

AI-050

補足資料：Module01

- Azure OpenAIにアクセスする
- Azure OpenAI Studio を使用する
- 生成 AI モデルの種類について調べる
- 生成 AI モデルをデプロイする
- プロンプトを使用してモデルから入力候補を取得する
- Azure OpenAI Studio のプレイグラウンドでモデルをテストする

- AIの一種で、新しいコンテンツを自動的に生成する能力を持つもの。

The image shows a chat interface with a light gray background. At the top right, there is a white speech bubble containing the text "Windowsについて教えてください" (Please teach me about Windows). To the right of this bubble is a small icon of a person. Below this, on the left, is a small icon of a robot head. To the right of the robot head is a large white speech bubble containing the AI-generated response. The response consists of three paragraphs of Japanese text. To the right of the AI response, there are two blue callout boxes with white text. The top callout box points to the prompt bubble and contains the text "プロンプト" (Prompt). The bottom callout box points to the AI response bubble and contains the text "AIが生成したコンテンツ" (Content generated by AI).

Windowsについて教えてください

プロンプト

Windowsは、マイクロソフト社が開発したオペレーティングシステム（OS）の一つです。パソコンの基本的な操作を管理する役割を持ち、アプリケーションソフトウェアが正しく動作するための環境を提供します。

初めてリリースされたのは1985年で、その後様々なバージョンがリリースされてきました。現在最新版はWindows 10ですが、2021年にはWindows 11も発表されました。

Windowsはその使いやすさと幅広いソフトウェアの互換性から、世界中で最も広く使用されているオペレーティングシステムの一つとなっています。

AIが生成したコンテンツ

- **生成AI** は **AI** の一種であり、その内部では、**ディープラーニング** のしくみを使用している。
- **生成AI** は、**ディープラーニング**の技術を用いて、既存のデータからパターンを学び、それを基に新しいコンテンツを生成する。

Artificial Intelligence



機械が人間のように思考し、学習し、問題を解決できるしくみ

Machine Learning



アルゴリズムが経験から学習し、その結果を改善していくしくみ

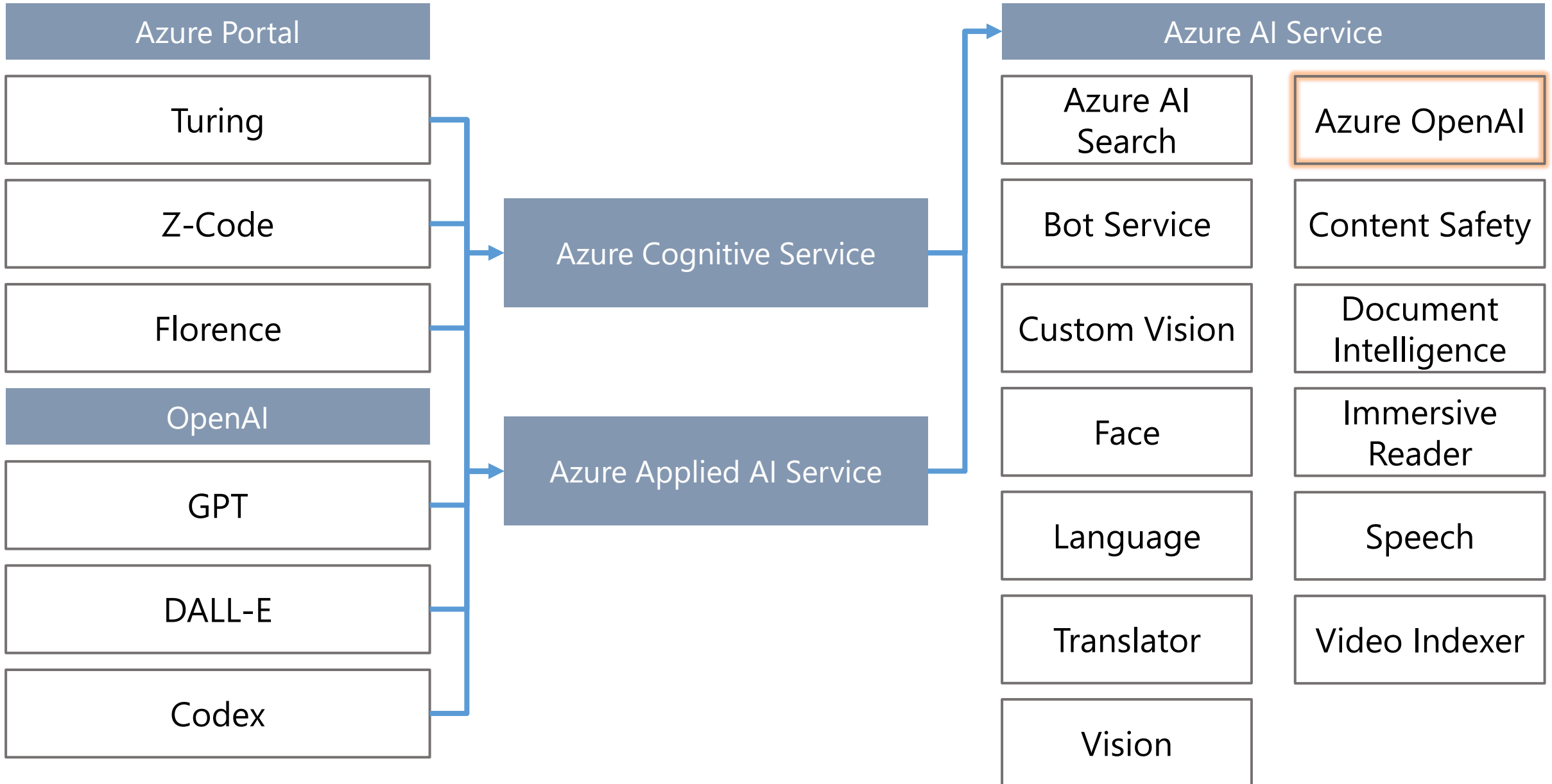
Deep Learning



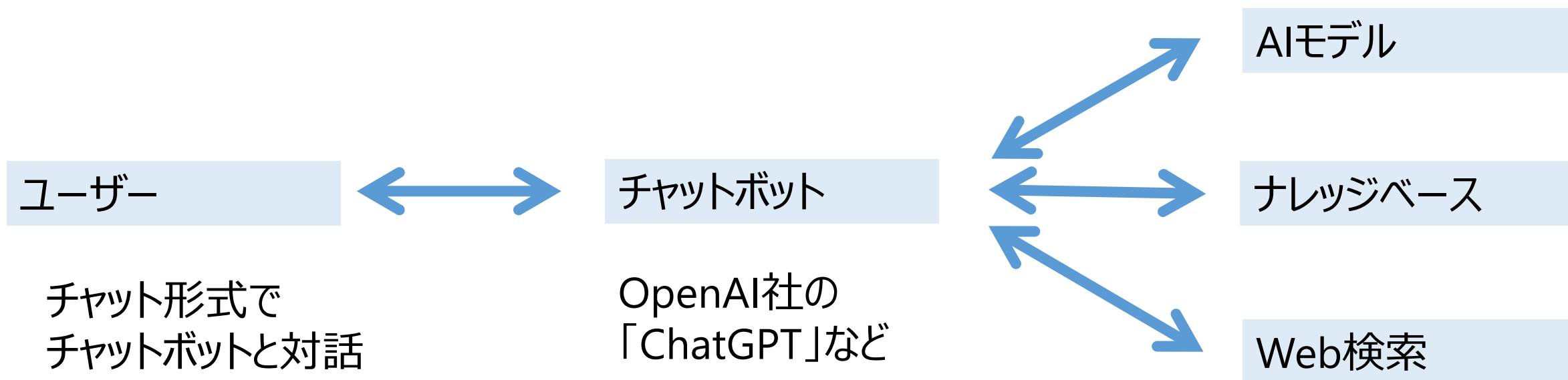
マシンラーニングのうち、人間の脳の神経回路網を模倣した「ニューラルネットワーク」を用いて学習を行うしくみ

- Azure OpenAI Serviceは、マイクロソフトのクラウドプラットフォーム「Microsoft Azure」上で提供されるサービスの一種。
- 「OpenAI社」が開発した**AIモデル**と同じものが、Azure上で提供される。
- ユーザーが独自に開発するアプリやシステムから、**API (Application Programming Interface)**を利用して、Azure上の**AIモデル**を呼び出して使用することが可能となる。
- 質問への回答・プランの提案・要約・翻訳などを行う「AIチャットボット」の開発で使用される。

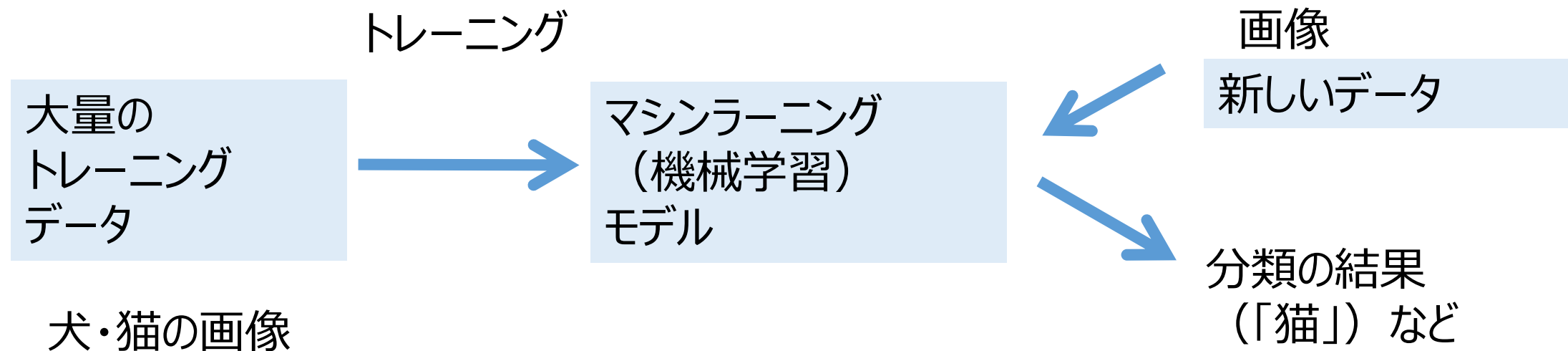
AI Model の系譜 (2024/02現在)




- チャットの形式でユーザーと対話ができるボット（ロボット）のプログラム。
- ユーザーの問い合わせに回答する。
- 「AIアシスタント」とも。



- マシンラーニングやAIにおける「モデル」とは、データから特定のパターンや構造を学習し、未知のデータに対する予測や分類を行うための数学的な表現のことを指す。
- トレーニングデータセットを通じて学習され、その結果得られるパラメータを用いて、新しいデータの予測や分類などを行う。



モデル※	対応API	概要	入力例	出力例
ChatGPT モデル	Chat Completions	ユーザの入力をチャット形式で返答する。 用途に最適化させるような指示や問いを出すことで自然言語での返答を得られる。	Microsoftについて教えてください	Microsoftは、アメリカ合衆国ワシントン州に本社を置く、ソフトウェアやクラウドサービスを開発、販売する会社です。
InstructGPT モデル	(Text) Completions	ユーザの入力したテキストの続きを予測する。 比較的ChatGPTよりも簡潔に回答を返す。 (喋りすぎない)	日本語: GPTは便利 英語:	GPT is convenient.
Embedding モデル	Embeddings	入力された単語や文章を数値データ(ベクトル)化する。自然言語が定量化されることで、文書同士の類似度を計算でき検索などに利用可能。	今日は晴れです	[0.89, -0.93, -0.26, 0.45] (「今日は晴れです」を定量的に表現したベクトル)
DALL-E モデル	Image generation	入力言語から画像を作成するモデル。 自然言語で柔軟に生成画像の画風や特徴を指示できる。	赤帽子をかぶったペンギン	
Whisper モデル	Speech to text	音声データをテキストへ変換するモデル。	「こんにちは」という音声ファイル	「こんにちは」というテキスト
TTS モデル	Text to Speech (preview)	テキストデータを音声へ変換するモデル。	こんにちは	「こんにちは」という音声

2024/5 時点

ChatGPT (Chat completions API で使用)

