

モジュール6 Azure OpenAI on your data



800 XP

Azure OpenAI Service で独自のデータを使用する

6分 残り • モジュール • 4/7 ユニットが完了しました

[フィードバック](#)

中級 AI エンジニア 開発者 ソリューション アーキテクト 学生 Azure OpenAI Service Azure

開発者は、Azure OpenAI で独自のデータを使用し、特定のデータ ソースを参照して応答を得ることができる、サポートされている AI チャット モデルを使用することができます。

学習の目的

このモジュールを完了すると、次のことができるようになります。

- データに対する Azure OpenAI の機能について説明する
- Azure OpenAI を独自のデータを使用するように構成する
- Azure OpenAI API を使用し、独自のデータに基づいて応答を生成する

[続行 >](#) [追加](#)

モジュール6

- Azure OpenAI **on your data**とは？
- Azure OpenAI **on your data**の特徴
- Azure OpenAI **on your data**の動作
- Azure OpenAI **on your data**のコスト
- 知識チェック
- 講師デモ


モジュール6

- Azure OpenAI **on your data**とは？
- Azure OpenAI **on your data**の特徴
- Azure OpenAI **on your data**の動作
- Azure OpenAI **on your data**のコスト
- 知識チェック
- 講師デモ




 Azure OpenAI

Playground

 Chat

 Completions

 DALL-E (Preview)

Management

 Deployments


 Models

 Data files

Chat playground

Assistant setup



 System message

 Add your data (preview)

Ask questions about your own data. Your data is stored securely in your Azure subscription. [Learn more about how your data is protected.](#)

 Add a data source



🏠 Azure OpenAI

プレイグラウンド

🗣️ チャット

🔗 入力候補

🖼️ DALL-E (プレビュー)

