	
Products	Product SKU	Japan East	Japan West	Korea Central	Korea South
	Language Understanding (LUIS)	●	●	●	
	Video Indexer	●	●	●	
	Azure AI Metrics Advisor	●		●	
	Azure <b>OpenAI</b> Service	●		●	
	Speaker recognition				
	Azure AI Speech	●	●	●	
	Language service	●			
	Azure Video Analyzer				
	Azure AI Video Indexer	●	●	●	
	Azure API Center				
	Azure Active Directory B2C				
	Azure Advisor				

Japan WestはOpenAI Serviceのリソースがない

# Azure OpenAI の作成 ...

- 1 基本
- 2 ネットワーク**
- 3 Tags
- 4 レビューおよび送信

**i** Configure network security for your Azure AI services resource.

- 種類 \*
- インターネットを含むすべてのネットワークがこのリソースにアクセスできます。
  - Selected networks, configure network security for your Azure AI services resource.
  - 無効になっています。どのネットワークもこのリソースにアクセスできません。このリソースにアクセスする唯一の方法となるプライベート エンドポイント接続を構成できます。

ホーム > Azure AI services | Azure OpenAI >

# Azure OpenAI の作成

基本 ネットワーク **3 Tags** 4 レビューおよび送信

タグは名前と値のペアで、同じタグを複数のリソースやリソース グループに適用することでリソースを分類したり、統合した請求を表示したりできるようにします。 [タグに関する詳細情報](#)

タグを作成してから別のタブでリソースの設定を変更すると、タグは自動的に更新されることにご注意ください。

名前 ⓘ	値 ⓘ	リソース
<input type="text"/>	: <input type="text"/>	Azure AI services

ホーム > Azure AI services | Azure OpenAI >

# Azure OpenAI の作成

- 基本
- ネットワーク
- Tags
- 4 レビューおよび送信**

Automation のテンプレートをダウンロードする

## 使用条件

"作成" をクリックすることで、お客様は (a) 上記の Marketplace のオファリングに関連する法律条項とプライバシーに関する声明に同意し、(b) Microsoft より、そのオファリングに関連する料金が、現在の支払い方法に対して Azure サブスクリプションと同じ請求頻度で請求されることを認め、かつ、(c) Microsoft がお客様の連絡先情報、使用量情報、取引に関する情報を、サポート、請求、その他の取引上のアクティビティを目的として、オファリングのプロバイダーと共有する可能性があることに同意するものとします。Microsoft は、サードパーティのオファリングに対する権利は提供しません。その他の詳細については、[Azure Marketplace 使用条件](#)を参照してください。

## 基本

サブスクリプション	1923306880-YamakataLab
リソース グループ	openai_test1
リージョン	Japan East
名前	gpt-4o-test-1
価格レベル	Standard S0

## ネットワーク

ホーム >

# Microsoft.CognitiveServicesOpenAI-20250129153123 | 概要

検索 削除 キャンセル 再デプロイ ダウンロード 最新の情報に更新

## 概要

- 入力
- 出力
- テンプレート

## デプロイが進行中です

デプロイ名 : Microsoft.CognitiveServi... 開始日時 : 2025/1/29 16:09:37  
 サブスクリプション : 1923306880-YamakataL... 関連付け ID : 48828a5b-d1e5-4953-9b8f-...  
 リソース グループ : openai\_test1

### デプロイの詳細

リソース	種類	状態
表示するリソースがありません。		

フィードバックの送信

🗨️ デプロイに関するエクスペリエンスについてお聞かせください



**Microsoft Defender for Cloud**  
 アプリおよびインフラストラクチャをセキュリティで保護します  
[次に移動: Microsoft Defender for Cloud >](#)

**無料の Microsoft チュートリアル**  
[今すぐ学習を開始する >](#)

**エキスパートとの共同作業**  
 Azure のエキスパートは、Azure での資産の管理を支援し、サポートの最前線となるサービスプロバイダー パートナーです。  
[Azure Expert の検索 >](#)



ホーム >

# Microsoft.CognitiveServicesOpenAI-20250129153123 | 概要

デプロイ

検索 削除 キャンセル 再デプロイ ダウンロード 最新の情報に更新

- 概要
- 入力
- 出力
- テンプレート

## ✓ デプロイが完了しました

デプロイ名 : Microsoft.CognitiveServi... 開始日時 : 2025/1/29 16:09:37  
サブスクリプション : 1923306880-YamakataL... 関連付け ID : 48828a5b-d1e5-4953-9b8f-...  
リソース グループ : openai\_test1

> デプロイの詳細

✓ 次の手順

**リソースに移動**

フィードバックの送信

🗣️ デプロイに関するエクスペリエンスについてお聞かせください



### Cost Management

予算内に収め、予期しない料金が請求されないように通知を受け取ります。  
[コストのアラートを設定 >](#)



### Microsoft Defender for Cloud

アプリおよびインフラストラクチャをセキュリティで保護します  
[次に移動: Microsoft Defender for Cloud >](#)

### 無料の Microsoft チュートリアル

[今すぐ学習を開始する >](#)

### エキスパートとの共同作業

Azure のエキスパートは、Azure での資産の管理を支援し、サポートの最前線となるサービスプロバイダー パートナーです。

[Azure Expert の検索 >](#)

ホーム > Microsoft.CognitiveServicesOpenAI-20250129153123 | 概要 >

gpt-4o-test-1 Azure OpenAI

検索 Go to Azure AI Foundry portal 削除

- 概要
- アクティビティ ログ
- アクセス制御 (IAM)
- タグ
- 問題の診断と解決
- > リソース管理
- > Security
- > 監視
- > オートメーション
- > ヘルプ

基本 JSON ビュー

リソース グループ (移動) <a href="#">openai_test1</a>	API Kind OpenAI
状態 アクティブ	価格レベル Standard
場所 Japan East	エンドポイント <a href="#">エンドポイントを表示するには、ここをクリックします</a>
サブスクリプション (移動) <a href="#">1923306880-YamakataLab</a>	キーの管理 <a href="#">キーを管理するにはここをクリック</a>
サブスクリプション ID a3b0857d-8c96-43a1-b2ee-d787b83bbe5e	
タグ (編集) <a href="#">タグの追加</a>	

開始する Develop Explore and deploy

### Build your own secure copilot and generative AI applications with Azure OpenAI Service

Deploy an Azure OpenAI model and start making API calls. Connect your own data, call functions, and improve workflow with Azure OpenAI language, image and speech models. You can access the service through REST APIs, Python SDK, or our web-based interface in the Azure OpenAI Studio.

[Learn More](#)



ホーム > Microsoft.CognitiveServicesOpenAI-20250129153123 | 概要 >

**gpt-4o-test-1** Azure OpenAI ✕

Go to Azure AI Foundry portal 削除

- 概要**
- アクティビティ ログ
- アクセス制御 (IAM)
- タグ
- 問題の診断と解決
- > リソース管理
- > Security
- > 監視
- > オートメーション
- > ヘルプ

[1923306880-YamakataLab](#) [キーを管理するにはここをクリック](#)

サブスクリプション ID  
a3b0857d-8c96-43a1-b2ee-d787b83bbe5e

タグ [\(編集\)](#)  
[タグの追加](#)

**開始する** Develop Explore and deploy

## Build your own secure copilot and generative AI applications with Azure OpenAI Service

Deploy an Azure OpenAI model and start making API calls. Connect your own data, call functions, and improve workflow with Azure OpenAI language, image and speech models. You can access the service through REST APIs, Python SDK, or our web-based interface in the Azure OpenAI Studio. [Learn More](#)

**Monitor your Azure OpenAI usage**  
Monitor, assess and optimize your usage of the Azure OpenAI service API with an out-of-the-box metrics dashboard

**Azure OpenAI Metrics Dashboard**

AI ソリューションを構築するためのより多くのモデル、機能、ツールをお試しになるか、追加の手順なしで既存のデプロイを導入してください。

# チャットプレイグラウンド

コードの表示 デプロイ インポート エクスポート プロンプトのサンプル

## セットアップ

非表示

デプロイ\* 新しいデプロイの作成

デプロイが存在しません

モデルに指示とコンテキストを与える

情報を見つけるのに役立つ AI アシスタントです。

変更の適用 プロンプトを生成する

セクションの追加

ソースデータが追加されている場合、一部のショットの例は使用されません。以前の例はすべてクリアされました。データを削除すると、例を追加できます。

データを追加する

パラメーター

## チャットの履歴



## サンプルプロンプトを使って始める

**Dialogue creation**  
Create a conversation between two characters who are meeting for the first time in a mysterious place.

**Character development**  
Describe a day in the life of a superhero who has lost their powers.

**Recipe creation**  
Invent a recipe for a dish that combines flavors from two different cuisines.

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)



## チャットプレイグラウンドのヘルプ

関連リソース

チャットのクイックスタート

ホーム

概要

モデルカタログ

プレイグラウンド

チャット

アシスタント プレビュー

リアルタイムオーディオ プレビュー

イメージ

入力候補

ツール

微調整

Azure OpenAI Service プレビュー

格納された入力候補 プレビュー

バッチ ジョブ

共有リソース

デプロイ

クォータ

安全性とセキュリティ

データ ファイル

ベクトルストア プレビュー

## 2. カタログからモデルを選ぶ

① AI ソリューションを構築するためのより多くのモデル、機能、ツールをお試しになるか、追加の手順なしで既存のデプロイを導入してください。

## チャットプレイグラウンド

</> コードの表示 🗑️ デプロイ → インポート → エクスポート 🔗 プロンプトのサンプル ...

セットアップ 🗑️ 非表示

デプロイ\* + 新しいデプロイの作成 ▼

デプロイが存在しません ▼

モデルに指示とコンテキストを与える ①

情報を見つけるのに役立つ AI アシスタントです。

📄 変更の適用 🔗 プロンプトを生成する 🔄

+ セクションの追加 ▼

① ソースデータが追加されている場合、一部のショットの例は使用されません。以前の例はすべてクリアされました。データを削除すると、例を追加できます。

> データを追加する

> パラメーター

チャットの履歴 🗑️ ⚙️ 🔊



### サンプルプロンプトを使って始める

- 🗨️ **Dialogue creation**  
Create a conversation between two characters who are meeting for the first time in a mysterious place.
- 👤 **Character development**  
Describe a day in the life of a superhero who has lost their powers.
- 🍳 **Recipe creation**  
Invent a recipe for a dish that combines flavors from two different cuisines.

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)

🗣️ ▶️

## チャットプレイグラウンドのヘルプ

- ▼ 関連リソース
- [チャットのクイックスタート](#)

- 🏠 ホーム
- 📄 概要
- 🗑️ **モデルカタログ**
- 📄 プレイグラウンド ^
- 🗣️ チャット**
- 🗣️ アシスタント プレビュー
- 🎧 リアルタイムオーディオ プレビュー
- 🖼️ イメージ
- 🗣️ 入力候補
- 🔧 ツール ^
- 🔧 微調整
- Azure OpenAI Service プレビュー
- 🗣️ Service 評価
- 🗣️ 格納された入力候補 プレビュー
- 📄 バッチ ジョブ
- 📄 共有リソース ^
- 🗣️ デプロイ
- 🗣️ クォータ
- 🛡️ 安全性とセキュリティ
- 📄 データ ファイル
- 🗣️ ベクトルストア プレビュー

- ホーム
- 概要
- モデル カタログ**
- ブレイグラウンド
- チャット
- アシスタント プレビュー
- リアルタイム オーディオ プレビュー
- イメージ
- 入力候補
- ツール
- 微調整
- Azure OpenAI Service 評価 プレビュー
- 格納された入力候補 プレビュー
- バッチ ジョブ
- 共有リソース
- デプロイ
- クォータ
- 安全性とセキュリティ
- データ ファイル
- ベクトル ストア プレビュー

## カスタム AI ソリューションを構築するための適切なモデルを見つける

### アナウンス

**o1 モデルを体験する**

o1 シリーズは、科学とコーディングの問題を解決するために強化された推論機能を備えています。

[モデルのチェックアウト](#) [ブログを読む](#)

**プロジェクトを作成して、より多くのモデルを取得する**

Meta、Mistral などのモデルを探索し、Azure AI サービスなどの新機能にアクセスできます。

[その他のモジュールを探索](#) [Learn more](#)

推論タスク ▼ タスクの微調整 ▼

検索 モデル 22

- |  |                                       |   |
|--|---------------------------------------|---|
| <b>o1</b><br>チャットの完了                               | <b>o1-mini</b><br>チャットの完了             | <b>gpt-4o-audio-preview</b><br>audio-generation |
| <b>gpt-4o-realtime-preview</b><br>audio-generation | <b>gpt-4o</b><br>チャットの完了              | <b>gpt-4o-mini</b><br>チャットの完了                   |
| <b>o1-preview</b><br>チャットの完了                       | <b>gpt-4</b><br>チャットの完了               | <b>gpt-4-32k</b><br>チャットの完了                     |
| <b>text-embedding-3-large</b><br>埋め込み              | <b>text-embedding-3-small</b><br>埋め込み | <b>tts</b><br>テキスト読み上げ                          |
| <b>tts-hd</b><br>テキスト読み上げ                          | <b>whisper</b><br>音声認識                | <b>dall-e-3</b><br>テキストから画像へ                    |
| <b>dall-e-2</b>                                    | <b>text-embedding-ada-002</b>         | <b>davinci-002</b>                              |

### モデルのヘルプ

- How to use the model catalog

Search by name, filter, or browse to find the right model for your use case. Click to see model details and specifications from the publisher and to deploy a model.