会話ベース
GPT-3.5 モデル

gpt-3.5-turbo (トークン 4k)
gpt-3.5-turbo-16k

ユーザの問い合わせやタスク指示に対して
会話ベースで返答が可能。GPT-4 に比べ高速・安価。

GPT-4 モデル

gpt-4 (トークン 8k)
gpt-4-32k

ユーザの入力に対して極めて高い解釈性を持ち、
指示に沿った返答が可能。

GPT-4 Turbo
モデル

gpt-4 (1106,0125,2024-04-
09モデル) (トークン 128k)

GPT-4の高速化・高精度化を実現し、
最新モデルはVisionとのマルチモーダルに対応

GPT-4o
モデル

gpt-4o (トークン 128k)

GPT-4 Turboの高速化・高精度・低コスト化を実現し、
最新モデルはVisionとのマルチモーダルに対応。
Voiceのマルチモーダル対応も将来的に可能。

InstructGPT (Completions API で使用)

指示応答ベース
GPT-3.5 モデル

gpt-3.5-turbo-instruct
(トークン 4k)

チャットモデルに比べ、いわゆる「喋り過ぎ」を抑制モデル。
パラメータ抽出などバックエンドの処理に使われるケースが
多い。

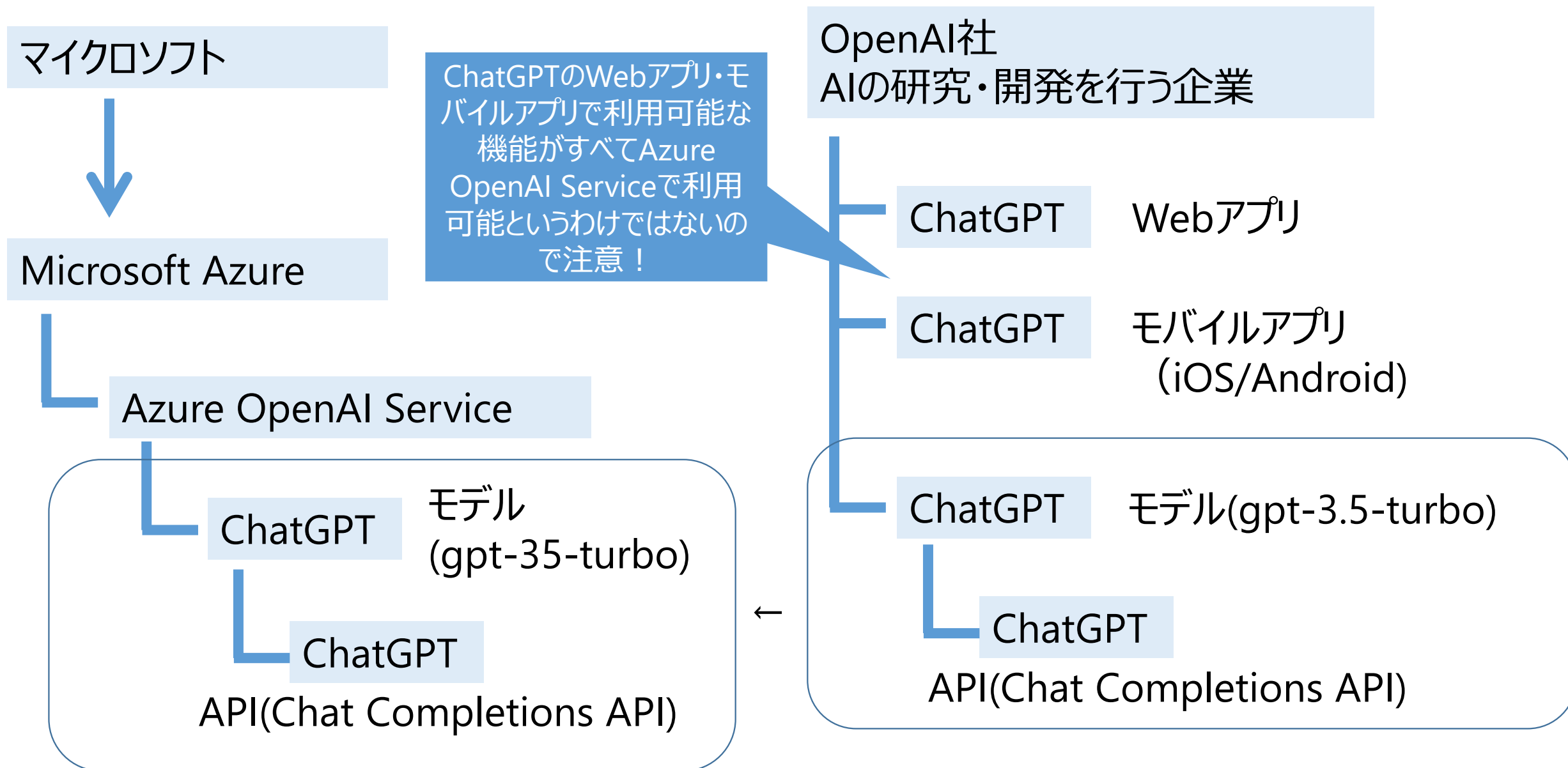
Embedding (Embeddings API で使用)

text-embedding-ada-002
text-embedding-3-small
text-embedding-3-large
(トークン 8k)

高速で安価。一般の Embedding モデルと比較し、
大きなトークンサイズでもベクトル化が実行可能。

- ChatGPT（チャットジーピーティー、Chat Generative Pre-trained Transformer）は、OpenAIが 2022/11/30 に公開した**チャットボット**。
<https://chat.openai.com/> からアクセスできる。
- iOS用・Android用のモバイルアプリ「ChatGPT」もある。
- 有料サービス「ChatGPT Plus」（月額20ドル）に加入すると、最新のAIモデル「GPT-4」の使用、Web検索、プラグインの利用、音声による会話、画像の入力などが可能となる。
- 世間一般でいう「ChatGPT」はこれらを指す（場合が多い）

- 2023/3/1に、OpenAI社は「ChatGPT」というモデルと、それを利用するための「ChatGPT」APIの提供を開始した。これらは「ChatGPT」アプリで使用されているものと同じである。
- 「ChatGPT」のAPIは <https://api.openai.com/v1/chat/completions> からアクセスできる（つまり ChatGPT API = **Chat Completions API** である）
- 「ChatGPT」APIが使用しているモデルは「gpt-3.5-turbo」という名前である。
- 2023/3/9 に、Azureでも ChatGPT モデル「**gpt-35-turbo**」と、それに対応するAPIが利用可能となった。
- その後、「**gpt-4**、**gpt-4-turbo**、**gpt-4o**」とモデルがアップデートしている
- **Azureで使える「ChatGPT」とは、これらのモデルやAPIを指す。**



- 2023/3/15 OpenAIが「**GPT-4**」モデルとAPIの提供を開始（ウェイトリストへの登録受付開始）。より高度なタスク処理が可能となった。
- 2023/3/21, Azure OpenAI Serviceで「**GPT-4**」モデルの提供を開始（プレビュー）。GPT-4の利用申請が別途必要。
- 2023/7/6 OpenAIが「**GPT-4**」モデルとAPIの一般提供を開始。すべてのユーザーが利用可能になった。
- 2023/9より、Azure OpenAI Serviceで「**GPT-4**」モデルがすべてのユーザーに対して提供開始。GPT-4の利用申請が不要になった。

- 2023/5 Microsoft Bing Chat (「新しいBing」とも) を一般公開。内部では **GPT-4** をベースとした独自の「Prometheus」モデルが使用されている
 - 2023/7 Microsoft Bing Chat Enterprise 発表
- 2023/9/26 Windows Updateで、Windows 11にAI機能「Copilot in Windows」が追加された。内部的には **Bing Chat** が使用されている。
- 2023/11/1 Microsoft 365 Copilotの一般提供開始。内部では **GPT-4** を使用している。

- 2023/11/6 **OpenAI DevDay** 開発者カンファレンスにおいて、多数の新機能が発表された。抜粋:
 - **GPT-4 Turbo**: GPT-4の改良版、より新しいデータで学習済み、より多くのトークン进行处理可能、より安価、など。
 - **GPTs** (ジーピーティーズ) : ユーザーの目的に合わせてカスタマイズ可能なGPT。Web検索、Code Interpreter、独自データファイル追加などを利用可能。

- GPT-4に、**画像**と、**画像に対する質問**を送信し、質問への回答を生成することができるようになった
- 使い方はChat Completions APIとほぼ同じ
- 画像データはURLまたはBase64エンコードで指定
- 2023/9/25 ChatGPTアプリへの画像入力が可能に
- 2023/11 gpt-4-vision-preview APIとしてAPIからも利用可能に
- Azure OpenAI Serviceでの対応予定: 2023年中と発表されている

- GPT-4oは、OpenAIが2024年5月13日に発表した新しいAIモデルです。"o"は"omni"を意味し、テキスト、音声、画像の組み合わせを入力として受け取り、それらを組み合わせて出力を生成します。
- 音声入力に対しては232ミリ秒、平均で320ミリ秒で応答することが可能です。
- 英語以外の言語においても性能が向上しており、日本語や中国語を含む20言語でトークナイザーが改善され、トークン使用量が圧縮されています。
- GPT-4oはGPT-4 Turboに比べて2倍の速さで、コストは半分、Rate Limitは5倍に設定されています

AI-050の範囲の内容（今後の進化はLearnを参照）

(1)コンテンツの生成

- ユーザーが指定した文章の続きを生成、ユーザー質問に回答

(2)コードの生成

- C#/Pythonなどのプログラムコードを生成

(3)画像の生成

- ユーザーが指定した文章から画像を生成

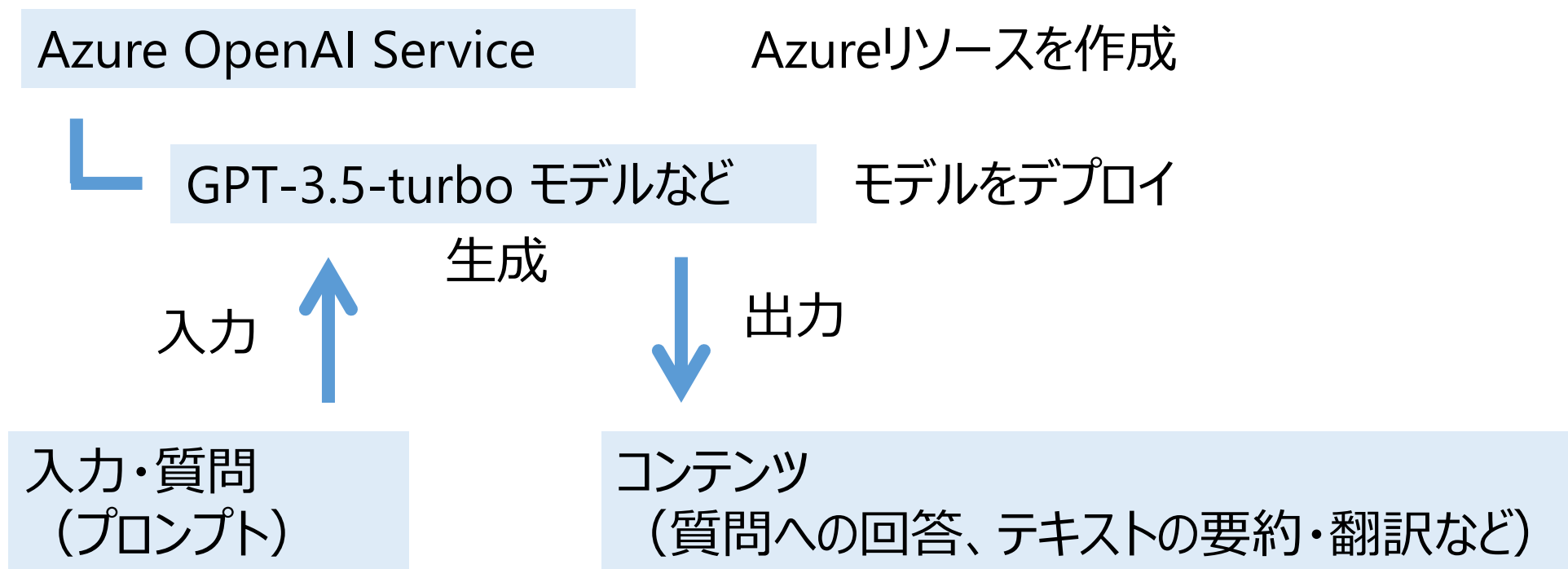
(4)音声をテキストに変換する

- wavファイルからテキストを文字起こし

(5)テキストをベクトルに変換する

- テキストの検索などに使用されるしくみ

- 自然言語による入力・質問から、コンテンツを生成。



以下の文章を3行で、箇条書きで要約してください。

====

生成的人工知能（せいせいてきじんこうちのう、英: generative artificial intelligence）または生成AIは、プロンプトに応答してテキスト、画像、または他のメディアを生成することができる人工知能システム的一种である[5][6]。生成AIモデルは、入力された訓練データの規則性や構造を学習し、同様の特性を持つ新しいデータを生成する[7][8]。ジェネレーティブAI、ジェネラティブAIともよばれる。

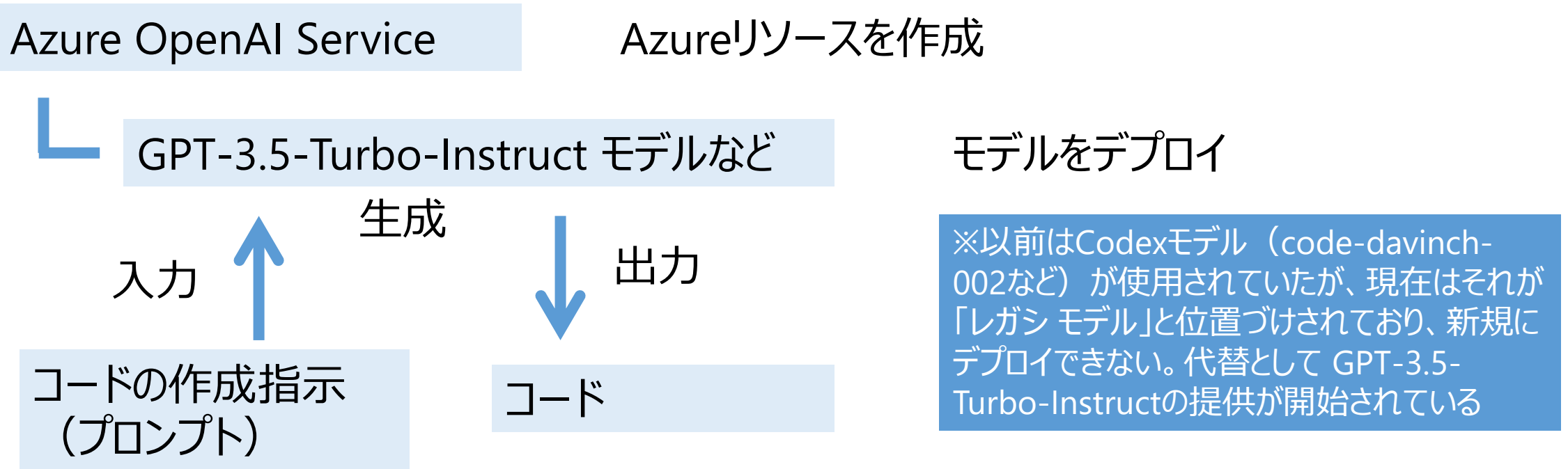
著名な生成AIシステムとして、OpenAIがGPT-3やGPT-4の大規模言語モデル[9]を使用して構築したチャットボットのChatGPT（および別形のBing Chat）や、GoogleがLaMDA基盤モデルに構築したチャットボットBardがある[10]。その他の生成AIモデルとして、Stable DiffusionやDALL-Eなどの人工知能アートシステムがあげられる[11]。

生成AIは、アート、執筆、ソフトウェア開発、ヘルスケア、金融、ゲーム、マーケティング、ファッションなど、幅広い業界で応用できる可能性がある[12][13]。生成AIへの投資は2020年代初頭に急増し、Microsoft、Google、Baiduなどの大企業だけでなく、多数の中小企業も生成AIモデルを開発している[5][14][15]。しかし、人をだましたり操作したりするフェイクニュースやディープフェイクの作成など、生成AIの悪用の可能性も懸念されている[16]。



- 生成的人工知能（生成AI）は、入力された訓練データを学習し、新しいデータを生成するシステムで、OpenAIのGPT-3やGoogleのLaMDAなどが存在する。
- アート、ソフトウェア開発、ヘルスケアなど、多岐にわたる業界での応用が可能で、2020年代初頭にはその投資が急増している。
- しかし、フェイクニュースやディープフェイクの作成など、生成AIの悪用の可能性も懸念されている。

- C#、Python、JavaScript、Perl、PHPなどさまざまなプログラミング言語のコードを生成可能
- 単体テストの生成、SQLクエリの生成、Azureリソースの定義ファイル（ARMテンプレート・Bicepファイル）の生成なども可能



例: コードの生成 (Completionsプレイグラウンド)

24

The screenshot shows the Azure AI Studio interface for the Completions playground. The left sidebar contains navigation links: Azure OpenAI, Playground, Chat, Completions (selected), DALL-E (Preview), Management, Deployments, Models, Data files, Quotas, and Content filters (Preview). The main area is titled 'Completions playground' and shows the 'deploy-gpt-35-turbo-0301' deployment selected. The input text is 'CSVファイルを生成するC#コードを記述してください。' followed by a code block for a CSV writer. The 'Parameters' panel on the right includes sliders for Temperature (1), Max length (tokens) (500), and Stop sequences (empty). It also has checkboxes for Pre-response text and Post-response text, both set to 'Enter text'. The 'Generate' button is at the bottom left, and the token count is 140.

```
using System;
using System.IO;
using System.Text;

class CSVWriter
{
    static void Main()
    {
        var sb = new StringBuilder();
        sb.AppendLine("ID,名前,年齢,性別");
        sb.AppendLine("1,山田太郎,20,男性");
        sb.AppendLine("2,鈴木花子,25,女性");
        sb.AppendLine("3,佐藤次郎,30,男性");

        File.WriteAllText("test.csv", sb.ToString());
    }
}
```

※CodexモデルはChatGPT/GPT4で代替可能。すでに**最新のモデルでは Completions プレイグラウンドは使用できない。**

C#を使用してCSVファイルを生成するコードを書いてください。



以下に、C#を使用してCSVファイルを生成する基本的なコードを示します。

```
using System;
using System.IO;
using System.Text;

class Program
{
    static void Main(string[] args)
    {
        // CSVファイルのパスを指定します
        string filePath = @"C:\temp\sample.csv";

        // データを作成します
        StringBuilder sb = new StringBuilder();
        sb.AppendLine("ID,名前,年齢");
        sb.AppendLine("1,山田太郎,20");
        sb.AppendLine("2,鈴木一郎,30");
        sb.AppendLine("3,佐藤次郎,40");

        // ファイルに書き込みます
        File.WriteAllText(filePath, sb.ToString(), Encoding.UTF8);

        Console.WriteLine("CSVファイルの生成が完了しました。");
    }
}
```

このコードは、指定したパスにCSVファイルを生成し、その中にいくつかのデータを書き込むものです。データはStringBuilderを使用して作成され、File.WriteAllTextメソッドを使用してファイルに書き込まれます。

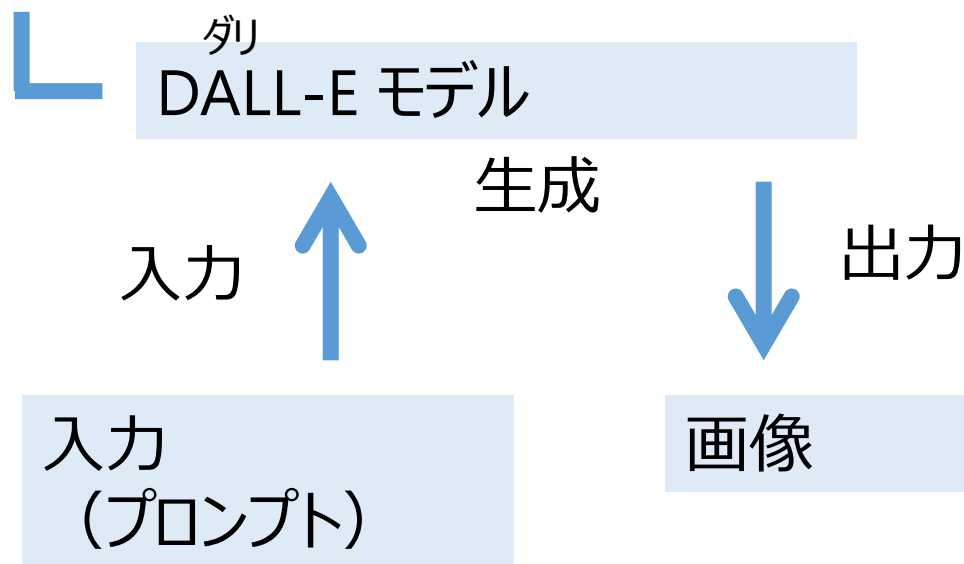
(3) 画像の生成

26

- プロンプトに入力されたテキストから画像を生成

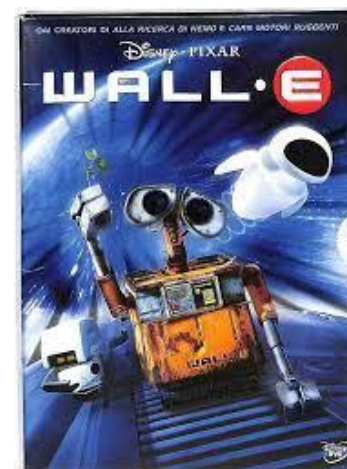
Azure OpenAI Service

Azureリソースを作成



※ユーザーによるDALL-Eモデルのデプロイは不要

※DALL-E（ダリ）は、Pixarのキャラクター WALL-E と、スペインの芸術家サルバドール・ダリ (Salvador Dali)の名前に由来。



プロンプト ⓘ

Tokyo urban landscape, realistic



Tokyo urban landscape, realistic

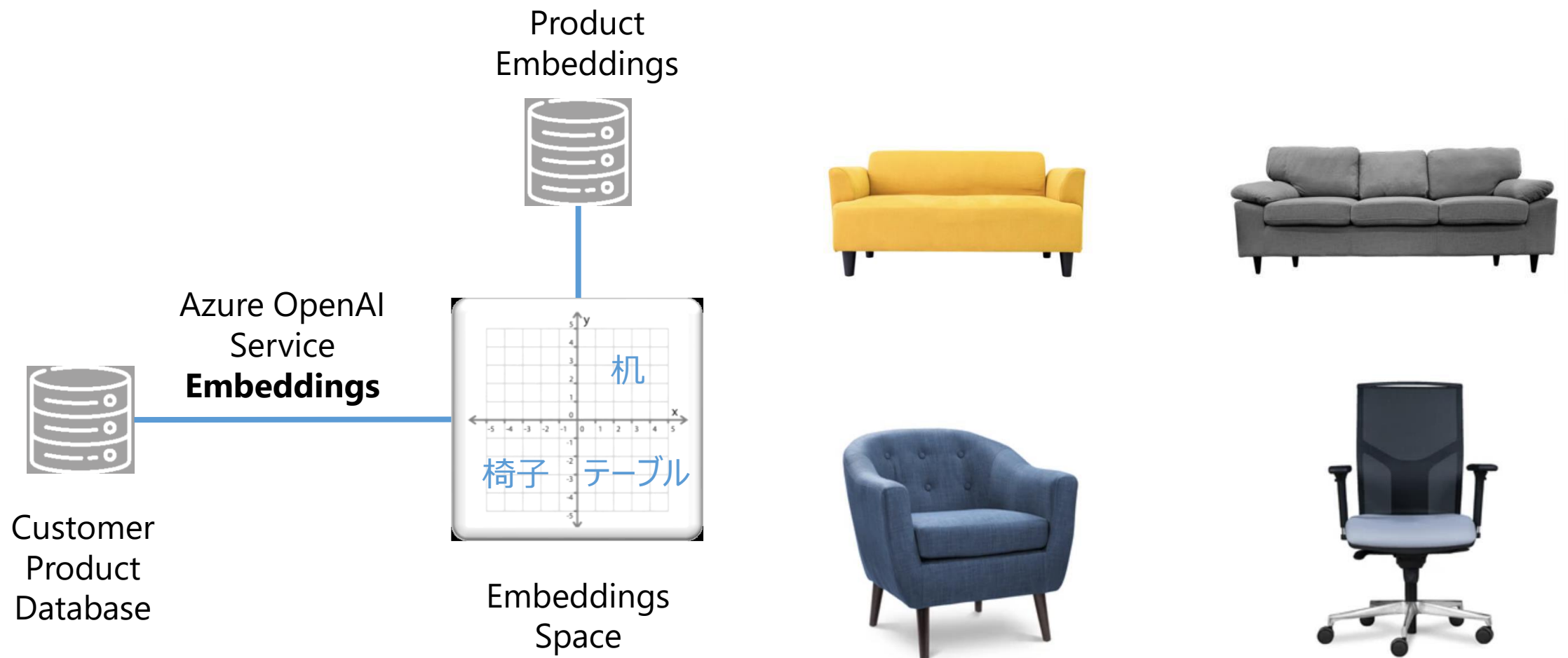


東京の都会の風景、写实的

- OpenAIが開発した音声認識のしくみ。speech to text（音声を文章に変換する）が可能。
- 「Whisper API」と、そこで使われるモデル「whisper-1」「whisper 2」「Whisper 3」がある。
- オープンソース化されている（ローカルPC上などでも実行可能）
- 2023/9/18 Azure OpenAI ServiceとAzure AI Speechでの Whisper モデル / Whisper API 提供 プレビュー開始。