Relevant resources

# 2025年2月現在、利用可能なモデル [\[Azureのドキュメント\]](#)

モデル	説明
o シリーズ モデル	高度な問題解決、増強された集中力と能力を備えた推論モデル。
GPT-4o、GPT-4o mini、GPT-4 Turbo	最新の最も能力の高い Azure OpenAI モデルであり、テキストと画像の両方を入力として受け入れることができるマルチモーダルバージョンを備えています。
GPT-4o audio	"音声出力" の会話のやり取り、またはオーディオ生成をサポートする
GPT-4	GPT-3.5 を基に改善され、自然言語とコードを理解し、生成できるモデルのセット。
GPT-3.5	GPT-3 を基に改善され、自然言語とコードを理解し、生成できるモデルのセット。
埋め込み	テキストを数値ベクトル形式に変換して、テキストの類似性を促進できるモデルのセット。
DALL-E	自然言語からオリジナルの画像を生成できるモデルのシリーズ。
Whisper	音声を文字起こしして音声テキスト変換を翻訳できる一連のモデル。
テキスト読み上げ (プレビュー)	テキストを音声に合成できるプレビュー段階の一連のモデル。

よく使われるのは  
GPT-4o

O1系は利用登録が  
必要



仮  
ご

gpt-4o - Azure OpenAI Service x Product Availability by Region x Azure OpenAI Service - Pricing | x Azure OpenAI Service モデル - Azu x Azure OpenAI Service モデル - Azu x

https://ai.azure.com/resource/models/gpt-4o/version/2024-11-20/registry/azure-openai?wsid=/subscriptions/a3b0857d-8c96-43a1-b2ee-d787b83bbe5e/resourceGroups/openai\_test1/p...

Azure AI Foundry | Azure OpenAI Service / gpt-4o-test-1 / モデル / gpt-4o

すべてのリソース 5

gpt-4o-test-1 (japaneast, S0)

### gpt-4o

デプロイ 微調整

詳細 既存のデプロイ コード サンプル ライセンス

gpt-4o offers a shift in how AI models interact with multimodal inputs. By seamlessly combining text, images, and audio, gpt-4o provides a richer, more engaging user experience.

Matching the intelligence of gpt-4 turbo, it is remarkably more efficient, delivering text at twice the speed and at half the cost. Additionally, GPT-4o exhibits the highest vision performance and excels in non-English languages compared to previous OpenAI models.

gpt-4o is engineered for speed and efficiency. Its advanced ability to handle complex queries with minimal resources can translate into cost savings and performance.

The introduction of gpt-4o opens numerous possibilities for businesses in various sectors.

クイックファクト OpenAI

gpt-4o chat-completion

最後のトレーニング

### モデルバージョン

リージョンの可用性に関する詳細情報

Japan East

モデルバージョン	可用性	ライフサイクル	最大要求	廃止の予定
2024-05-13	グローバル標準、プロビジ	一般提供	Input: 128000	2025年5月20日 (火)

モデル ID

コードでモデルを配置するときにこのモデル ID を参照する

```
azureml://registries/azure-openai/models/gpt-4o/versions/2024-11-20
```

### データ、メディア、言語

Property	説明
サポートされるデータ型	入力 出力

リージョンによってはすべてのモデルが選べるわけではない

- ホーム
- 概要
- モデルカタログ
- ブレイクラウンド
- チャット
- アシスタント プレビュー
- リアルタイム オーディオ プレビュー
- イメージ
- 入力候補
- ツール
- 微調整
- Azure OpenAI Service 評価
- 格納された入力候補
- バッチ ジョブ
- 共有リソース
- デプロイ
- クォータ
- 安全性とセキュリティ
- データ ファイル
- ベクトルストア

## gpt-4o

デプロイ 微調整

詳細 既存のデプロイ コード

gpt-4o offers a shift in how AI models process text, images, and audio, gpt-4o provides a significant performance boost, matching the intelligence of gpt-4o but at half the cost. Additionally, gpt-4o is engineered for speed and efficiency, meaning non-English languages compared to previous models. The introduction of gpt-4o opens up new possibilities for AI applications that require speed and minimal resources can translate it into a wide range of languages.

### モデル バージョン

リージョンの可用性に関する詳細

Japan East

### モデル gpt-4o をデプロイする

デプロイ名\*

デプロイの種類

Global Standard: 最も高い制限が適用された API 呼び出しあたりの支払。グローバル展開の種類の詳細をご覧ください。  
データはリソースの Azure 地域の外部でグローバルに処理される可能性があります。データストレージは AI リソースの Azure 地域に残ります。データ所在地の詳細をご覧ください。

#### デプロイの詳細

モデルバージョン	2024-11-20	AI リソース	東アジア 1 (東アジア 1) 10233-m58are1c-eastus2
容量	1分あたりのトークン数 (TPM)	リソースの場所	East US 2

Japan Eastにはリソースがない

処理はEast US 2で行われる

### データ、メディア、言語

Property	説明
サポートされるデータ型	入力 出力

# リージョンによるモデルの対応

- [Azure OpenAI Service モデル - Azure OpenAI | Microsoft Learn](#)

## Global-Standard モデルの提供状況

🔍 テーブルを展開する

リージョン	o1-preview、 2024-09-12	o1-mini、 2024-09-12	gpt-4o、 2024年5月13日	gpt-4o、 2024-08-06	gpt-4o、 2024-11-20	gpt-4o-mini、 2024-07-18	gpt-4o-realtime-preview、 2024-10-01	gpt-4o-realtime-preview、 2024-12-17
australiaeast	-	-	✓	✓	-	✓	-	-
brazilsouth	-	-	✓	✓	-	✓	-	-
canadaeast	-	-	✓	✓	-	✓	-	-
eastus	✓	✓	✓	✓	✓	✓	✓	✓
eastus2	✓	✓	✓	✓	✓	✓	✓	✓
francecentral	-	-	✓	✓	-	✓	-	-
germanywestcentral	-	-	✓	✓	-	✓	-	-
japaneast	-	-	✓	✓	-	✓	-	-
koreacentral	-	-	✓	✓	-	✓	-	-

Japan East(こは2024-11-20の  
gpt-4o(はない  
→ East US 2が自動選択される

# 2025年2月現在、GPT-o1系はアクセス要求が必要

- 「デプロイ」をクリックすると「このモデルを使用するには登録が必要です」と出る→「アクセスの要求」をクリック

モデル o1 をデプロイする

このモデルを使用するには登録が必要です  
アクセスは、Microsoft の資格条件に従って付与されます。現時点では、アクセスの資格がない可能性があります。既にアクセスを要求している場合は、申請を確認していますので、しばらくお待ちください。

[アクセスの要求](#)

取り消し



フォームを  
入力して登録

Request Access to the Azure OpenAI Service  
o\* family of models

\* 必須

Please read all instructions carefully and complete form as instructed

The Azure OpenAI Service o\* family of models require registration. Due to high demand, customers may remain on the waitlist until space is available.

Other models (listed [here](#)) do not require registration. [Learn more about limited access to Azure OpenAI Service here.](#)

Please visit the [Overview of Responsible AI Practices](#) and [Azure OpenAI's transparency note](#), which provides information and guidelines for responsible use of the service as well as system limitations that may be applicable to your scenario.

Possible causes for a denied application:

1) You submitted an invalid subscription2) You submitted with a personal email (Example: [@gmail.com](#), [@yahoo.com](#), [@hotmail.com](#), etc.)

Updates about your request: Please add [maccount@microsoft.com](#) and [csqate@microsoft.com](#) to your Safe Senders list to ensure you receive all correspondence pertaining to your registration, including the verification email from Microsoft Vetting Service.

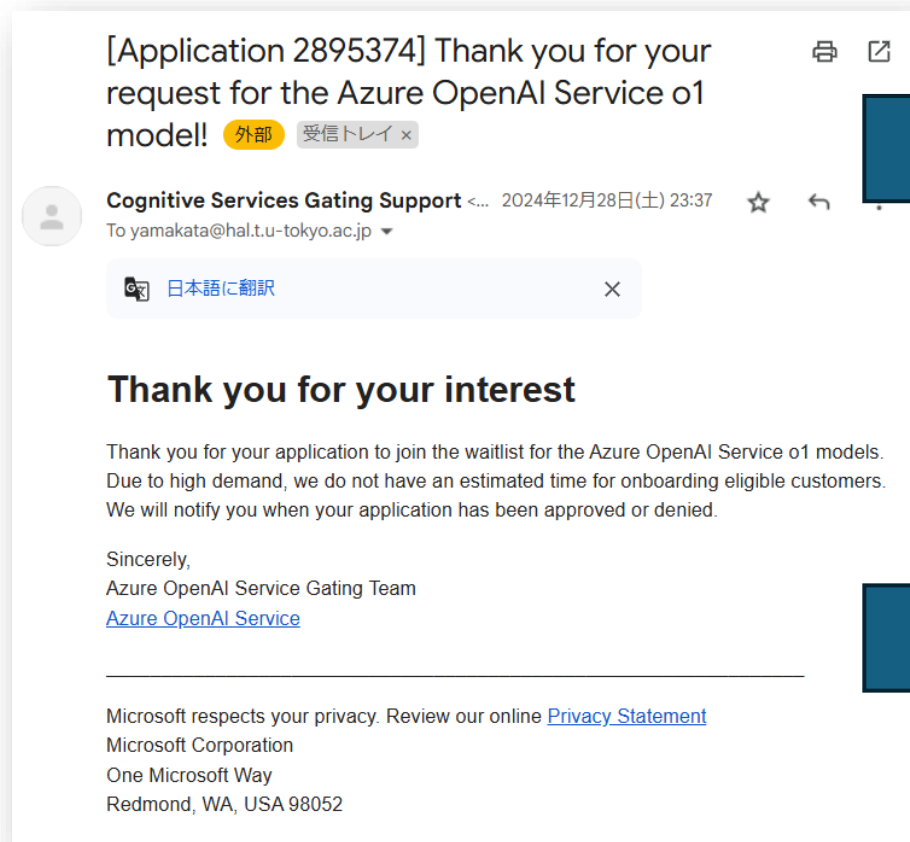
1  
Your First Name \*

回答を入力してください

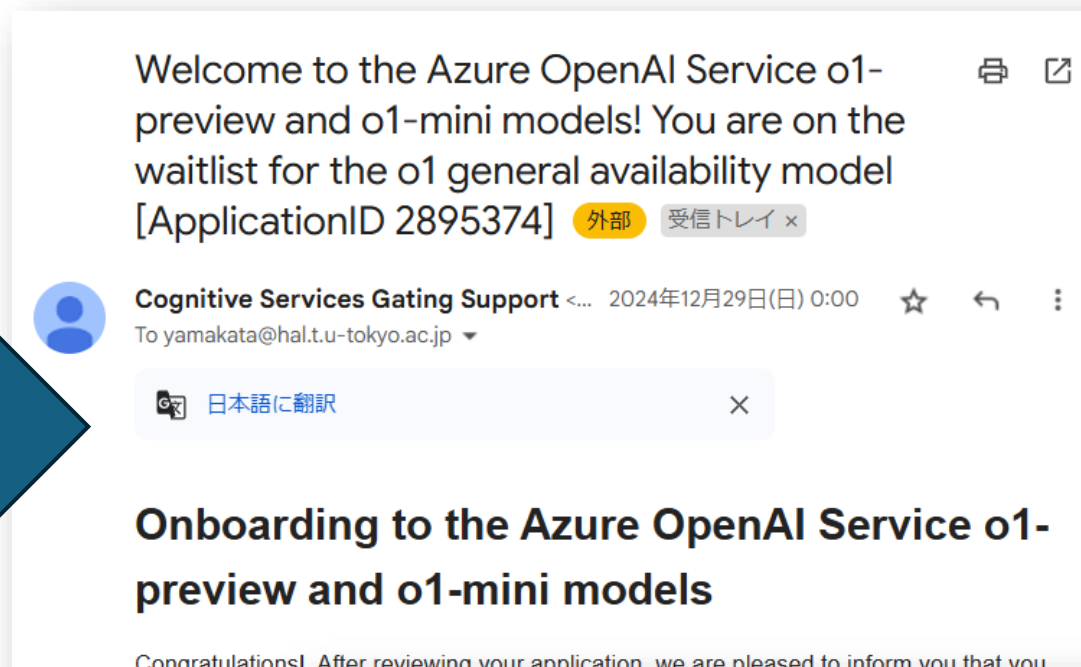
# しばらくすると承認される

私の場合…

- O1-previewとo1-miniはすぐ承認
- O1は15日後に承認



30分後



18日後



- ホーム
- 概要
- モデル カタログ
- ブレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイム オーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
  - Azure OpenAI Service 評価 プレビュー
  - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ**
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトル ストア プレビュー