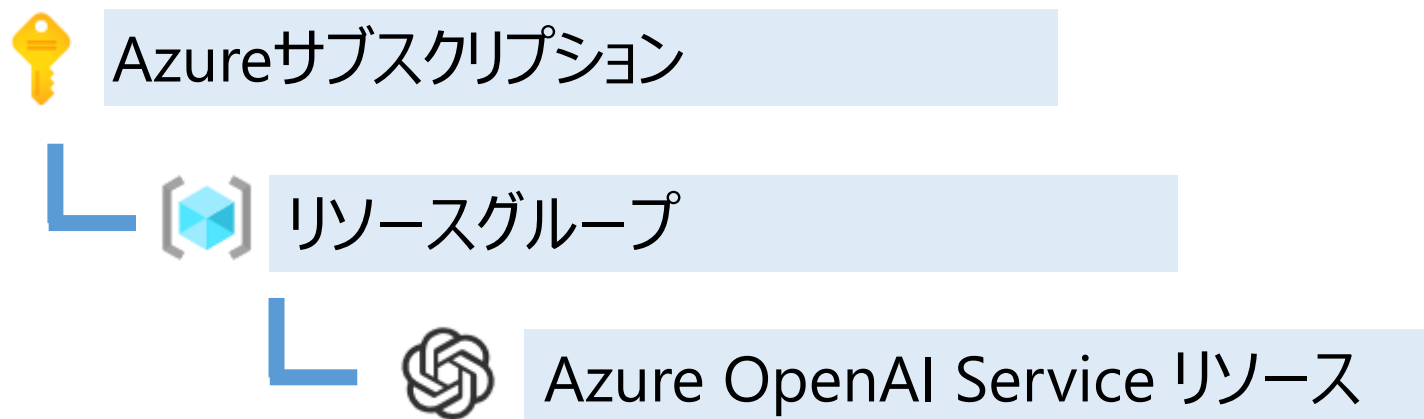
(5) テキストをベクトルに変換（Embeddings、「埋め込み」） 29

- ※なにかをどこかに「埋め込む」といったものではない。
- 自然言語処理（NLP）において、文章を浮動小数点の高次元数値ベクトル（1536次元）に変換するしくみ、またはその数値表現。
 - 数値ベクトル: [1.23, 4.56, 7.89, ...] のようなもの。
- 2つの文章の数値ベクトル（のコサイン類似度）を計算することで、2つの文章の類似度を判定することができる。
- たとえば、「議事録のデータベースを使用して、ユーザーが入力した質問に関係性が高い議事録を検索する」といった場面で使用できる。



それぞれの製品の特長などを数値化してグラフにマッピングし、近しいものを選ぶ

- Azureサブスクリプションでは、事前に利用申請を済ませておく。
- Azureの他のリソースと同様に、Azure portal、Azure CLI、Azure PowerShell、ARMテンプレート、Bicepなどを使用して、リソースを作成できる。



The screenshot displays the Microsoft Azure portal interface. On the left, the navigation pane shows 'すべてのサービス' (All Services) highlighted with a red box. Below it, 'お気に入り' (Favorites) is also visible. The main content area is titled 'すべてのサービス | AI + Machine Learning'. A search bar for 'サービスのフィルター' (Service Filter) is present. The 'Azure AI サービス' (Azure AI Services) section lists various services, with 'Azure OpenAI' highlighted by a red box and marked with a blue star. The '機械学習' (Machine Learning) section is also visible at the bottom.

Microsoft Azure リソース、サービス、ドキュメントの検索 (G+)

すべてのサービス | AI + Machine Learning

すべて サービスのフィルター

お気に入り

最近使用したもの

推奨

Azure AI サービス

- Azure AI services
- Azure AI Video Indexer
- Bot Service
- Computer Vision
- Custom Vision
- Face API
- Language
- Azure OpenAI** ★
- 音声サービス
- Azure AI services multi-service account
- Anomaly Detector
- Cognitive Search
- Content Moderator
- Document intelligences
- Immersive Reader
- Metrics Advisor
- Personalizer
- 翻訳

機械学習

- Bonsai
- Azure Machine Learning
- Intelligent Recommendations アカウント
- Azure Synapse Analytics

カテゴリー

- AI + Machine Learning**
- 分析
- コンピューティング
- コンテナ
- データベース
- DevOps
- 全般
- ハイブリッド + マルチクラウド
- ID
- 統合
- モノのインターネット
- Management and governance
- 移行
- Mixed Reality

リソースの作成

ホーム

ダッシュボード

すべてのリソース

リソース グループ

App Service

関数アプリ


SQL データベース

Azure Cosmos DB

Virtual Machines

ロード バランサー

ストレージ アカウント

 Microsoft Azure

リソース、サービス、ドキュメントの検索 (G+/)

すべてのサービス > Azure AI services



 **Azure AI services** | Azure OpenAI  ...

Azure AI services

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

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


2 件中 1 ~ 2 件のレコードを表示しています。

<input type="checkbox"/> 名前 ↑↓	サブタイプ ↑↓	場所 ↑↓
<input type="checkbox"/>  aoaigpt3eastus92837... OpenAI		East US
<input type="checkbox"/>  aoaigpt4canadaeast9... OpenAI		Canada East


Azure AI services

 Azure OpenAI

 Cognitive Search

 Computer Vision

 Face API

 Custom Vision

[すべてのサービス](#) > [Azure AI services](#) | [Azure OpenAI](#) >

Azure OpenAI の作成 ...

① 基本 ② ネットワーク ③ タグ ④ レビューおよび送信

GPT-3 モデルを利用した OpenAI の言語生成機能により、新しいビジネス ソリューションを実現します。そのモデルは、何兆もの単語で事前トレーニングされており、推論時にいくつかの短い例を示すことで、シナリオに簡単に適応できます。概要作成からコンテンツ、コード生成まで、さまざまなシナリオに適用できます。

[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 * ⓘ

インスタンスの詳細

リージョン ⓘ

名前 * ⓘ

価格レベル * ⓘ

aoairg3

[新規作成](#)

East US

myopenaiservice928374

Standard S0

サブスクリプションを選択

リソースグループを選択または作成

リージョンを選択

世界中で一意となる（重複しない）リソース名を入力

価格レベルを選択（現在 Standard S0のみ）

すべてのサービス > Azure AI services | Azure OpenAI >

Azure OpenAI の作成 ...

✓ 基本 **2 ネットワーク** ③ タグ ④ レビューおよび送信

i Configure network security for your Azure AI services resource.

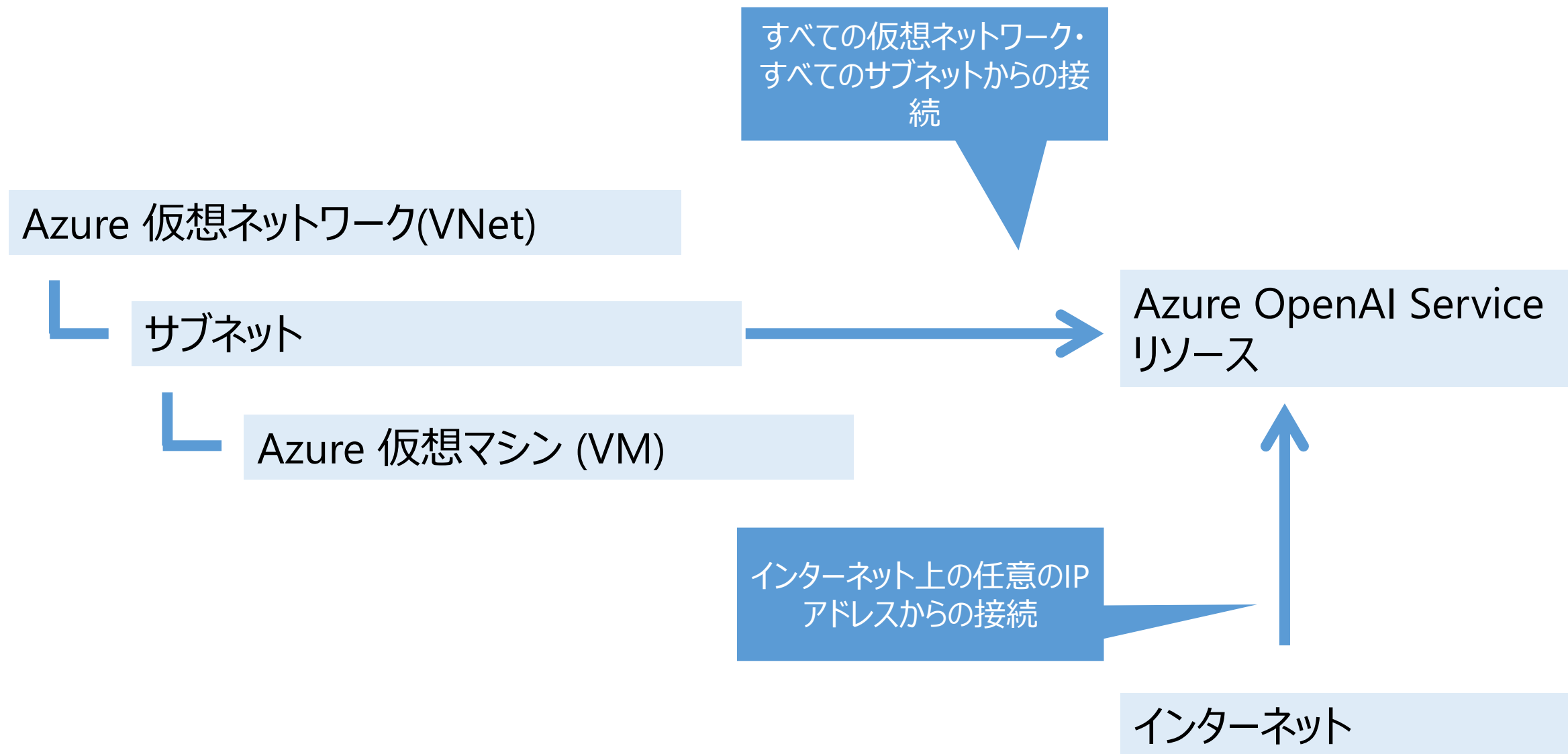
Azure OpenAI Serviceにアクセス可能なネットワークを選択。

デフォルトでは、インターネットを含むすべてのネットワークから接続が可能

選択したネットワーク・指定したIPアドレス範囲からの接続のみ

プライベート
エンドポイント接続のみ

- ☒ 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。



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。

仮想ネットワークを構成します

Virtual network * ⓘ

(新規) vnet01



[新規作成](#)

Subnets *

(新規) subnet-1 (10.0.0.0/26)