## ← gpt-4o-test

詳細    メトリック    リスクと安全性

ブレイグラウンドで開く    編集    削除

**配置情報**

<b>名前</b> gpt-4o-test	<b>プロビジョニングの状態</b> 成功
<b>展開の種類</b> グローバル標準	<b>作成日</b> 2025-01-30T08:19:22.6269308Z
<b>作成者</b> 1923306880@utac.u-tokyo.ac.jp	<b>変更日</b> Jan 30, 2025 5:19 PM
<b>更新者</b> 1923306880@utac.u-tokyo.ac.jp	<b>バージョン更新ポリシー</b> 新しい既定のバージョンが使用可能になったら

---

<b>レート制限 (1分あたりのトークン数)</b> 10,000	<b>レート制限 (1分あたりの要求数)</b> 60
--------------------------------------	--------------------------------

---

<b>モデル名</b> gpt-4o	<b>モデルバージョン</b> 2024-11-20
<b>ライフサイクルの状態</b> GenerallyAvailable	<b>作成日</b> Dec 3, 2024 9:00 AM
<b>更新日</b> Dec 3, 2024 9:00 AM	<b>モデルの廃止日</b> Aug 20, 2025 9:00 AM

**エンドポイント**

**ターゲット URI**  
https://19233-m58are1c-eastus2.cognitiveservices.azure.com/...

**キー**  
.....

**監視と安全性**

**コンテンツ フィルター**  
DefaultV2

**アプリケーション開発に役立つリンク**

[コード サンプル リポジトリ](#)   [チュートリアル](#)

## デプロイのヘルプ

▼ **Azure AI Foundry でのデプロイとは?**

Azure AI Foundry では、大規模言語モデル (LLM)、フロー、Web アプリのデプロイがサポートされています。LLM またはフローをデプロイすると、それを Web サイトやアプリケーション、またはその他の運用環境で使用できるようになります。

[詳細情報](#)

> **関連リソース**

### 3. チャットを行う



- ホーム
- 概要
- モデル カタログ
- ブレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイム オーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
  - Azure OpenAI Service 評価 プレビュー
  - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ**
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトル ストア プレビュー

## ← gpt-4o-test

詳細   メトリック   リスクと安全性

**ブレイグラウンドで開く**   編集   削除

### 配置情報

<b>名前</b>	gpt-4o-test	<b>プロビジョニングの状態</b>	成功
<b>展開の種類</b>	グローバル標準	<b>作成日</b>	2025-01-30T08:19:22.6269308Z
<b>作成者</b>	1923306880@utac.u-tokyo.ac.jp	<b>変更日</b>	Jan 30, 2025 5:19 PM
<b>更新者</b>	1923306880@utac.u-tokyo.ac.jp	<b>バージョン更新ポリシー</b>	新しい既定のバージョンが使用可能になったら

---

<b>レート制限 (1分あたりのトークン数)</b>	10,000	<b>レート制限 (1分あたりの要求数)</b>	60
----------------------------	--------	--------------------------	----

---

<b>モデル名</b>	gpt-4o	<b>モデルバージョン</b>	2024-11-20
<b>ライフサイクルの状態</b>	GenerallyAvailable	<b>作成日</b>	Dec 3, 2024 9:00 AM
<b>更新日</b>	Dec 3, 2024 9:00 AM	<b>モデルの廃止日</b>	Aug 20, 2025 9:00 AM

### エンドポイント

**ターゲット URI**  
https://19233-m58are1c-eastus2.cognitiveservices.azure.com/...

**キー**  
.....

### 監視と安全性

**コンテンツ フィルター**  
DefaultV2

### アプリケーション開発に役立つリンク

- [コード サンプル リポジトリ](#)
- [チュートリアル](#)

## デプロイのヘルプ

### Azure AI Foundry でのデプロイとは?

Azure AI Foundry では、大規模言語モデル (LLM)、フロー、Web アプリのデプロイがサポートされています。LLM またはフローをデプロイすると、それを Web サイトやアプリケーション、またはその他の運用環境で使用できるようになります。

[詳細情報](#)

> [関連リソース](#)



- ホーム
- 概要
- モデルカタログ
- プレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイム オーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
  - Azure OpenAI Service 評価 プレビュー
  - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトル ストア プレビュー

## チャットプレイグラウンド

</> コードの表示 | デプロイ | インポート | エクスポート | プロンプトのサンプル ...

セットアップ 非表示

デプロイ\* + 新しいデプロイの作成

gpt-4o-test (version:2024-11-20)

モデルに指示とコンテキストを与える

情報を見つけるのに役立つ AIアシスタントです。

変更の適用 | プロンプトを生成する

+ セクションの追加

データを追加する

自分のデータについて質問します。データは、指定したデータソースに保存されたままになります。データの保護方法について説明します。

+ データソースの追加

パラメーター

過去のメッセージを含む

チャットの履歴 | 応答形式 テキスト



### サンプルプロンプトを使って始める

- Historical fiction**  
Write a scene set in ancient Rome, focusing on the daily life of a common citizen.
- Character development**  
Describe a day in the life of a superhero who has lost their powers.
- Recipe creation**  
Invent a recipe for a dish that combines flavors from two different cuisines.

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)

21/128000 送信するトークン

## チャットプレイグラウンドのヘルプ

- 関連リソース
  - チャットのクイックスタート

# チャットプレイグラウンド

コードの表示 デプロイ インポート エクスポート プロンプトのサンプル

セットアップ 非表示

デプロイ\* 新しいデプロイの作成

gpt-4o-test (version:2024-11-20)

下の [変更の適用] ボタンを選択してください。

モデルに指示とコンテキストを与える

あなたは小学校の教師です。高学年の小学生の質問にわかりやすく答えてください。

変更の適用

このAIはどのような立場のなのか、どのような状況ののかを書く (コンテキスト)

## サンプルプロンプトを使って始める

- Historical fiction**  
Write a scene set in ancient Rome, focusing on the daily life of a common citizen.
- Character development**  
Describe a day in the life of a superhero who has lost their powers.
- Recipe creation**  
Invent a recipe for a dish that combines flavors from two different cuisines.

データを追加する

自分のデータについて質問します。データは、指定したデータソースに保存されたままになります。データの保護方法について説明します。

データソースの追加

パラメーター

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)

21/128000 送信するトークン

## チャットプレイグラウンドのヘルプ

関連リソース  
チャットのクイックスタート

- ホーム
- 概要
- モデルカタログ
- プレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイムオーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
  - Azure OpenAI Service 評価 プレビュー
  - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトルストア プレビュー

# チャットプレイグラウンド

</> コードの表示 | デプロイ | インポート | エクスポート | プロンプトのサンプル

セットアップ 非表示

デプロイ\* 新しいデプロイの作成

gpt-4o-test (version:2024-11-20)

下の [変更の適用] ボタンを選択してください。

モデルに指示とコンテキストを与える

あなたは小学校の教師です。高学年の小学生の質問にわかりやすく教えてください

変更の適用 | プロンプトを生成する

セクションの追加

データを追加する

自分のデータについて質問します。データは、指定したデータソースに保存されたままになります。データの保護方法について説明します。

データソースの追加

パラメーター

チャットの履歴 | 応答形式 テキスト

**システムメッセージを更新しますか?**

システムメッセージを更新すると、新しいチャットセッションが開始されます。以前のメッセージは新しいAPI要求には含まれません。

今後、このメッセージを表示しない

**続行** | キャンセル

## チャットプレイグラウンドのヘルプ

関連リソース  
チャットのクイックスタート

- ホーム
- 概要
- モデルカタログ
- プレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイムオーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
  - Azure OpenAI Service 評価
  - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトルストア プレビュー

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)

39/128000 送信するトークン

# チャット プレイグラウンド

Azure AI Foundry | Azure OpenAI Service / 19233-m58are1c-eastus2 / チャット プレイグラウンド

すべてのリソース 5

## チャット プレイグラウンド

</> コードの表示 | デプロイ | インポート | エクスポート | プロンプトのサンプル ...

### セットアップ

非表示

デプロイ \* [+ 新しいデプロイの作成](#)

gpt-4o-test (version:2024-11-20)

モデルに指示とコンテキストを与える

あなたは小学校の教師です。高学年の小学生の質問にわかりやすく教えてください。

変更の適用 | プロンプトを生成する

[+ セクションの追加](#)

### チャットの履歴

応答形式 テキスト

## サンプルプロンプトを使って始める

- Historical fiction**  
Write a scene set in ancient Rome, focusing on the daily life of a common citizen.
- Character development**  
Describe a day in the life of a superhero who has lost their powers.
- Recipe creation**  
Invent a recipe for a dish that combines flavors from two different cuisines.

### データを追加する

自分のデータについて質問します。データは、指定したデータソースに保存されたままになります。データの保護方法について説明します。

[+ データソースの追加](#)

### パラメーター

過去のメッセージを含む

大規模言語モデルの仕組みについて教えてください。|

67/128000 送信するトークン

### チャット プレイグラウンドのヘルプ

関連リソース

[チャットのクイックスタート](#)

- ホーム
- 概要
- モデルカタログ
- プレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイムオーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
  - Azure OpenAI Service 評価 プレビュー
  - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトルストア プレビュー

# チャットプレイグラウンド

</> コードの表示 | デプロイ | インポート | エクスポート | プロンプトのサンプル ...

セットアップ 非表示

デプロイ \* 新しいデプロイの作成  
 gpt-4o-test (version:2024-11-20)

モデルに指示とコンテキストを与える

あなたは小学校の教師です。高学年の小学生の質問にわかりやすく教えてください。

変更の適用 | プロンプトを生成する | 戻る

+ セクションの追加

- > データを追加する
- > パラメーター

チャットの履歴 リセット | 応答形式 テキスト

大規模言語モデルの仕組みについて教えてください。

大規模言語モデル (例えばChatGPTなど) は、たくさんの文章を学習して、言葉と言葉のつながりやパターンを理解するように作られたAIの一種です。簡単に言うと、たくさんの本や記事、ウェブページなどを読んで、どんな言葉がどんなふうにつながるのかわかっている「とても大きな辞書みたいなもの」と考えるとわかりやすいかもしれません。

### 仕組みを簡単に説明すると

- 学習する (トレーニング)**  
AIは人間が書いたたくさんの文章を見て、「次にどんな言葉が来そうか」を予測する練習をします。例えば、「今日は空がとても〇〇です」という文章を見たとき、「〇〇」の部分には「青い」や「きれい」といった言葉が来そうだと予測するんです。
- パターンを覚える**  
AIは言葉と言葉の関係や、文の流れをたくさん覚えます。例えば、「犬」という言葉が出てきたら「かわいい」「散歩」「吠える」などの関連する言葉を覚えたりします。
- 質問に答える**  
質問されたり指示を出されたりすると、覚えたパターンを使って「一番自然な答え」を考えます。ただし、AIは実際に考えているわけではなく、あくまで「学んだパターンに基づいて次の言葉を予測している」だけです。

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)

838/128000 送信するトークン

チャットプレイグラウンドのヘルプ

関連リソース

[チャットのクイックスタート](#)

# チャットプレイグラウンド

</> コードの表示 | デプロイ | インポート | エクスポート | プロンプトのサンプル ...

セットアップ 非表示

デプロイ\* 新しいデプロイの作成

gpt-4o-test (version:2024-11-20)

モデルに指示とコンテキストを与える

あなたはAzureの専門家です。初めてAzureを使うユーザーに、わかりやすくAzureの使い方を教えてください。

変更の適用 | プロンプトを生成する

セクションの追加

データを追加する

自分のデータについて質問します。データは、指定したデータソースに保存されたままになります。データの保護方法について説明します。

データソースの追加

パラメーター

チャットの履歴 削除 | 応答形式 テキスト

Azureで仮想マシンを起動して13BのLlama-3モデルをLoRA fine-tuningしたいと思います。リソースは何を選んだらいいですか？

Azureで13BのLlama-3モデルをLoRA (Low-Rank Adaptation) を使ってファインチューニングする場合、以下の手順とリソース選定のポイントを考慮してください。

## 1. 仮想マシンの選定

Llama-3のような大規模な言語モデルを扱う場合、GPU搭載の仮想マシン (VM) が必要です。LoRAは従来のファインチューニングよりも13BモデルはGPUメモリや計算リソースを大量に必要とします。

### 推奨される仮想マシンタイプ

Azureでは、以下のGPU搭載VMシリーズが適しています。

- 1. NDシリーズ (NVIDIA A100搭載)
  - 例: ND96asr\_v4

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)

2104/128000 送信するトークン

## チャットプレイグラウンドのヘルプ

関連リソース  
チャットのクイックスタート

途中までしか生成されない場合は、パラメータを調整

- ホーム
- 概要
- モデルカタログ
- プレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイム オーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
  - Azure OpenAI Service 評価 プレビュー
  - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトルストア プレビュー



- ホーム
- 概要
- モデルカタログ
- プレイグラウンド
  - チャット
  - アシスタント プレビュー
  - リアルタイム オーディオ プレビュー
  - イメージ
  - 入力候補
- ツール
  - 微調整
    - Azure OpenAI Service 評価 プレビュー
    - 格納された入力候補 プレビュー
  - バッチ ジョブ
- 共有リソース
  - デプロイ
  - クォータ
  - 安全性とセキュリティ
  - データ ファイル
  - ベクトル ストア プレビュー

## チャットプレイグラウンド

</> コードの表示    デプロイ    インポート    エクスポート    プロンプトのサンプル

自分のデータについて質問します。データは、指定したデータソースに保存されたままになります。データの保護方法について説明します。

+ データソースの追加

パラメーター

過去のメッセージを含む ①

最大応答 ①

温度 ①

上位 P ①

シーケンスの停止 ①

頻度のペナルティ ①

プレゼンスのペナルティ ①

### チャットの履歴

応答形式 テキスト

Azureで仮想マシンを起動して13BのLlama-3モデルをLoRA fine-tuningしたいと思います。リソースは何を選んだらいいですか？

**Azureで13BのLlama-3モデルをLoRA (Low-Rank Adaptation) を使ってファインチューニングする場合、以下の手順とリソース選定のポイントを考慮してください。**

#### 1. 仮想マシンの選定

Llama-3のような大規模な言語モデルを扱う場合、GPUを搭載した仮想マシン (VM) が必要です。LoRAは従来のファインチューニングより軽量ですが、それでも13BモデルはGPUメモリや計算リソースを大量に使用します。

#### 推奨される仮想マシンタイプ

Azureでは、以下のGPU搭載VMシリーズが適しています：

- NDシリーズ (NVIDIA A100搭載)
  - 例: ND96asr\_v4

ここにユーザー クエリを入力します。(Shift + Enter で新しい行)

2104/128000 送信するトークン

### チャットプレイグラウンドのヘルプ

関連リソース

チャットのクイックスタート

# 4. 外部のプログラムから 呼び出す

Google Colaboratoryのノートブックから、APIを介してOpenAI Service  
のモデルにプロンプトを送り、結果を受け取る



### ← gpt-4o-test

詳細 メトリック リスクと安全性

プレイグラウンドで開く 編集 削除

配置情報	
名前	プロビジョニングの状態 成功
gpt-4o-test	
展開の種類	作成日
グローバル標準	2025-01-30T08:19:22.6269308Z
作成者	変更日
1923306880@utac.u-tokyo.ac.jp	Jan 30, 2025 5:19 PM
更新者	バージョン更新ポリシー
1923306880@utac.u-tokyo.ac.jp	新しい既定のバージョンが使用可能になったら
レート制限 (1分あたりのトークン数)	レート制限 (1分あたりの要求数)
10,000	60
モデル名	モデルバージョン
gpt-4o	2024-11-20
ライフサイクルの状態	作成日

エンドポイント

ターゲット URI  
https://19233-m58are1c-eastus2.cognitiveservices.azure.com/...

キー  
.....

でコピーしてエンドポイントにセット

フロー、Web アプリのデプロイがサポートされています。LLM またはフローをデプロイすると、それを Web サイトやアプリケーション、またはその他の運用環境で使用できるようになります。

をクリックして api キーを表示させよう。をクリックしてコピー

- ホーム
- 概要
- モデル カタログ
- プレイグラウンド
- チャット
- アシスタント プレビュー
- リアルタイム オーディオ プレビュー
- イメージ
- 入力候補
- ツール
- 微調整
- Azure
- OpenAI プレビュー
- Service 評価
- 格納された入力候補 プレビュー
- バッチ ジョブ
- デプロイ
- 安全性とセキュリティ
- データ ファイル
- ベクトル ストア プレビュー

**注意：** エンドポイントの情報が知られてしまうと、誰でもあなたのリソースを使ってしまう（支払い義務はあなたに発生）ので、絶対に公開しないこと！

# Google Colaboratory サンプルコード

- ECCSアカウントにログインして[こちら](#)をColabで開いてください。
- 保存したい場合は「ドライブにコピー」をクリックして、ご自分のGoogleドライブにコピーしてください。
- 「2. パラメータの設定」にある以下の部分を書き換えてください。

```
[ ] import os
```

```
# 以下の[YOUR TARGET URL]をエンドポイントのターゲットURLに書き換える  
os.environ["AZURE_OPENAI_ENDPOINT"] = "[YOUR TARGET URL]"
```

```
# 以下の[YOUR KEY]をエンドポイントのキーに書き換える  
os.environ["AZURE_OPENAI_KEY"] = "[YOUR KEY]"
```

ターゲットURLに書き換え

キーに書き換え

```
▶ from openai import AzureOpenAI

client = AzureOpenAI(
    azure_endpoint = os.getenv("AZURE_OPENAI_ENDPOINT"),
    api_key = os.getenv("AZURE_OPENAI_KEY"),
    api_version="2024-08-01-preview",
)

response = client.chat.completions.create(
    model=os.getenv("AZURE_OPENAI_MODEL"),
    messages=[
        {"role": "system", "content": "あなたはAzureの専門家です。初めてAzureを使うユーザに、わかりやすくAz"},
        {"role": "user", "content": "Azure OpenAI Serviceでデプロイしたgpt-4oのモデルにGoogle Colab上で動か"},
    ],
    max_tokens=800,
    temperature=0.7,
    top_p=0.95,
    frequency_penalty=0,
    presence_penalty=0,
    stop=None,
    stream=False
)
```

コンテキストとユーザクエリを設定。

{"role": "system", "content": "あなたはAzureの専門家です。初めてAzureを使うユーザに、わかりやすくAz"},  
{"role": "user", "content": "Azure OpenAI Serviceでデプロイしたgpt-4oのモデルにGoogle Colab上で動か"}

```
print(response.choices[0].message.content)
```

⇒ Azure OpenAI ServiceでデプロイしたGPT-4 (gpt-4o) モデルにGoogle Colabからアクセスするには、以下の

---

### ### 1. \*\*Azure OpenAI Serviceの準備\*\*

まず、Azure上でOpenAI Serviceを設定し、GPT-4のモデルをデプロイします。

#### #### 手順

##### 1. \*\*Azureポータルにログイン\*\*:

- [Azureポータル](<https://portal.azure.com>)にアクセスし、アカウントでログインします。

##### 2. \*\*Azure OpenAIリソースを作成\*\*:

- Azureポータルの検索バーで「OpenAI」を検索し、「Azure OpenAI」を選択します。
- 「+ 作成」をクリックしてリソースを作成します。
- サービス名やリージョンを選択し、リソースをデプロイします。

##### 3. \*\*モデルをデプロイ\*\*:

- 作成したAzure OpenAIリソースに移動し、「モデルデプロイ」を選択。
- 使用したいモデル (例: `gpt-4` または `gpt-4o`) を選択してデプロイします。
- デプロイ名 (例: `gpt-4o-model`) を指定してデプロイを完了させます。

モデルが返してきたメッセージ  
を書き出すコード。

## 5. その他の話題

# 5. 1 OpenAI Serviceの価格

料金の見積もり

# Azure OpenAI Serviceの価格 (こちらを参照)

Home / [Azure pricing](#) / Azure OpenAI Service pricing

## Azure OpenAI Service pricing

[Request a pricing quote](#) [Try Azure for free](#)

[AI Services Family](#) [Overview](#) [Pricing table](#) [Purchase options](#) [Resources](#) [FAQ](#) [Azure OpenAI Service](#)

### Azure OpenAI Service pricing overview

Azure OpenAI Service delivers enterprise-ready generative AI featuring powerful models from OpenAI, enabling organizations to innovate with text, audio, and vision capabilities. Beyond the cutting-edge models, companies choose Azure OpenAI Service for built-in data privacy, regional/area/global flexibility, and seamless integration into the Azure ecosystem including Fabric, Cosmos DB and Azure AI Search. Companies of all sizes can confidently scale AI solutions to enhance customer experience, automate workflows, and unlock creative potential, driving measurable impact and competitive differentiation.

To help customers in the journey, we offer pricing and cost management solutions to meet your needs. including:

- **Standard (On-Demand):** Pay-as-you-go for input and output tokens.
- **Provisioned (PTUs):** [Allocate throughput](#) with predictable costs, with monthly and annual reservations available to reduce overall spend.
- **Batch API:** Language models are also now available in the [Batch API for global deployments and three regions](#), that returns completions within 24 hours for a 50% discount on Global Standard Pricing.

[Chat with Sales](#)

# OpenAI GPTの料金

## GPT-4o

GPT-4o is the most advanced multimodal model that's faster and cheaper than GPT-4 Turbo with stronger vision capabilities. The model has 128K context and an October 2023 knowledge cutoff.