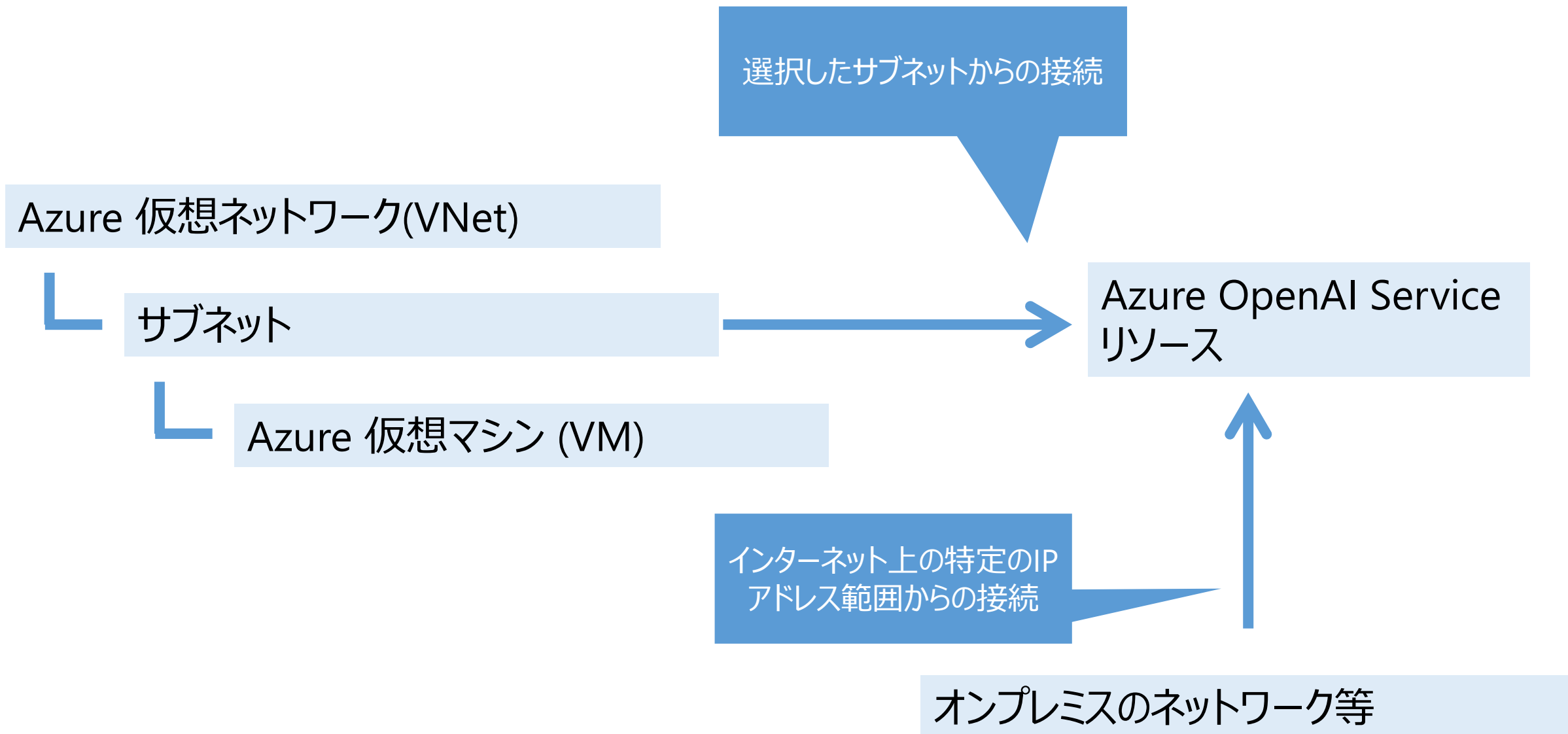


ファイアウォール

インターネットまたはオンプレミスのネットワークからのアクセスを許可するための IP 範囲を追加します。

[詳細情報](#)

アドレスの範囲



種類 *

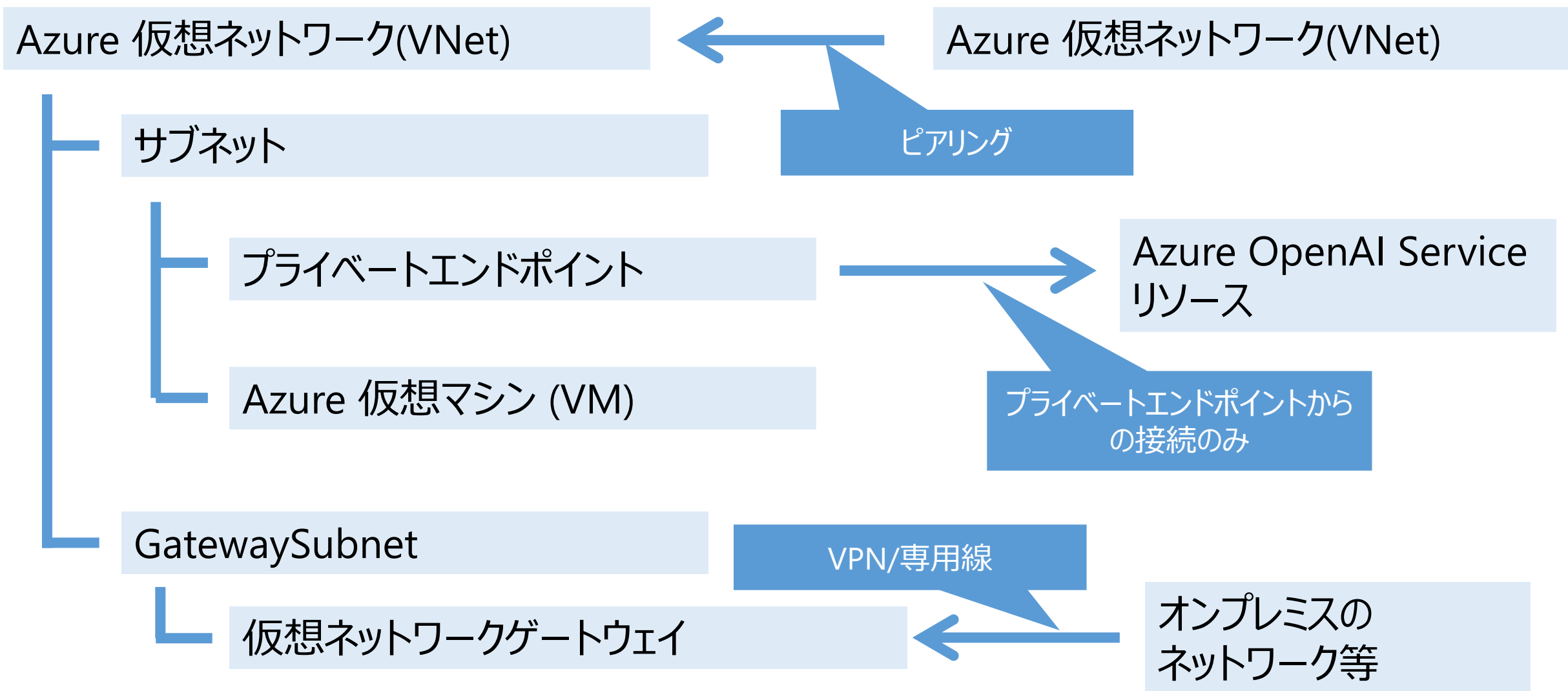
- ☐ インターネットを含むすべてのネットワークがこのリソースにアクセスできます。
- ☐ Selected networks, configure network security for your Azure AI services resource.
- ☒ 無効になっています。どのネットワークもこのリソースにアクセスできません。このリソースにアクセスする唯一の方法となるプライベート エンドポイント接続を構成できます。

プライベート エンドポイント

Create a private endpoint to allow a private connection to this resource. Please make sure that the private endpoint has the same location as this resource. Additional private endpoint connections can be created within the Azure AI services account or private link center.

[+ Add Private Endpoint](#) [Delete](#)

<input type="checkbox"/>	Subscription	Private endpoint	Resource group	Region	Target su
--------------------------	--------------	------------------	----------------	--------	-----------



すべてのサービス > Azure AI services

Azure AI services | Azure OpenAI

Azure AI services

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

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

2 件中 1 ~ 2 件のレコードを表示しています。 グループ化なし

<input type="checkbox"/> 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 ↑↓
<input type="checkbox"/>  aoaig ...	OpenAI	East US	 aoaig ...	S0	Succeeded
<input type="checkbox"/>  aoaig ...	OpenAI	Canada East	 aoai ...	S0	Succeeded



Azureサブスクリプション

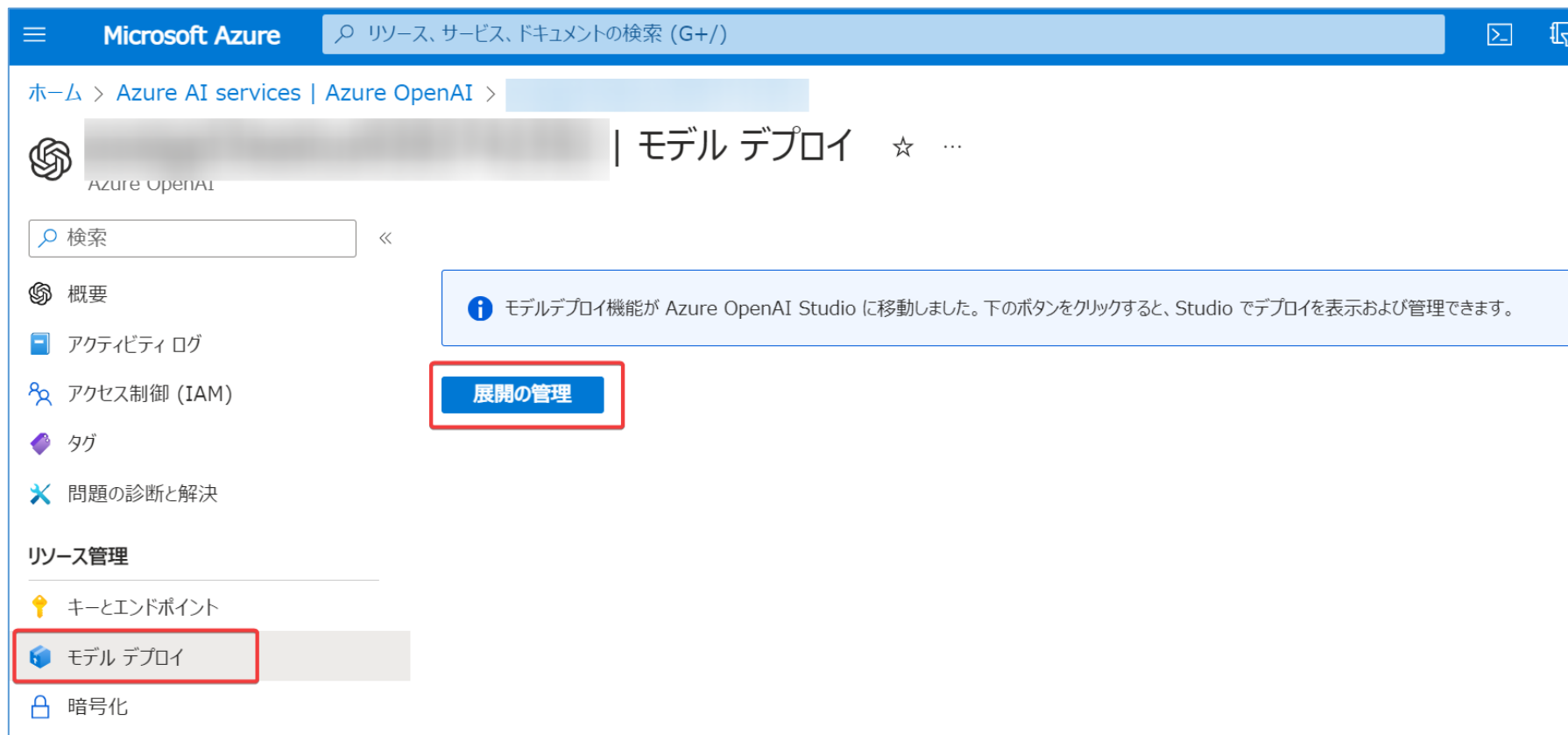


リソースグループ

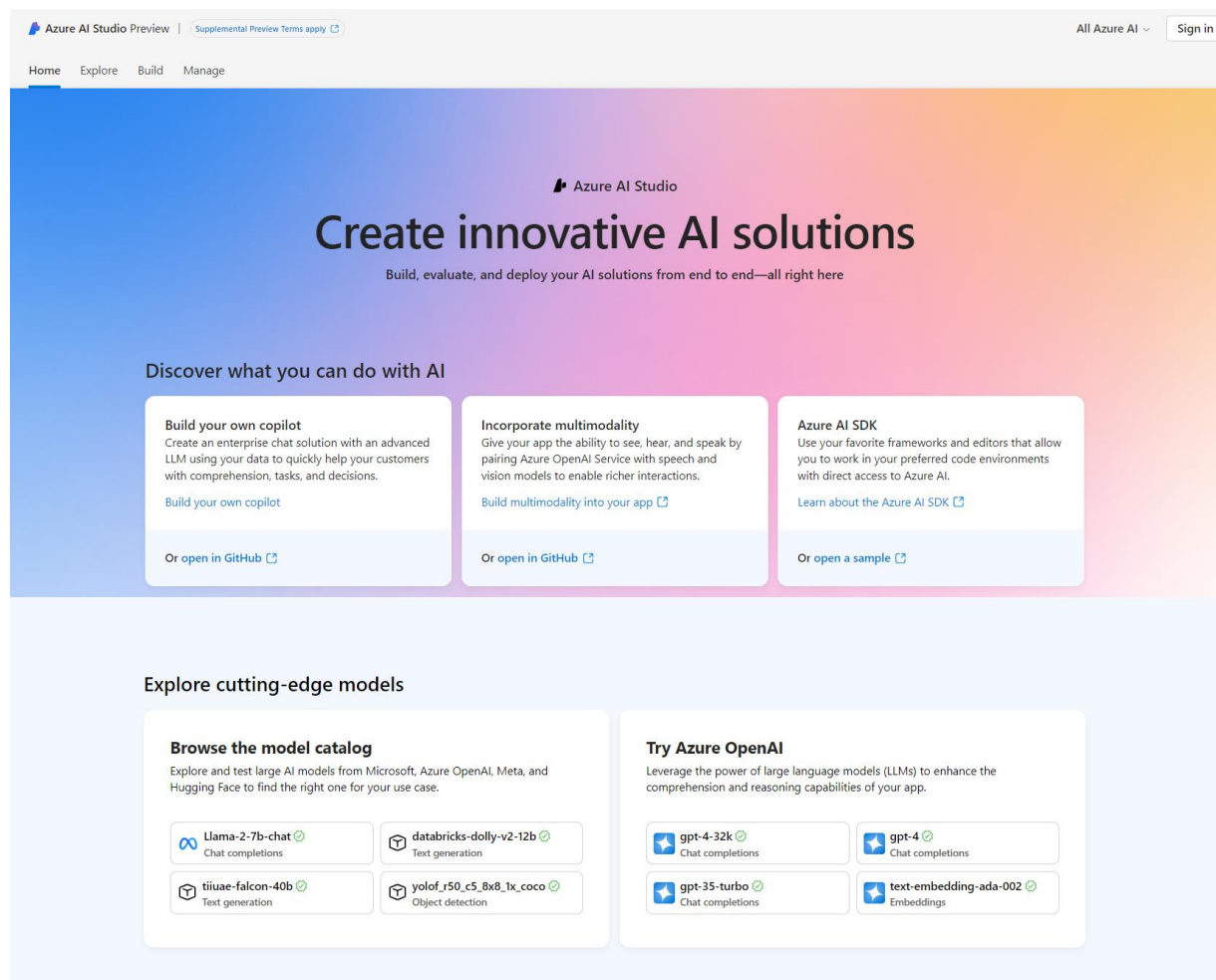


Azure OpenAI Service リソース

- モデルのデプロイや動作確認は、Azure portal内ではなく、「Azure OpenAI Studio」という専用サイトから行う。



<https://oai.azure.com/>



Azureの主なAI機能を集約したような管理画面。ゆくゆくはAzure OpenAI Studioの機能はこちらに移行していくものと思われる。

- モデルの検索とデプロイ
- プレイグラウンドの利用
- AIプロジェクトの作成

<https://ai.azure.com/>

Azure OpenAI Studio でモデルを選択しデプロイする 44

Azure AI | Azure OpenAI Studio

Naoki Abe
ailearn20230926 (米国東部, S0)

Azure OpenAI

プレイグラウンド

チャット

入力候補

DALL-E

アシスタント (プレビュー)

管理

デプロイ

モデル

データ ファイル

クォータ

コンテンツ フィルター (プレビュー)

Azure OpenAI Studio > モデル

プライバシーと Cookie

Azure AI Studio の探索

モデル

Azure OpenAI は、異なる機能と価格ポイントを備えたモデルによって動作します。提供された基本モデルの 1 つをデプロイして [Playground](#) で試すか、カスタム モデルを特定のユース ケースとデータでトレーニングして、パフォーマンスを向上させ、より正確な結果を得られます。
[さまざまな種類の基本モデルに関する詳細情報](#)

ベース モデル

デプロイ

カスタム モデルの作成

列のオプション

最新の情報に更新

検索

モデル名 ↑ ↓	モデル バージ...	作成日時 ↓	状態 ↓	デプロ...
dall-e-2	2.0	2024/4/15 9:00	成功	はい
dall-e-3	3.0	2023/8/11 9:00	成功	はい
gpt-4	turbo-2024-04-09		成功	はい
gpt-4	0613	2023/6/19 9:00	成功	はい
gpt-4	1106-Preview	2023/11/15 9:00	成功	はい
gpt-4	0125-Preview	2024/1/25 9:00	成功	はい
gpt-4-32k	0613	2023/6/19 9:00	成功	はい
gpt-4o	2024-05-13		成功	はい
gpt-35-turbo	0125	2024/2/15 9:00	成功	はい
gpt-35-turbo	1106	2023/11/15 9:00	成功	はい
gpt-35-turbo	0613	2023/6/19 9:00	成功	はい



Azureサブスクリプション



リソースグループ



Azure OpenAI Service リソース

gpt-35-turbo
(0301)

gpt-35-turbo-instruct
(0914)

DALL-E

ベースモデル

※DALL-Eのベースモデルは
一覧には表示されない。
デプロイも不要。

デプロイ: モデルの種類、モデルバージョンなどを選択して、どのモデルをどのように使用するかという**設定を作る**こと。モデルのデプロイによってモデルが新たに作られるわけではない。各デプロイは Azure OpenAI Service環境に用意された「ベースモデル」を指す。

deploy1

TPM=10K

deploy2

TPM=20K

TPM=Tokens Per Minute
1分あたりのトークン上限数。
トークン処理量 (=コスト)
を制限

生成 AI モデルの種類

モデル名 ▾

モデルバージョン ▾

gpt-35-turbo

0613

gpt-35-turbo

0301

gpt-35-turbo-16k

0613

gpt-35-turbo-instruct

0914

text-embedding-ada-002

2

text-embedding-ada-002

1

GPT（テキストコンテンツの生成、
質問への回答、要約など）を行う
モデル
※コード生成も可能

コードの生成に特化したモデル
※以前使われていた Codex モデル
の代替となる

埋め込み（embedding）計算
を行うモデル

モデル名 ▾

モデルバージョン ▾

gpt-35-turbo

0613

gpt-35-turbo

0301

gpt-35-turbo-16k

0613

0613バージョンは、0301バージョンよりも、モデルがシステムメッセージにより従いやすくなり、操作性が向上している。

0301バージョンは2024/7/5に廃止予定。

名前に -16k が付いたモデルは、一連の会話において、以前のモデルよりも多くの「トークン」を処理できる。より長い文章の入力や生成に対応が可能。

gpt-35-turbo: 4,096 トークン
gpt-35-turbo-16k: 16,384 トークン

- 機械学習で、文章を分割・デジタル化して記録する際の単位。
- たとえばhelloという単語は、ASCII文字コードを使って記録すると5バイトになるが、トークンとしてはこれを1つの整数値として、効率よく記録している。
- 1単語 = 1トークンとは限らない。1単語1トークンの場合もあれば、複数の単語をまとめて1トークンとする場合もあるし、長い単語を分割して複数トークンとする場合もある。
- AIのサービスでは、トークンを使用して、あるトークンの次に出現する確率が高いトークンを計算しながら、文章を生成している。
- また、入力したトークンと、出力されたトークンを計算し、それによって料金を計算している。

GPT-3 Codex

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🍌

Sequences of characters commonly found next to each other may be grouped together: 1234567890

Clear

Show example

Tokens	Characters
64	252

Many words map to one token, but some don't: indivisible.

Unicode characters like emojis may be split into many tokens containing the underlying bytes: 🍌

Sequences of characters commonly found next to each other may be grouped together: 1234567890

TEXT

TOKEN IDS

<https://platform.openai.com/tokenizer>

- Chat プレイグラウンド
- Completions プレイグラウンド
- DALL-E プレイグラウンド
- アシスタント プレイグラウンド

プレイグラウンド：直訳すると「遊び場」。

これらのプレイグラウンド画面を使って、各モデルの動作をすばやく確認することができる。

これはエンドユーザー向けではなく、開発者向けの画面。



- 内部では「Chat Completions API」を使用

The screenshot shows the Azure OpenAI Studio Chat playground interface. The left sidebar contains a navigation menu with the following items: Azure OpenAI, プレイグラウンド (Playground), チャット (Chat), 入力候補 (Suggestions), DALL-E (Preview), 管理 (Management), デプロイ (Deploy), モデル (Model), Data files, Quotas, and Content filters (Preview). The 'チャット' (Chat) item is highlighted with a red box, and a red arrow points from it to the 'システム メッセージ' (System message) section.

The main area is divided into three panels:

- アシスタントのセットアップ (Assistant Setup):** This panel contains a 'システム メッセージ' (System message) section. It has a dropdown menu set to 'Default' and a text area containing the message: 'あなたはAzure OpenAI Serviceのプロフェッショナルです。日本語で回答してください。' (You are a professional of Azure OpenAI Service. Please answer in Japanese). This text area is highlighted with a red box. Below it is an 'Add an example' button.
- チャットセッション (Chat Session):** This panel shows the chat history. It includes a 'チャットをクリアする' (Clear chat) button, a 'コードの表示' (Show code) button, and a 'Show raw JSON' toggle. The chat history shows a single message from the assistant: 'Azure OpenAI Serviceについて教えてください。' (Tell me about Azure OpenAI Service). The response text is highlighted with a red box. Below the chat history is a text input field with the placeholder 'Type user query here. (Shift + Enter for new line)' and a 'Send' button (a triangle icon). A red arrow points from the 'システム メッセージ' text area to the input field, and another red arrow points from the 'Send' button to the chat history.
- Configuration (Configuration):** This panel contains settings for the chat session. It includes a 'デプロイ' (Deploy) button, a 'パラメーター' (Parameters) dropdown menu set to 'deploy-gpt-4-32k', a 'セッションの設定' (Session settings) section with a '過去のメッセージを含' (Include previous messages) slider set to 10, and a '現在のトークン数' (Current token count) section showing 'Input tokens progress indicator' at 354/32768.