Model	Pricing (1M Tokens)	Pricing with Batch API (1M Tokens)
GPT-4o-2024-1120 Global	Input: <b>¥393.88751</b> Cached Input: <b>¥196.9438</b> Output: <b>¥1,575.5501</b>	Input: <b>¥196.94376</b> Output: <b>¥787.7751</b>

## o1

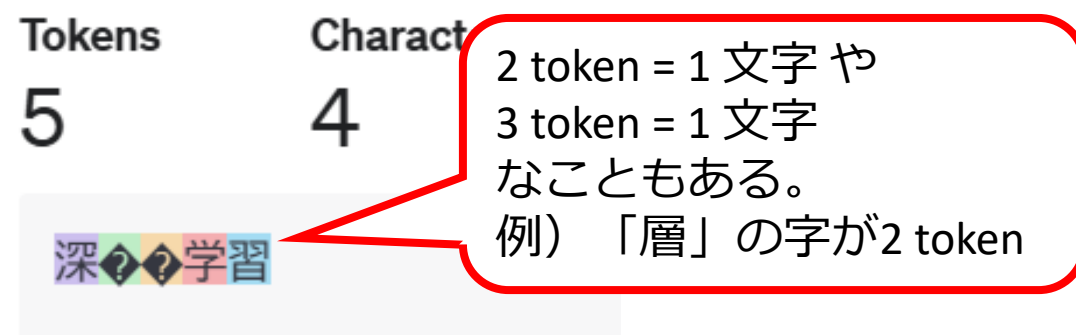
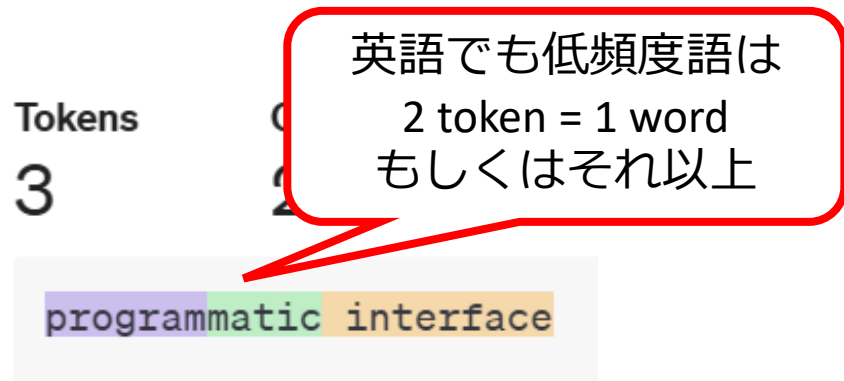
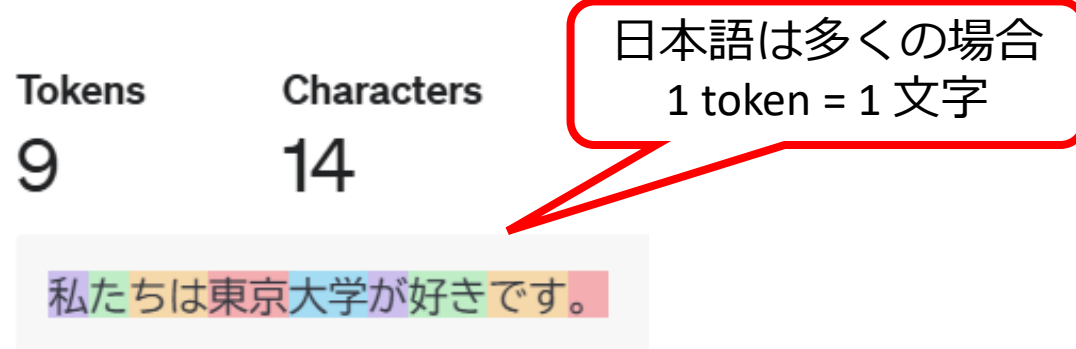
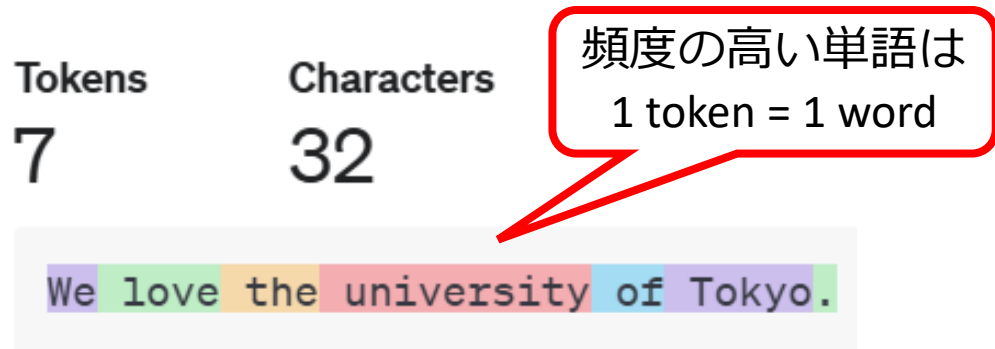
o1 is the new reasoning model series for complex tasks. The model has 200K context and an October 2023 knowledge cutoff.

Model	Pricing (1M Tokens)
o1 1217 Global	Input: <b>¥2,363.3251</b> Cached Input: <b>¥1,181.6626</b> Output: <b>¥9,453.300060</b>



# Azure OpenAI Serviceの価格

- 入力文 (Input) と出力文 (Output) のtoken数によって決まる
- OpenAIにおけるtoken数の調べ方：[Web版tokenizer](#)
- 英語は100 tokens  $\approx$  75 words だが、低頻度語は対応するtoken数が増える
- 日本語は全言語から見れば低頻度  $\rightarrow$  英語よりtoken数が圧倒的に多い



「デプロイ>[名前]>メトリック」で  
トークン数を確認できる。

## gpt-4o-test

メトリック

Azure Monitor を表示 01/24/2025 - 01/31/2025 最終日 7日 1か月 ...



### デプロイのヘルプ

Azure AI Foundry でのデプロイとは?

Azure AI Foundry では、大規模言語モデル (LLM)、フロー、Web アプリのデプロイがサポートされています。LLM またはフローをデプロイすると、それを Web サイトやアプリケーション、またはその他の運用環境で使用できるようになります。

[詳細情報](#)

> 関連リソース

クリップボード (24/24)  
アイテムは収集されません。クリップボードの容量以上のサイズです。

デプロイ

## 5.2 1分当たりのトークン制限

Azure OpenAI Serviceにはクォータ制限がある

### クォータの使用状況の監視および追跡

サブスクリプションとリージョン別にクォータを表示し、デプロイ全体の使用状況を追跡します。デプロイを作成するにはクォータが必要であり、トラフィックのニーズに応じて柔軟にサイズを変更できます。

#### サブスクリプション

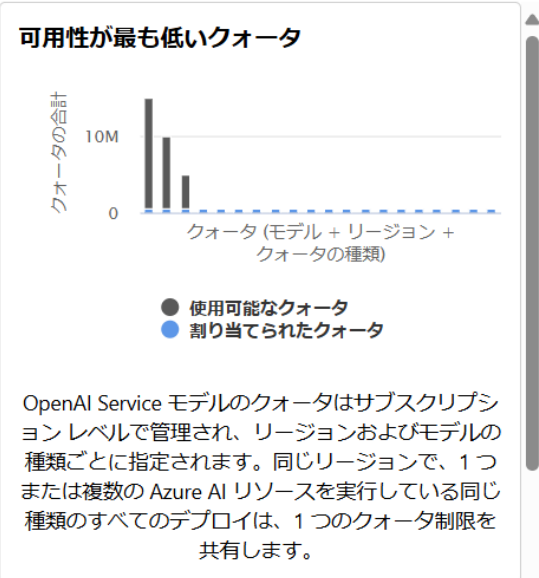
1923306880-YamakataLab

Azure OpenAI Service Standard + バッチ Azure OpenAI Service がプロビジョニングされました

クォータの要求 最新の情報に更新 ビューのリセット

グループ化 クォータの種類、モデルとリージョン

デプロイ	モデル名	バージョン	クォータ...	リージョン	リソース	クォ...
> GlobalStandard						



#### リージョンとクォータの種類全体におけるモデルの可用性

- ホーム
- 概要
- モデル カタログ
- プレイグラウンド
- チャット
- アシスタント プレビュー
- リアルタイムオーディオ プレビュー
- イメージ
- 入力候補
- ツール
- 微調整
- Azure OpenAI Service 評価 プレビュー
- 格納された入力候補 プレビュー
- バッチ ジョブ
- 共有リソース
- クォータ**
- 安全性とセキュリティ
- データ ファイル
- ベクトルストア プレビュー

## クォータの使用状況の監視および追跡

サブスクリプションとリージョン別にクォータを表示し、デプロイ全体の使用状況を追跡します。デプロイを作成するにはクォータが必要であり、トラフィックのニーズに応じて柔軟にサイズを変更できます。

### サブスクリプション

1923306880-YamakataLab

Azure OpenAI Service Standard + バッチ Azure OpenAI Service がプロビジョニング

クォータの要求 最新の情報に更新 ビューのリセット

デプロイ	モデル名	バージョン	クォータ
GlobalStandard			
gpt-4o			
East US 2	gpt-4o		20K
gpt-4o		2024-11-20	GlobalStandard East US 2 aifoundrytest0... 10KTPM
gpt-4o-test	gpt-4o	2024-11-20	GlobalStandard East US 2 19233-m58a... 10KTPM
o1			
East US 2	o1		GlobalStandard 10K の 5M TPM

- Kは1000、TPMはTokens-per-Minuteで、1分当たりのトークン数
- 初期状態では1分に10000トークンの制限がかかっている
- モデル、リージョンによってクォータは異なる (参照)

クォータ (モデル + リージョン + クォータの種類)

● 使用可能なクォータ  
● 割り当てられたクォータ

OpenAI Service モデルのクォータはサブスクリプションレベルで管理され、リージョンおよびモデルごとに指定されます。同じリージョンで、1つの複数の Azure AI リソースを実行している同じ種別のすべてのデプロイは、1つのクォータ制限を共有します。

リージョンとクォータの種類全体における  
の可用性

### クォータの使用状況の監視および追跡

サブスクリプションとリージョン別にクォータを表示し、デプロイ全体の使用状況を追跡します。デプロイを作成するにはクォータが必要であり、トラフィックのニーズに応じて柔軟にサイズを変更できます。

#### サブスクリプション

1923306880-YamakataLab

Azure OpenAI Service Standard + バッチ Azure OpenAI Service がプロビジョニングされました

#### クォータの要求

**このデプロイに割り当てられたクォータの編集**

容量 Tokens Per Minute (thousands) - gpt-4o - GlobalStandard 1120K/29990K

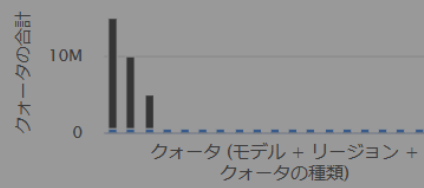
Slider: 0 to 29990K (current value: 1120K)

確認 キャンセル

スライダーで希望するサイズまで引き上げ

#### グループ化 クォータの種類、モデルとリージョン

#### 可用性が最も低いクォータ



OpenAI Service モデルのクォータはサブスクリプションレベルで管理され、リージョンおよびモデルごとに指定されます。同じリージョンで、1つのクォータを複数の Azure AI リソースを実行している同じ種類のすべてのデプロイは、1つのクォータ制限を共有します。

#### リージョンとクォータの種類全体における可用性

デプロイ	モデル	作成日時	クォータの種類	リージョン	クォータ ID	制限	操作
gpt-4o		2024-11-20	GlobalStandard	East US 2	aifoundrytest0...	10KTPM	✎
gpt-4o-test	gpt-4o	2024-11-20	GlobalStandard	East US 2	19233-m58ar...	10KTPM	✎
o1			GlobalStandard				
East US 2	o1		GlobalStandard			10K の 5M TPM	

## クォータの使用状況の監視および追跡

サブスクリプションとリージョン別にクォータを表示し、デプロイ全体の使用状況を追跡します。デプロイを作成するにはクォータが必要であり、トラフィックのニーズに応じて柔軟にサイズを変更できます。

### サブスクリプション

1923306880-YamakataLab

Azure OpenAI Service Standard + バッチ Azure OpenAI Service がプロビジョニングされました

クォータの要求

最新の情報に更新

ビュー

検索

デプロイ ↑

モデル名

GlobalStandard

gpt-4o

GlobalStandard

East US 2

gpt-4o

GlobalStandard

1.13M の 30M TPM

gpt-4o

2024-11-20

GlobalStandard

10KTPM

gpt-4o-test

gpt-4o

2024-11-20

GlobalStandard

East US 2

1.12MTPM

> o1

GlobalStandard

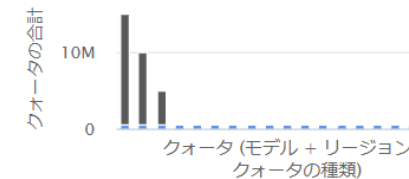
- 初期状態では30M TPMが上限
- もっと増やしたいときは クォータの引き上げを要求 (承認されるまで数日かかる！)

増えた！

1.12MTPM

グループ化 クォータの種類、モデルとリ

### 可用性が最も低いクォータ



OpenAI Service モデルのクォータはサブスクリプションレベルで管理され、リージョンおよびモデルごとに指定されます。同じリージョンで、1複数の Azure AI リソースを実行している同じすべてのデプロイは、1つのクォータ制限を共有します。

リージョンとクォータの種類全体における可用性



# クォータの 引き上げを要求

- クォータ増加リクエスト  
フォーム
- 氏名やメールアドレス、組  
織名、要求タイプ・サイズ
- なぜクォータ引き上げをし  
たいのかを説明  
(Justification)
- 人力で判定している？  
承認まで数日かかる  
(早めにやっておく)

The screenshot shows a web browser window displaying the Microsoft Azure OpenAI Service Request for Quota Increase form. The page has a dark header with the Microsoft Azure logo and the title "Azure OpenAI Service: Request for Quota Increase". Below the title, there is a paragraph of text explaining the form's purpose and that requests are processed in order of receipt, with priority given to high-traffic users. A section marked with an asterisk as "必須" (required) contains instructions to read carefully and answer completely before submitting. It also includes a "Please Note" section advising that the form is for Microsoft personnel and a link for more information. The first form field is labeled "1. First Name \*" and has a placeholder text "Enter your first name."