(2) Completions プレイグラウンド

52

- 内部では「Completions API」を使用（現在はリタイア）

Azure AI | Azure OpenAI Studio

Azure OpenAI

プレイグラウンド

チャット

入力候補

DALL-E (Preview)

管理

デプロイ

モデル

Data files

Quotas

Content filters (Preview)

Azure Studio > Completions playground

Completions playground

デプロイ

deploy-gpt-35-turbo-0301

例

クイズを生成する

コードの表示

以下の文章に基づき、多岐選択式のクイズを生成してください。

====

GPT 3.5 モデルを使用すると、自然言語プロンプトに基づいて、自然言語とコード補完を生成できます。特に、GPT-35-turbo モデルは、チャットベースの対話に対して最適化されており、ほとんどの生成 AI シナリオで適切に機能します。

埋め込みモデルは、テキストを数値ベクトルに変換し、テキストソースを比較して類似点を探すなどの言語分析シナリオで役立ちます。

DALL-E モデルは、自然言語プロンプトに基づいて画像を生成するために使用されます。現在、DALL-E モデルはプレビュー段階です。DALL-E モデルは Azure OpenAI Studio インターフェイスに一覧表示されないため、明示的にデプロイする必要があります。

====

例:

Q1: 自然言語の生成を利用するにはどのモデルを使用しますか？

A: GPT 3.5

B: 埋め込み

C: DALL-E

Q2: GPT-35-turbo モデルが適している用途は何ですか？

A: チャットベースの対話

B: 大量のコード補完

C: 画像の生成

Q3: 埋め込みモデルはどのように役立ちますか？

A: 言語分析シナリオでのテキストソースの比較

B: アプリケーションの展開

C: テキストの構造解析

Q4: DALL-E モデルはどのように活用されるでしょうか？

A: 自然言語プロンプトに基づいた画像の生成

B: 自然言語プロンプトに基づいた音声の生成

C: 自然言語プロンプトに基づいた文章の生成

生成

元に戻す

再生成

トークン: 848

Parameters

温度 ①

0.8

最大長 (トークン) ①

500

シーケンスの停止 ①

シーケンスの停止

上位の確率 ①

1

頻度のペナルティ ①

0

ブレゼンス ペナルティ ①

0.5

応答前のテキスト ①

☐ テキストの入力

応答後のテキスト ①

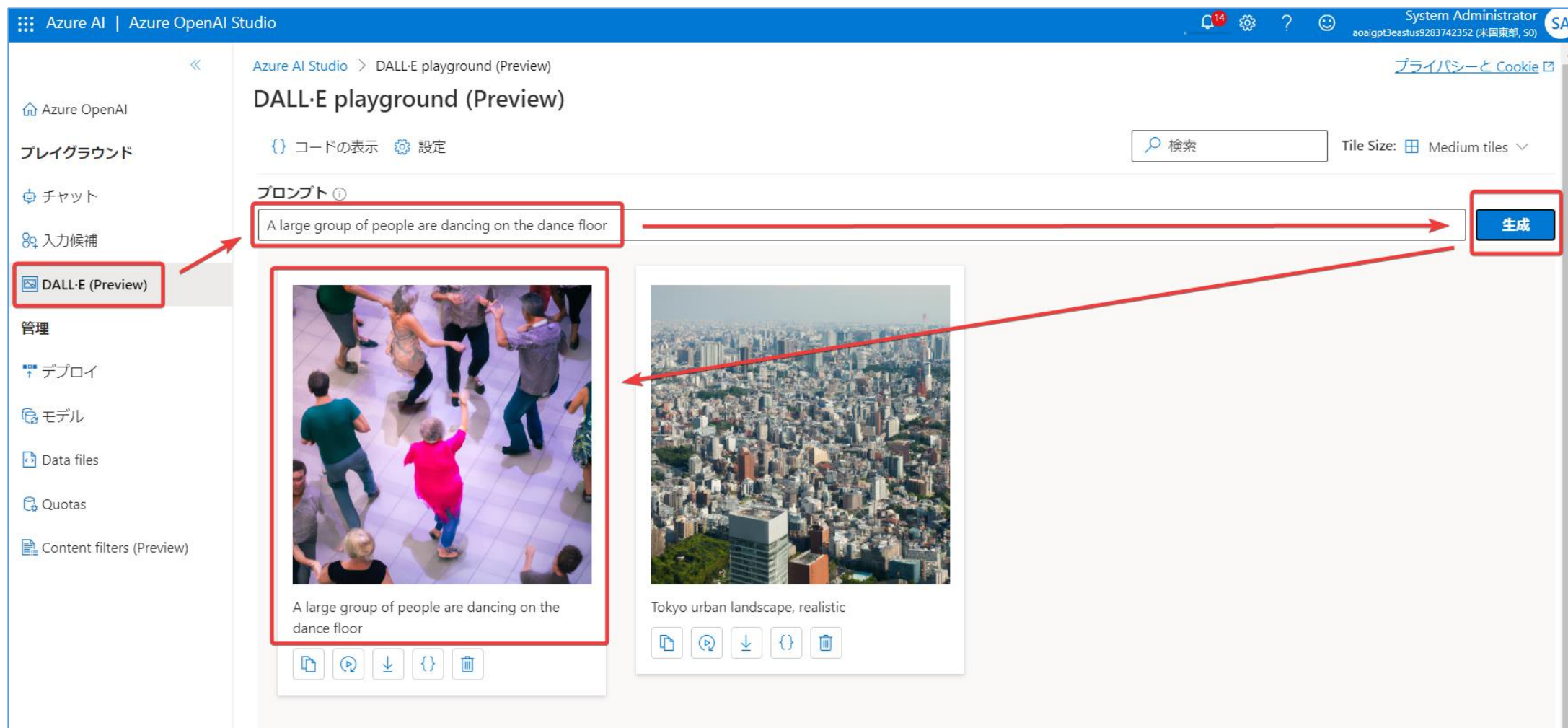
☐ テキストの入力

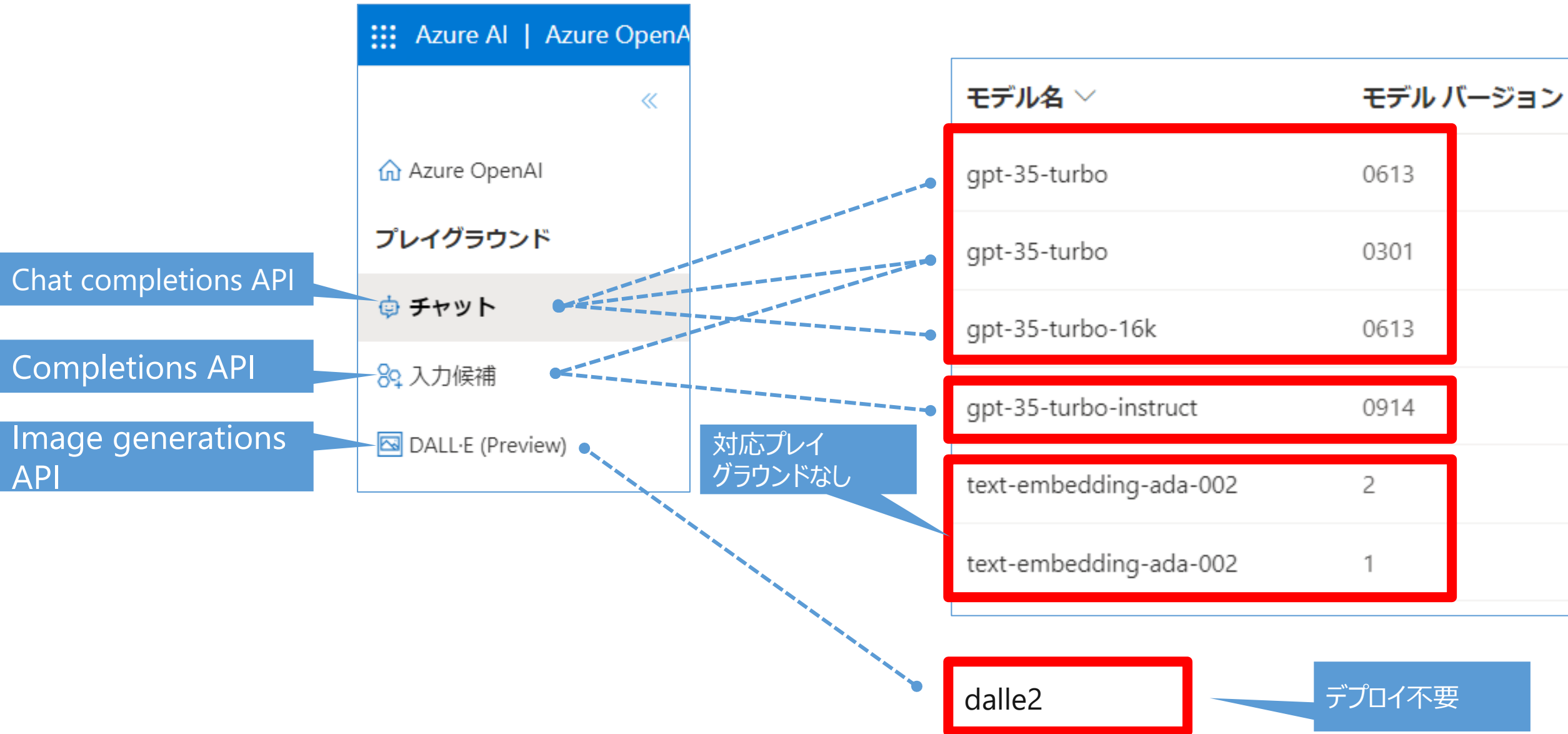
詳細情報

(3) DALL-E プレイグラウンド

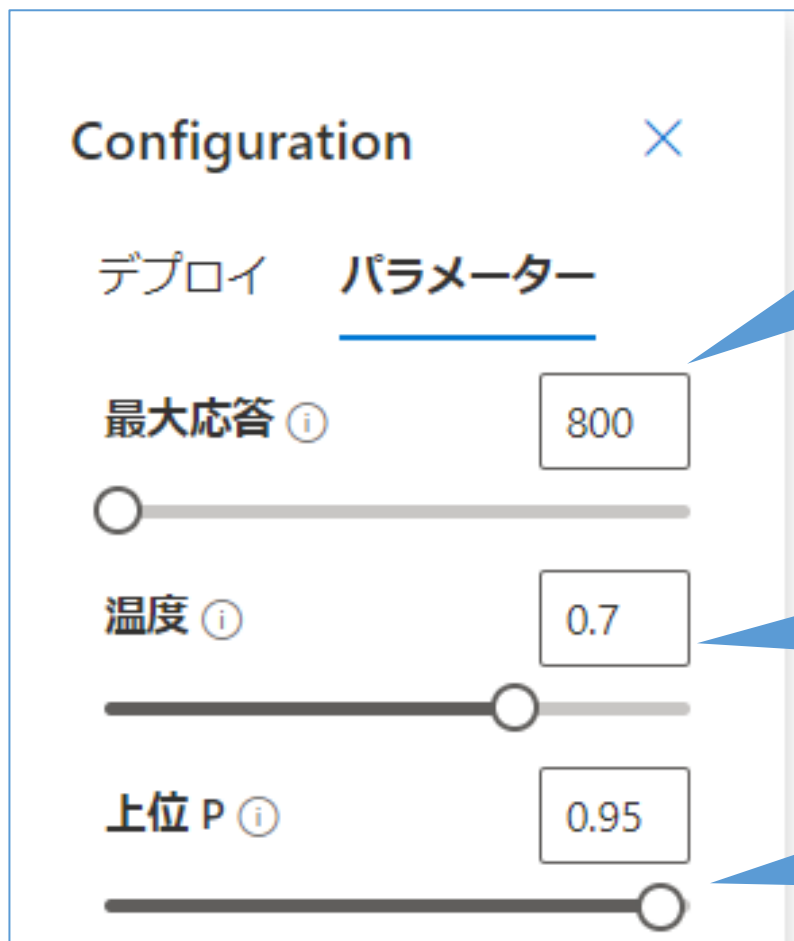
53

- 内部では「Image generation API」を使用





- 2020/6 に Completions API を導入。
- 2023/3 に Chat Completions APIを導入。
 - これまでのユースケースの大部分が Chat Completions API で実行できる。
 - 2023/7時点で、GPT の API 利用の 97% が Chat Completions API を使用。
 - システムメッセージとユーザーメッセージを分離でき、プロンプトインジェクション攻撃のリスクを減らせる。
- **今後はユーザーが Chat Completions API を採用することを推奨。**
- 2024/1 に、**Completions API 用の古いモデル**は使用できなくなる。
 - ※ 古いモデルの代替として、Completions API で使用できる gpt-3.5-turbo-instruct モデルが提供されている。
 - ※ **Completions API** が非推奨 / 廃止となるかどうかの言及はない。
 - ※ Azure OpenAI Service側でも、Completions API 用の古いモデルは「レガシ モデル」となり、新規デプロイはできなくなっている。



The image shows a 'Configuration' dialog box with a close button (X) in the top right corner. It has two tabs: 'デプロイ' (Deploy) and 'パラメーター' (Parameters), with the latter being selected. There are three parameter settings, each with a label, an information icon (i), a slider, and a numeric input field:

- 最大応答 (Maximum Response):** The slider is at the far left, and the input field shows '800'.
- 温度 (Temperature):** The slider is positioned in the middle, and the input field shows '0.7'.
- 上位 P (Top P):** The slider is positioned towards the right, and the input field shows '0.95'.