Microsoft Azure

## Azure OpenAI Service: Request for Quota Increase

This form is used to submit requests for quota increases in quota for the standard deployment type. Quota increase requests are being accepted and will be filled in the order they are received. Priority will be given to customers who generate traffic that consumes the existing quota allocation, and your request may be denied if this condition is not met.

\* 必須

Read the instructions carefully and answer each question completely before submitting the request

Use this form to request an increase due to your forecasted usage for Azure OpenAI Service. Microsoft will use the information you provide to assess your usage volume and patterns, allowing us to allocate the necessary GPU capacity to support your work. We will make every effort to accommodate your request; however, allocation is based on our current capacity and future deployments, and is subject to availability.

**Please Note:** For [Microsoft personnel](#), do not fill out this form. Find more information [here](#).

1. First Name \*

Enter your first name.



## 5.3 Azure OpenAI Service のデータ、 プライバシー、セキュリティ

データはAzureやOpenAIに使われるのか？

# Data, privacy, and security for Azure OpenAI Service - Azure AI services | Microsoft Learn

## **プロンプト (入力) と完成品 (出力)、埋め込み、トレーニング データ:**

- 他のお客様はご利用いただけません。
- OpenAIには利用できません。
- OpenAIモデルの改善には使用されません。
- はAzure OpenAI Service 基盤モデルのトレーニング、再トレーニング、または改善には使用されません。
- お客様の許可または指示なしに、Microsoft または第三者の製品またはサービスを改善するために使用することはありません。
- 微調整された Azure OpenAI モデルは、お客様専用に使使用できます。

Azure OpenAI Service は、Microsoft が Azure サービスとして運営しています。Microsoft は Microsoft の Azure 環境で OpenAI モデルをホストしており、本サービスは OpenAI が運営するサービス (ChatGPT や OpenAI API など) と対話しません。

# 目的

- 学習済みモデルを利用する2つの選択肢
  - OpenAIのモデルをすぐに使いたい → Azure OpenAI Service
  - OpenAIを含め様々なモデルを使いたい → Azure AI Foundry
- Azure AI Foundryのロードマップ
  1. Hubを作成してからProjectを作成
  2. モデルカタログからモデルを選択（たとえば"deepseek-R1"を選ぶ）
  3. チャットプレイグラウンドからチャットを実行

# DeepSeekを使おう

OpenAI以外のモデルを使いたいならAI Foundryを利用

# 2025年2月現在のモデル一覧

最新版は[こちら](#)を参照

会社	モデル	タイプ
AI21 Labs	AI21-Jamba-1.5-Mini, AI21-Jamba-1.5-Large	chat-completion
Azure OpenAI	o3-mini, o1, o1-preview, o1-mini, gpt-4o, gpt-4o-mini	chat-completion
Azure OpenAI	gpt-4o-realtime-preview	real-time
Azure OpenAI	text-embedding-3-large, text-embedding-3-small	embeddings
Cohere	Cohere-embed-v3-english, Cohere-embed-v3-multilingual	embeddings
Cohere	Cohere-command-r-plus-08-2024, Cohere-command-r-08-2024, Cohere-command-r-plus, Cohere-command-r	chat-completion
Core42	jais-30b-chat	chat-completion
DeepSeek	DeepSeek-R1	chat-completion
Meta	Llama-3.3-70B-Instruct, Llama-3.2-11B-Vision-Instruct, Llama-3.2-90B-Vision-Instruct, Meta-Llama-3.1-405B-Instruct, Meta-Llama-3.1-8B-Instruct, Meta-Llama-3.1-70B-Instruct, Meta-Llama-3-8B-Instruct, Meta-Llama-3-70B-Instruct	chat-completion
Microsoft	Phi-3-mini-128k-instruct, Phi-3-mini-4k-instruct, Phi-3-small-8k-instruct, Phi-3-medium-128k-instruct, Phi-3-medium-4k-instruct, Phi-3.5-vision-instruct, Phi-3.5-MoE-instruct, Phi-3-small-128k-instruct, Phi-3.5-mini-instruct, Phi-4	chat-completion
Mistral AI	Ministral-3B, Mistral-large, Mistral-small, Mistral-Nemo, Mistral-large-2407, Mistral-Large-2411, Codestral-2501	chat-completion
NTT Data	Tsuzumi-7b	chat-completion

[オープンソース版](#)

# 1. HubとProjectを作成

### Azure サービス

+ リソースの作成

[Icon] リソース グループ

### リソース

最近 お気に入り

名前
AI-Foundry-test0
AI-Foundry-hub0
1923306880-3791_ai
rg-1923306880-1245_ai
deepseek0
openai_test1
19233-m58are1c-eastus2
1923306880-0528
deepseek00567883523
1923306880-YamakataLab
gpt-4o-test-1
ai-19233068803791ai761221281178

すべて表示

### 移動

検索: AI Found

すべて サービス (13)

サービス

- Azure AI Foundry
- AI Search
- Community Training
- Azure OpenAI

Marketplace

- Azure AI Foundry

ドキュメント

- Azure AI Foundry アーキテクチャ - Azure AI Foundry
- Azure AI Foundry SDK を使い始める方法 - Azure AI Foundry
- Azure AI Foundry ポータルでモデル カタログについて確認する - Azure AI Foundry
- Azure AI サービスのドキュメント

Microsoft Entra ID

- AI-Foundry-test0 サービス プランシブル
- ai-foundry-project0 サービス プランシブル

Microsoft Entra ID で検索を続行してください

すべてのサブスクリプションを検索しています。 フィードバックの送信

Azure AI Foundry  
を選択

クォータ

Azure Machine...

その他のサービス

### 最終表示日

10 時間前
10 時間前
11 時間前
11 時間前
11 時間前
11 時間前
11 時間前
11 時間前
11 時間前
22 時間前
3 日前
4 日前

まずHubを作る  
(Projectから始めても、既存のHubがなければ作ることになる)

+ Create ビューの管理 更新 CSV

**Project**  
Collaborate, organize, and track work to build AI apps.

**Hub**  
Grouping container for projects. Provides security, connectivity, and compute management.

	リソース グループ ↑↓	種類 ↑↓	場所 ↑↓	サブスクリプション ↑↓	
	openai_test1	Azure AI services	East US 2	1923306880-YamakataLab	...
<input type="checkbox"/>	rg-1923306880-1245_ai	Azure AI hub	East US 2	1923306880-YamakataLab	...
<input type="checkbox"/>	ai-foundry-project0	AI-Foundry-hub0	East US 2	1923306880-YamakataLab	...
<input type="checkbox"/>	AI-Foundry-test0	AI-Foundry-hub0	East US 2	1923306880-YamakataLab	...
<input type="checkbox"/>	aifoundrytest00903255598	AI-Foundry-hub0	East US 2	1923306880-YamakataLab	...



ホーム > Azure AI Foundry >

# Azure AI ハブ

Azure AI ハブ リソースの作成

基本 | ストレージ | ネットワーク | 暗号化 | ID | タグ | 確認および作成

**組織**  
デプロイされたリソースとコストを管理するサブスクリプションを選択してください。フォルダーなどのリソース グループを使用してすべてのリソースを整理し管理します。AI ハブは、チームがプロジェクト作業、モデル エンドポイント、コンピューティング、(データ) 接続、セキュリティ設定を共有し、使用状況を管理するためのコラボレーション環境です。

サブスクリプション \* ①

リソース グループ \* ①   
[新規作成](#)

リージョン \* ①

**リソースの詳細**

名前 \* ①  ✓

フレンドリ名 ①

既定のプロジェクト リソース グループ ①

Same as hub resource group

**Azure AI サービスのベース モデル**

OpenAI を含む AI サービスに接続する \* ①   
[新規作成](#)

インターフェースは  
Azure OpenAI Service と  
概ね同じ

**確認および作成**

< 前へ | 次へ: ストレージ

ホーム > Azure AI Foundry >

# Azure AI ハブ

Azure AI ハブ リソースの作成

✔ 検証に成功しました

基本 | ストレージ | ネットワーク | 暗号化 | ID | タグ | 確認および作成

## 基本

サブスクリプション: 1923306880-YamakataLab  
リソース グループ: (新規) AI-Foundry-hub0  
名前: AI-Foundry-test0

## リソース

リージョン: East US 2  
AI サービス: (新規) aifoundrytest00903255598  
ストレージ アカウント: (新規) aifoundrytest02974364556  
キー コンテナー: (新規) aifoundrytest02322244864  
Application Insights: なし  
コンテナー レジストリ: なし

## ネットワーク

接続方法: すべてのネットワークからのパブリック アクセスを有効にする  
ネットワークの分離: パブリック

## 暗号化

暗号化の種類: Microsoft マネージド キー

## ID

ID の種類: システム割り当て済み

**作成**

< 前へ

次へ >

Automation のテンプレートをダウンロードする

ホーム >

# Microsoft.MachineLearningServices | 概要

検索 削除 キャンセル 再デプロイ ダウンロード 最新の情報に更新

- 概要
- 入力
- 出力
- テンプレート

## デプロイが進行中です

**デプロイ名** : Microsoft.MachineLearningServices  
**サブスクリプション** : 1923306880-YamakataLab  
**リソースグループ** : AI-Foundry-hub0  
**開始日時** : 2025/2/8 23:22:50  
**関連付け ID** : 3df3a6f4-cacf-4654-86c7-36b50a2049c2

### デプロイの詳細

リソース	種類	状態	操作の詳細
aifoundrytest02974364556	ストレージ アカウント	Accepted	<a href="#">操作の詳細</a>
aifoundrytest02322244864	キー コンテナ	OK	<a href="#">操作の詳細</a>



**Microsoft Defender for Cloud**  
 アプリおよびインフラストラクチャをセキュリティで保護します  
[次に移動: Microsoft Defender for Cloud >](#)

**無料の Microsoft チュートリアル**  
[今すぐ学習を開始する >](#)

**エキスパートとの共同作業**  
 Azure のエキスパートは、Azure での資産の管理を支援し、サポートの最前線となるサービスプロバイダー パートナーです。  
[Azure Expert の検索 >](#)

ホーム >

# Microsoft.MachineLearningServices | 概要

デプロイ

検索 削除 キャンセル 再デプロイ ダウンロード 最新の情報に更新

- 概要
- 入力
- 出力
- テンプレート

## ✓ デプロイが完了しました

デプロイ名	: Microsoft.MachineLearningServices	開始日時	: 2025/2/8 23:22:50
サブスクリプション	: 1923306880-YamakataLab	関連付け ID	: 3df3a6f4-cacf-4654-86c7-36b50a2049c2
リソースグループ	: AI-Foundry-hub0		

> デプロイの詳細

∨ 次の手順

**リソースに移動**



### Cost Management

予算内に収め、予期しない料金が請求されないように通知を受け取ります。

[コストのアラートを設定 >](#)



### Microsoft Defender for Cloud

アプリおよびインフラストラクチャをセキュリティで保護します

[次に移動: Microsoft Defender for Cloud >](#)

### 無料の Microsoft チュートリアル

[今すぐ学習を開始する >](#)

### エキスパートとの共同作業

Azure のエキスパートは、Azure での資産の管理を支援し、サポートの最前線となるサービスプロバイダー パートナーです。

[Azure Expert の検索 >](#)

ホーム > Microsoft.MachineLearningServices | 概要 >

# AI-Foundry-test0

Azure AI hub


検索 + Create project ↓ Download config.json Delete

- 概要
- アクティビティ ログ
- アクセス制御 (IAM)
- タグ
- 問題の診断と解決
- イベント
- > 設定
- > 監視
- > オートメーション
- > サポート + トラブルシューティング

要点 JSON ビュー

リソース グループ	: <a href="#">AI-Foundry-hub0</a>	Project resource group (...)	: <a href="#">AI-Foundry-hub0</a>
場所	: East US 2	Storage	: <a href="#">aifoundrytest02974364556</a>
サブスクリプション	: <a href="#">1923306880-YamakataLab</a>	Container Registry (edit)	: ...
サブスクリプション ID	: a3b0857d-8c96-43a1-b2ee-d787b83bbe5e	Application Insights (edit)	: ...
Key Vault	: <a href="#">aifoundrytest02322244864</a>	Provisioning State	: Succeeded

### Govern the environment for your team in AI Foundry

 Your Azure AI hub provides enterprise-grade security, and a collaborative environment to build AI solutions. Centrally audit usage and cost, and set up connections to your company resources that all projects can use. [learn more about the Azure AI Foundry](#)

[Launch Azure AI Foundry](#)

- 管理センター
- All hubs + projects
- Quota
- ハブ (AI-Foundry-test0)
- Overview**
- Users
- Models + endpoints
- Connected resources
- Compute

## AI Foundry test0

**+ 新しいプロジェクト** 最新の情報に更新 ...



### 最初のプロジェクトを作成する

プロジェクトを使用すると、AI 資産を整理し、さまざまなツール、データ接続などにまたがる "状態" (システムが以前のイベントや相互作用を記憶している状態) を保存できます。

[Azure AI Foundry 内のプロジェクトの詳細を確認する](#)

### 説明

ハブの説明をここに追加してください

### ハブのプロパティ

名前	場所
AI-Foundry-test0	eastus2
サブスクリプション	リソースグループ
1923306880-YamakataLab	AI-Foundry-hub0

[Azure Portal で管理する](#)

[API エンドポイントとキーを取得する](#)

[サブスクリプション クォータを表示する](#)

**ハブの削除**

### ユーザー 1

[すべて表示](#) →

AI-Foundry-test0 Azure AI Administrator

+ 新しいユーザー

## ハブの管理

### Overview

You can use the management center to create and configure hubs and projects within those hubs.

[Relevant resources](#)

**入門チュートリアルの起動**

- 管理センター
  - All hubs + projects
  - Quota
- ハブ (AI-Foundry-test0)
  - Overview**
  - Users
  - Models + endpoints
  - Connected resources
  - Compute

## AI Foundry test0

[+ 新しいプロジェクト](#) 最新の情報に更新 ...

説明

ハブの説明をここに追加してください

### ハブのプロパティ

名前	場所
is2	
グループ	
foundry-	

ユーザー 1 [すべて表示](#) →

AI-Foundry-test0 Azure AI Administrator

[+ 新しいユーザー](#)

## ハブの管理

- Overview
  - You can use the management center to create and configure hubs and projects within those hubs.
- Relevant resources

[入門チュートリアルの起動](#)

### プロジェクトの作成

プロジェクトは、作業のための管理が簡単なコンテナであり、コラボレーション、組織、そしてデータやその他のサービスとの接続のための鍵となります。

**現在のハブ**

AI-Foundry-test0

**プロジェクト名** \*

[作成](#) [キャンセル](#)

最初の  
プロジェクトを使用すると、  
などにまたがる "状態" (シス  
態) を保存できます。  
[Azure AI Foundry 内のプロジェクトの詳細を確認する](#)

Azure OpenAI Service  
と概ね同じ

- 概要
- モデル カタログ
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプトフロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
- モデル + エンドポイント
- データとインデックス
- Web Apps

## AI-Foundry-Project0

プロジェクトの説明を追加してください (省略可能)

### エンドポイントとキー すべてのエンドポイントを表示する

#### API キー

.....

#### 含まれる機能

Azure AI 推論

Azure OpenAI Service

Azure AI サービス

次のエンドポイントを使用して、Azure OpenAI Service モデルを呼び出します:

#### Azure OpenAI Service エンドポイント

https://aifoundrytest00903255598.openai.azure.

{x} API ドキュメント

### プロジェクトの詳細

#### プロジェクト接続文字列

eastus2.api.azureml.ms;a3b0857d-8c96-43a...

#### サブスクリプション

1923306880-YamakataLab

#### サブスクリプション ID

a3b0857d-8c96-43a1-b2ee-d787b83bbe5e

#### 場所

eastus2

### プロジェクト設定の管理

- ユーザーの追加
- クォータの表示
- リソースの接続
- コストの追跡

管理センターで開く

## プロジェクトのヘルプ

### プロジェクトとは

プロジェクトは、AI のカスタマイズとオーケストレーションのためのツールを備えた組織のコンテナです。プロジェクトでは、作業の整理、プロンプトフローなどのさまざまなツールでの状態の保存、他のユーザーとの共同作業を行うことができます。

[プロジェクトで作業を整理する方法の詳細](#)

関連リソース

その他のチュートリアル

入門チュートリアルの起動

## 次の手順で基本を抑える



ステージ1

定義と探索

適切なモデルを選択する



ステージ2

ビルドとカスタマイズ



ステージ3

評価と改善

トレースの作成

管理センター



## 2. モデルカタログから モデルを選択

- 概要
- モデル カタログ**
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプトフロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
  - モデル + エンドポイント
  - データとインデックス
  - Web Apps

## AI-Foundry-Project0

プロジェクトの説明を追加してください(省略可能)

### エンドポイントとキー

[すべてのエンドポイントを表示する](#)

#### API キー

.....

#### 含まれる機能

Azure AI 推論

Azure OpenAI Service

Azure AI サービス

次のエンドポイントを使用して、Azure OpenAI Service モデルを呼び出します:

#### Azure OpenAI Service エンドポイント

https://aifoundrytest00903255598.openai.azure.

{x} API ドキュメント

### プロジェクトの詳細

#### プロジェクト接続文字列

eastus2.api.azureml.ms;a3b0857d-8c96-43a...

#### サブスクリプション

1923306880-YamakataLab

#### サブスクリプション ID

a3b0857d-8c96-43a1-b2ee-d787b83bbe5e

#### 場所

eastus2

### プロジェクト設定の管理

- ユーザーの追加
- クォータの表示
- リソースの接続
- コストの追跡

[管理センターで開く](#)

## プロジェクトのヘルプ

### プロジェクトとは

プロジェクトは、AI のカスタマイズとオーケストレーションのためのツールを備えた組織のコンテナです。プロジェクトでは、作業の整理、プロンプトフローなどのさまざまなツールでの状態の保存、他のユーザーとの共同作業を行うことができます。

[プロジェクトで作業を整理する方法の詳細](#)

[関連リソース](#)

[その他のチュートリアル](#)

[入門チュートリアルの起動](#)

### 次の手順で基本を抑える



ステージ1

定義と探索

適切なモデルを選択する



ステージ2

ビルドとカスタマイズ



ステージ3

評価と改善

トレースの作成

[管理センター](#)

## カスタム AI ソリューションを構築するための適切なモデルを見つける

### 新着情報

**4o Mini Audio is here!**

Advanced audio at a fraction of the cost

[モデルのチェックアウト](#)

[ブログを読む](#)

**2411/NeMo/3B で使用できるミストラル 微調整**

カスタマイズにより、Mistral AI モデルのパフォーマンスを向上させます。

[Check out models](#) [ブログを読む](#)

**o3-mini はここにあります!**

o3-mini には、高いパフォーマンスを必要とするシナリオに適したコスト効率の高い o1 機能が...

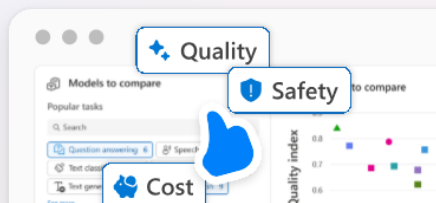
[モデルのチェックアウト](#)

[ブログを読む](#)

**New model benchmarks available now in model catalog**

Model benchmarks are integrated into model catalog for easier navigation. Compare benchmarks across models and datasets available in the industry to assess which one...

[Compare with benchmarks](#) [How model benchmarks are scored](#)



コレクション | 業界 | 機能 | デプロイ オプション | 推論タスク

タスクの微調整 | ライセンス

[モデルの比較](#)

検索 モデル 1831

**o3-mini**  
チャットの完了

**DeepSeek-R1**  
チャットの完了

**Phi-4**  
チャットの完了

**gpt-4o-mini-audio-preview**  
オーディオの生成

**gpt-4o-mini-realtime-preview**  
オーディオの生成

**o1**  
チャットの完了

### モデルのヘルプ

**How to use the model catalog**

Search by name, filter, or browse to find the right model for your use case. Click to see model details and specifications from the publisher and to deploy a model.

[Relevant resources](#)

Microsoft 以外のモデルには異なる使用条件が適用されることに注意してください。デプロイする前にドキュメントをよくお読みください。

[入門チュートリアル](#)の起動

- 概要
- モデル カタログ**
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプト フロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイ アセット
- モデル + エンドポイント
- データとインデックス
- Web Apps

[管理センター](#)

クリップボード (24/24)  
アイテムは収集されません。クリップボードの容量以上のサイズです。

- 概要
- モデル カタログ**
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプト フロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイ アセット
  - モデル + エンドポイント
  - データとインデックス
  - Web Apps
- 管理センター

## ← DeepSeek-R1 プレビュー

**▶ デプロイ**

**詳細** 既存のデプロイ コード サンプル ライセンス

Learn more: [\[original model announcement\]](#)

DeepSeek-R1 excels at reasoning tasks using a step-by-step training process, such as language, scientific reasoning, and coding tasks. It features 671B total parameters with 37B active parameters, and 128k context length.

DeepSeek-R1 builds on the progress of earlier reasoning-focused models that improved performance by extending Chain-of-Thought (CoT) reasoning. DeepSeek-R1 takes things further by combining reinforcement learning (RL) with fine-tuning on carefully chosen datasets. It evolved from an earlier version, DeepSeek-R1-Zero, which relied solely on RL and showed strong reasoning skills but had issues like hard-to-read outputs and hallucinations. To address these limitations, DeepSeek-R1 incorporates a small amount of cold-start

[詳細表示](#)

### データ、メディア、言語

Property	説明
サポートされるデータ型	入力 text 出力 text
サポートされている言語	en, zh

### 使用状況のガイダンス

#### Content Filtering

Prompts and completions are passed through a default configuration of Azure AI Content Safety classification models to detect and prevent the output of harmful content. Learn more about [Azure AI Content Safety](#). Configuration options for content filtering vary when you deploy a model for production in

## モデルのヘルプ

▼ **モデルデプロイの詳細**

さまざまな種類のモデルのデプロイプロセスの詳細を取得します:

- [Azure OpenAI Service モデルをデプロイする](#)
- [Phi-3 モデルをデプロイする](#)
- [Cohere モデルをデプロイする](#)
- [Meta Llama モデルをデプロイする](#)
- [Mistral モデルをデプロイする](#)
- [JAIS モデルをデプロイする](#)
- [Jamba-Instruct モデルをデプロイする](#)
- [TimeGEN-1 モデルをデプロイする](#)

**入門チュートリアル**の起動

### クイック ファクト

**DeepSeek-R1**

chat-completion

最後のトレーニング **使用できません**

### モデル ID

コードでモデルを配置するときにこのモデルを参照する

```
azureml://registries/azureml-deepseek/models/DeepSeek-R1/versions/1
```

- 概要
- モデル カタログ**
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプトフロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
  - モデル + エンドポイント
  - データとインデックス
  - Web Apps
- 管理センター

## DeepSeek-R1 プレビュー

**デプロイ**

**詳細** 既存のデプロイ コード

Learn more: [\[original model announcement\]](#)

DeepSeek-R1 excels at reasoning, and coding tasks. It features a long context length.

DeepSeek-R1 builds on the previous generation by extending Chain-of-Thought (CoT) reasoning with Reinforcement Learning (RL) with fine-tuning on a large dataset, which relied solely on RL at inference time.