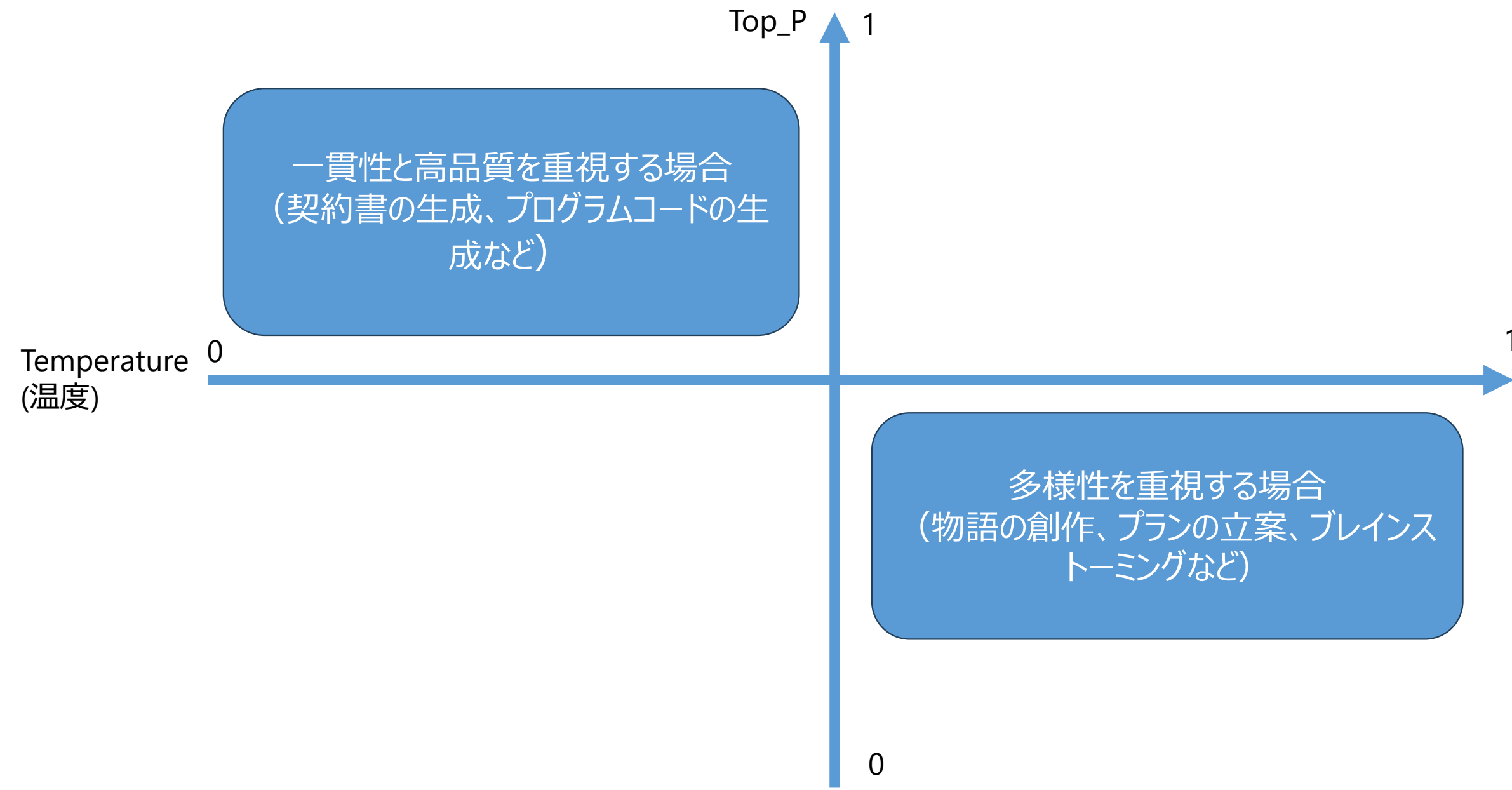
応答あたりのトークン数に制限を設定。
数字を大きくすると、より長い回答が得られる。
数字を小さくすると、生成された回答が長い場合に途中で出力が強制的に打ち切られる。
(打ち切られないようにするにはプロンプトで「100文字以内で出力」などと指示)

Temperature (0～1): ランダム性を制御。数値を小さくすると、ランダムの度合いが下がり、同じ回答を返しやすくなる。数値を大きくすると、ランダムの度合いが上がり、リクエストのたびに異なる生成結果が得られる (いろいろなストーリーを生成する)

Top_probability (0～1): 数値を小さくすると、モデルのトークンの選択がより可能性が高いトークンに絞られる。数値を大きくすると、確率が高いトークンと低いトークンの両方からモデルが選択できるようになる。(言葉選びが自由になる)

- TemperatureとTop-Pをチューニングする方法は、目的によって変わります。
- **一貫性と高品質を重視するなら、低いTemperatureと高いTop-Pを試すことが有効**です。
- **一方、多様性を重視するなら、高いTemperatureと低いTop-Pを試すことが有効**です。

ただし、TemperatureとTop-Pは交互に影響し合うため、**最適な値を見つけるには多くの試行錯誤が必要**です。



System message ⓘ

あなたは日本の昔の物語を創作する作家です。

「温度」と
「上位P」を変化させて
ストーリーの変化を確認

Configuration ×

デプロイ パラメーター

最大応答 ⓘ

800



温度 ⓘ

0.7



上位 P ⓘ

0.95



- Temperature=**0**、Top P = 0.95 の場合**ワンパターン**な作品案を生成

※個人的な主観です

Chat session

 Clear chat  Playground Settings  View code  Show raw JSON 

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。



1. "月夜の狐"という物語は、一人の若者が狐に化かされ、その後、狐との奇妙な友情を育む物語です。
2. 若者は狐の助けを借りて村を救い、その過程で自分自身の勇氣と知恵を発見します。
3. 最終的に、若者と狐は互いの違いを超えた深い絆を築き、村人たちにもその絆を理解させることで、村全体が一つになる物語です。

作品案「月夜の狐」を生成

Chat session

 Clear chat  Playground Settings  View code  Show raw JSON 

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。



1. "月夜の狐"という物語は、人間に化けることができる狐のキツネが主人公で、人間界での生活と恋愛を描く。
2. キツネは人間の女性に恋をし、その愛を守るために様々な困難に立ち向かう。
3. しかし、最終的には自分の本当の姿を受け入れ、人間と狐の世界をつなぐ架け橋となる。

何度生成しなおしても「月夜の狐」が出てくる（ストーリーは若干変化）

- Temperature=0、Top P = 0 の場合、**ワンパターンだが、意外性のある作品**案を生成

※個人的な主観です

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。



1. "月夜の狐"という物語は、人間に化けることができる狐のキツネが主人公で、人間との共生を模索する物語です。
2. キツネは、人間の女性と恋に落ち、その愛を守るために様々な困難に立ち向かいます。
3. しかし、最終的には自分の本当の姿を受け入れ、人間と狐との間に生まれた子供たちと共に新たな生活を始めることを決意します。

人間と狐の間に生まれた子供たち！？（*°▽°）

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。



1. "月夜の狐"という物語は、一人の若者が狐に化かされるが、その狐との交流を通じて人間の愛と善意を学び、成長していく物語である。
2. 若者は狐に化かされてしまうが、狐が人間に化けていた理由や、狐が抱えていた悲しみを知り、互いに理解し合うことで友情が芽生える。
3. 最終的には、若者が狐のために自己犠牲をすることで、狐は人間の善意を理解し、二人は互いに敬意を持って別れる。

ちょっと悲しい結末になってしまう場合も(ノД`)

- Temperature=**1**、Top P = 0 の場合**バリエーション豊かな**作品案を生成

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。箇条書きの1行目にはタイトルも含めてください。



- タイトル：「月夜の狐と賢者の笛」
- ある村に伝わる不思議な笛の音色が、月夜にだけ聞こえるという話。
- 狐が人間に化けて賢者に弟子入りし、笛の真の力を探る冒険が始まる。

「月夜の狐」シリーズの
新たな章がスタート・・・！？

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。箇条書きの1行目にはタイトルも含めてください。



- タイトル：「星を継ぐ者と竜の神託」
- 星の位置によって未来を予言する能力を持つ孤児の少年が、天に舞う竜から神託を受ける。
- 少年は村を襲う災厄を防ぐため、竜の試練を乗り越えて真の予言者へと成長する物語。

「星を継ぐ者と竜の神託」
SFファンタジーっぽい
新しいストーリーができました

- Temperature=**1**、Top P = 0.95 の場合**様々なモチーフ**の作品案を生成

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。



1. 「闇を照らす月の琴」は、貧しい木こりが発見した魔法の琴を巡る冒険と愛の物語である。
2. 主人公の木こりが繰り広げる冒険を通じて、心の豊かさや人間の絆、自己犠牲の大切さを描いている。
3. さらに、美しい月の描写や琴の音色を通じて、読者に日本の自然美や伝統文化への鑑賞を喚起する。

++ 闇を照らす月の琴 ++
「月夜の狐」ではなくなった。

新しい日本の昔話の作品を考えます。3行の箇条書きで作品の概要を出力してください。



1. "藍色の龍と女君主" - 平安時代に、優れた智恵と勇気を持つ女性君主が謎の病で苦しむ村人たちを救う物語。
2. 主人公は、神秘的な藍色の龍から秘密の治療法を得るため、数々の困難に立ち向かう。
3. 物語は、主人公の決断力と助けを求める心が、人々を結束させ、ついに村を救う力になることを描く。

「藍色の龍と女君主」
これも「月夜の狐」ではなくなった。

※個人的な主観です

Top_P ↑ 1

ワンパターンな
ストーリー生成、かつ
意外性があまりない
オーソドックスな展開

多様なパターンな
ストーリーが生成されるが
意外性があまりない
オーソドックスな展開

Temperature 0
(温度)

ワンパターンな
ストーリー生成だが
生成されたそれぞれの
ストーリーに意外性がある

多様なパターンの
ストーリーが生成され、かつ
生成されたそれぞれの
ストーリーに意外性がある

作品案生成といったケースでは
Temperature高め、
Top_P低めがよいと思われる

0

文章の生成や要約、質問への回答などの目的でデプロイされる Azure OpenAI モデルは何ですか?

- a. gpt-35-turbo-instruct
- b. gpt-35-turbo
- c. text-embedding-ada-002

文章の生成や要約、質問への回答などの目的でデプロイされる Azure OpenAI モデルは何ですか?

a. gpt-35-turbo-instruct

コード生成用のモデル。

b. gpt-35-turbo

正解。チャット用のモデル

c. text-embedding-ada-002

埋め込み (Embedding) 用のモデル。

返される回答のランダム性またはクリエイティブ性を変更するには、どのパラメーターを調整すればよいですか？

a. Temperature

b. 最大長(トークン)

返される回答のランダム性またはクリエイティブ性を変更するには、どのパラメーターを調整すればよいですか？

a. Temperature

正解。回答のランダム性の度合いを設定

b. 最大長(トークン)

テキスト（トークン）の長さを制限。

ユーザーによる質問の入力と、チャットボットからの回答の出力を繰り返す「会話イン、メッセージアウト」シナリオをサポートする Azure OpenAI Studio プレイグラウンドはどれですか？

- a. Completions（入力候補）プレイグラウンド
- b. Chat プレイグラウンド
- c. DALL-E プレイグラウンド

ユーザーによる質問の入力と、チャットボットからの回答の出力を繰り返す「会話イン、メッセージアウト」シナリオをサポートする Azure OpenAI Studio プレイグラウンドはどれですか？

a. Completions（入力候補）プレイグラウンド

テキストイン・テキストアウト（文章の続きを生成する）

b. Chat プレイグラウンド

正解。対話的なシナリオをサポート。

c. DALL-E プレイグラウンド

テキストイン・画像アウト
（文章に従って画像を生成する）