**データ、メディア、**

**Property**

サポートされるデータ型

サポートされている言語

**使用状況のガイド**

### Content Filtering

Prompts and completions are passed through a default configuration of Azure AI Content Safety classification models to detect and prevent the output of harmful content. Learn more about [Azure AI Content Safety](#). Configuration options for content filtering vary when you deploy a model for production in

### DeepSeek-R1 用のサーバーレス API デプロイ

**概要** 価格と使用条件

 DeepSeek-R1 は、ファーストパーティの従量課金サービスとして、Microsoft により提供されています。  
[サービスとしてのモデルの詳細について説明します。](#)

**現在のプロジェクト リソース**  
AI-Foundry-Project0

① サーバーレス API は以下のリージョンでご利用いただけます: [eastus2, westus3, northcentralus, eastus, southcentralus, westus]。

**デプロイ名 \***  
DeepSeek-R1-vlxjf

**コンテンツ フィルター (プレビュー)**  
 有効

① コンテンツ フィルターは既定の構成を使用し、Azure AI Content Safety を通じて請求されます。  
[詳細情報](#)

**デプロイ** キャンセル

### モデルのヘルプ

▼ **モデルデプロイの詳細**

さまざまな種類のモデルのデプロイプロセスの詳細を取得します:

- [Azure OpenAI Service モデルをデプロイする](#)
- [Phi-3 モデルをデプロイする](#)
- [Cohere モデルをデプロイする](#)
- [Meta Llama モデルをデプロイする](#)
- [Mistral モデルをデプロイする](#)
- [JAIS モデルをデプロイする](#)
- [Jamba-Instruct モデルをデプロイする](#)
- [TimeGEN-1 モデルをデプロイする](#)

① **入門チュートリアルの起動**

- 概要
- モデル カタログ**
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプトフロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
  - モデル + エンドポイント
  - データとインデックス
  - Web Apps
- 管理センター

## DeepSeek-R1 プレビュー

[デプロイ](#)

**詳細** 既存のデプロイ コード サンプル ライセンス

Learn more: [\[original model announcement\]](#)

DeepSeek-R1 excels at reasoning, reasoning, and coding tasks. It fe length.

DeepSeek-R1 builds on the prog extending Chain-of-Thought (Co learning (RL) with fine-tuning on Zero, which relied solely on RL ar

[詳細表示](#)

### データ、メディア、

Property
サポートされるデータ型
サポートされている言語

### 使用状況のガイダンス

#### Content Filtering

Prompts and completions are passed through a default configuration of Azure AI Content Safety classification models to detect and prevent the output of harmful content. Learn more about [Azure AI Content Safety](#). Configuration options for content filtering vary when you deploy a model for production in

### DeepSeek-R1 用のサーバーレス API デプロイ

概要 **価格と使用条件**

#### 価格

DeepSeek R1 の使用は現在 0 ドルで価格が設定されており、使用はレート制限の対象となり、いつでも変更される可能性があります。価格が変更される可能性があり、引き続きご利用になるには新しい価格が適用されます。モデルはプレビュー中です。引き続き使用するには、新しいデプロイが必要になる場合があります。

#### 法的事項

ライセンス情報は、モデル カタログのモデル詳細ページで確認できます

[デプロイ](#) キャンセル

無料な代わりに  
使用レートが制限

### モデルのヘルプ

▼ **モデルデプロイの詳細**

さまざまな種類のモデルのデプロイプロセスの詳細を取得します:

- [Azure OpenAI Service モデルをデプロイする](#)
- [Phi-3 モデルをデプロイする](#)
- [Cohere モデルをデプロイする](#)
- [Meta Llama モデルをデプロイする](#)
- [Mistral モデルをデプロイする](#)
- [JAIS モデルをデプロイする](#)
- [Jamba-Instruct モデルをデプロイする](#)
- [TimeGEN-1 モデルをデプロイする](#)

**1 入門チュートリアルの起動**

- 概要
- モデル カタログ
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプト フロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
- モデル + エンドポイント**
- データとインデックス
- Web Apps

管理センター

# DeepSeek-R1-vlxjf

詳細 実行する

プレイグラウンドで開く | 最新の情報に更新 | 編集 | 削除

### 配置情報

<b>名前</b> DeepSeek-R1-vlxjf	
<b>プロビジョニングの状態</b> 成功	<b>最終更新日時</b> Feb 8, 2025 11:27 PM
<b>作成者</b> 1923306880@utac.u-tokyo.ac.jp	<b>作成日</b> Feb 8, 2025 11:27 PM

---

**モデル**  
[DeepSeek-R1](#)

### アプリケーション開発に役立つリンク

- [コード サンプルリポジトリ](#)
- [チュートリアル](#)

### エンドポイント

**ターゲット URI**  
<https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com>

**キー**  
.....

**コンピューティングの種類**  
Consumption

**Swagger URI**  
<https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com/swagger.json>

### API ルート

Azure AI model inference: Chat Completion  
<https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com/chat/completions>

deepseek-r1: Chat Completion  
<https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com/v1/chat/completions>

### 監視と安全性

Azure AI Content Safety  
有効

### 3.チャットプレイグラウンド からチャットを実行



# DeepSeek-R1-vlxjf

詳細 実行する

[プレイグラウンドで開く](#) [最新の情報に更新](#) [編集](#) [削除](#)

## 配置情報

<b>名前</b> DeepSeek-R1-vlxjf	
<b>プロビジョニングの状態</b> 成功	<b>最終更新日時</b> Feb 8, 2025 11:27 PM
<b>作成者</b> 1923306880@utac.u-tokyo.ac.jp	<b>作成日</b> Feb 8, 2025 11:27 PM

**モデル**  
[DeepSeek-R1](#)

## アプリケーション開発に役立つリンク

- [コード サンプル リポジトリ](#)
- [チュートリアル](#)

## エンドポイント

<b>ターゲット URI</b> https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com
<b>キー</b> .....
<b>コンピューティングの種類</b> Consumption
<b>Swagger URI</b> https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com/swagger.json
<b>API ルート</b> Azure AI model inference: Chat Completion https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com/chat/completions deepseek-r1: Chat Completion https://DeepSeek-R1-vlxjf.eastus2.models.ai.azure.com/v1/chat/completions

## 監視と安全性

Azure AI Content Safety  
有効

- 概要
- モデル カタログ
- プレイグラウンド
- AI サービス
- ビルドとカスタマイズ
- エージェント [プレビュー](#)
- コード [プレビュー](#)
- 微調整
- プロンプト フロー
- 評価と改善
- トレース [プレビュー](#)
- 評価
- 安全性とセキュリティ
- マイアセット
- モデル + エンドポイント**
- データとインデックス
- Web Apps

管理センター

- 概要
- モデル カタログ
- プレイグラウンド**
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプトフロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
- モデル + エンドポイント
- データとインデックス
- Web Apps
- 管理センター

## ← チャットプレイグラウンド

</> コードの表示

**セットアップ** 非表示

**デプロイ\*** [新しいデプロイの作成](#)

DeepSeek-R1-vlxjf

> パラメーター

### チャットセッション

ここから入力を開始

## チャットプレイグラウンドのヘルプ

**チャットプレイグラウンドを探索する**

チャットプレイグラウンドのガイドツアーに参加し、AIソリューションの構築に役立つツールとコントロールについて学習します。

[ツアーを起動する](#)

- ### ▼ 関連リソース
- [データなしでプレイグラウンドでチャット](#)
  - [エンタープライズチャット Web アプリをデプロイする](#)
  - [Azure AI Foundry でプロンプトフローを使用して質問と回答のコパイロットを構築してデプロイする](#)
  - [Azure AI Foundry プレイグラウンドでプロンプトを手動で評価する](#)

[入門チュートリアルを起動](#)

- 概要
- モデル カタログ
- プレイグラウンド**
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプト フロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
- モデル + エンドポイント
- データとインデックス
- Web Apps
- 管理センター

## ← チャットプレイグラウンド

</> コードの表示

**セットアップ** 非表示

**デプロイ\*** [新しいデプロイの作成](#)

DeepSeek-R1-vlxjf

> パラメーター

### チャットセッション

東京大学の良いところを教えてください。

嗯，用户问的是东京大学的好处在哪里。首先，我需要确定用户的需求是什么。可能是一个准备申请东大的学生，或者是对日本教育体系感兴趣的人。他们可能想了解东大的优势，以便做出升学或研究的决定。接下来，我得回想自己知道的关于东京大学的信息。东大是日本顶尖的大学，世界排名也很高，学术方面应该很强。然后，学术自由，教授资源，国际交流这些可能是重点。另外，校园环境和传统文化也是吸引人的地方。

不过，用户可能想知道具体的点，比如课程设置，研究机会，就业支持等。也有可能他们关心学生生活，比如社团活动，校园设施。需要覆盖学术和非学术的优点。

还要考虑结构是否清晰，分点说明会比较清楚。比如分成学术声誉、教育资源、国际性、校园文化、就业支持等部分。另外，可能需要提到知名校友，增强说服力。

但要注意不要只停留在表面，应该深入每个点的具体内容。例如，学术自由方面，东大允许学生探索兴趣，跨学科研究的机会

日本語で教えてください。

ここから入力を開始

## チャットプレイグラウンドのヘルプ

**チャットプレイグラウンドを探索する**

チャットプレイグラウンドのガイドツアーに参加し、AIソリューションの構築に役立つツールとコントロールについて学習します。

[ツアーを起動する](#)

- ▼ **関連リソース**
- [データなしでプレイグラウンドでチャット](#)
  - [エンタープライズチャット Web アプリをデプロイする](#)
  - [Azure AI Foundry でプロンプトフローを使用して質問と回答のコパイロットを構築してデプロイする](#)
  - [Azure AI Foundry プレイグラウンドでプロンプトを手動で評価する](#)

**入門チュートリアルの起動**

- 概要
- モデル カタログ
- プレイグラウンド**
- AI サービス
- ビルドとカスタマイズ
- エージェント プレビュー
- コード プレビュー
- 微調整
- プロンプト フロー
- 評価と改善
- トレース プレビュー
- 評価
- 安全性とセキュリティ
- マイアセット
- モデル + エンドポイント
- データとインデックス
- Web Apps
- 管理センター

## ← チャットプレイグラウンド

</> コードの表示

**セットアップ** 非表示

**デプロイ\*** [新しいデプロイの作成](#)

DeepSeek-R1-vlxjf

> パラメーター

### チャットセッション

#### 6. 地理的優位性

- 国際機関との連携**：本部がある本郷キャンパスはWHO健康開発総合研究センターやJICA研究所と徒歩圏内。インターンシップ先として直接連携。
- 起業支援エコシステム**：東大ベンチャー支援機構（TVP）を通じ、学生起業家に最大300万円の資金助成。東大発スタートアップは年間100社以上誕生。

#### 7. 問題点を逆手に取る機会

- 「競争が激しくストレス」→ 自己管理能力が飛躍的に向上
- 「官僚養成校の批判」→ 社会システムを変える立場に立つ覚悟を養える

#### ～データで見る東大の実力～

- 2024年QS世界大学ランキング：日本1位（世界25位）
- 卒業生の初任給中央値：約32万円（文系）／約35万円（理系）
- 留学生比率：14%（政府の「留学生30万人計画」を大幅に先行達成）

こうした環境で得られるのは「単なる学歴ではなく、社会構造を動かす思考回路」と言えるでしょう。一方、自主性が求められる環境のため、受け身の姿勢では活かしきれない点にも注意が必要です。ご自身のキャリアビジョンと照らし合わせて検討されることをお勧めします。

ここから入力を開始

## チャットプレイグラウンドのヘルプ

### チャットプレイグラウンドを探索する

チャットプレイグラウンドのガイドツアーに参加し、AIソリューションの構築に役立つツールとコントロールについて学習します。

[ツアーを起動する](#)

### 関連リソース

[データなしでプレイグラウンドでチャット](#)

[エンタープライズチャット Web アプリをデプロイする](#)

[Azure AI Foundry でプロンプトフローを使用して質問と回答のコパイロットを構築してデプロイする](#)

[Azure AI Foundry プレイグラウンドでプロンプトを手動で評価する](#)

[入門チュートリアルを起動](#)