管理

🚀 デプロイ

📁 モデル

📄 データ ファイル

Azure AI Studio > チャット プレイグラウンド

チャット プレイグラウンド

アシスタントのセットアップ

🔄 システム メッセージ

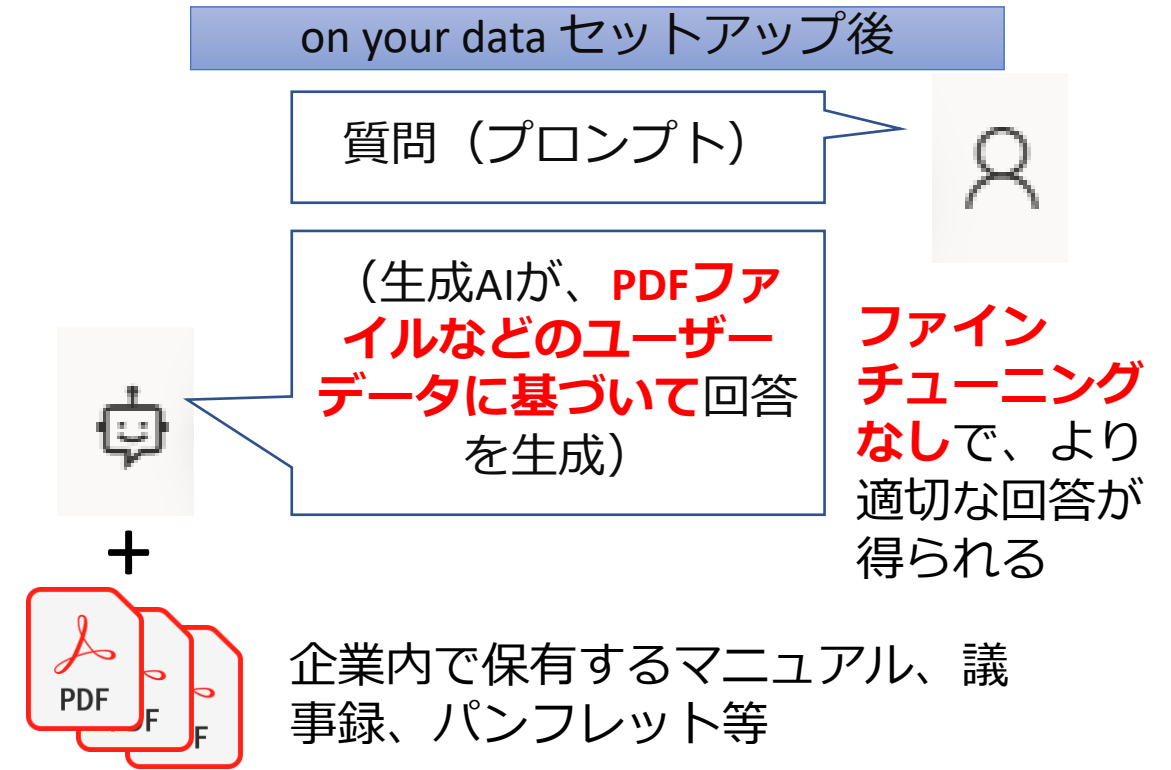
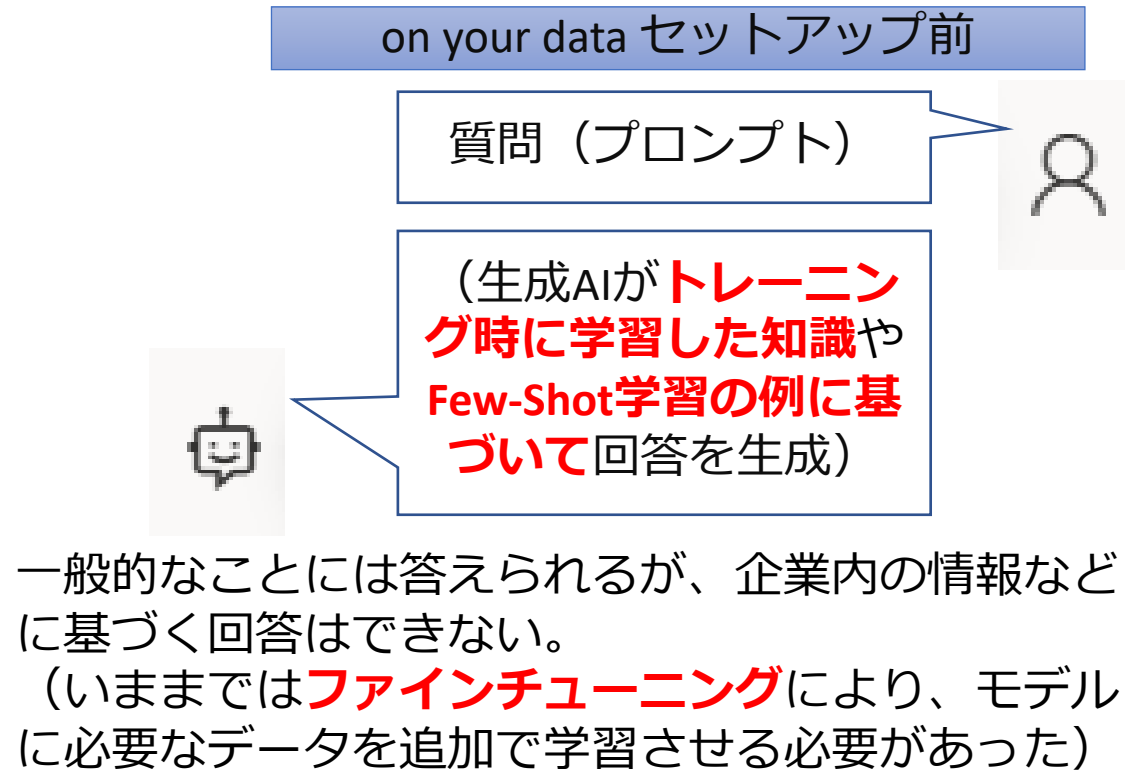
📄 データの追加 (プレビュー)

自分のデータについて質問します。データは Azure サブスクリプションに
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

Azure OpenAI on your dataとは？

- 生成モデル（gpt-35-turbo / gpt-4等）が、**ユーザーが与えたデータに基づいて回答を生成できるようにする**仕組み
- 2023/6/27 パブリックプレビュー



Azure OpenAI on your dataとは？

- データ形式として.txt / .md / .html / Word / PowerPoint / PDFに対応
- データは**ストレージアカウント**に記録される※
 - ※データソースで「ファイルのアップロード」を指定した場合。
 - 既存のAzure AI Search、Azure Blob Storage、Azure Cosmos DB for MongoDB vCoreなどをデータソースとして利用することも可能
- データの追加・変更・削除も可能（インデックスの手動 / 自動更新が可能）
- 内部的には**Azure AI Search**を使用して高速な検索を実現している
- **データの追加はAzure OpenAI Studioから行う**ことが推奨されている
 - 投入したデータが適切な大きさのかたまり（チャンク）に分割されて、インデックスが作成されるため。

[Azure OpenAI Service で独自のデータを使用する - Azure OpenAI | Microsoft Learn](#)

<https://learn.microsoft.com/ja-jp/azure/ai-services/openai/use-your-data-quickstart?tabs=command-line>

[Public Preview: Azure OpenAI Service On Your Data | Azure の更新情報 | Microsoft Azure](#)

モジュール6

- Azure OpenAI **on your data**とは？
- Azure OpenAI **on your data**の特徴
- Azure OpenAI **on your data**の動作
- Azure OpenAI **on your data**のコスト
- 知識チェック
- 講師デモ

Azure OpenAI on your dataの特徴

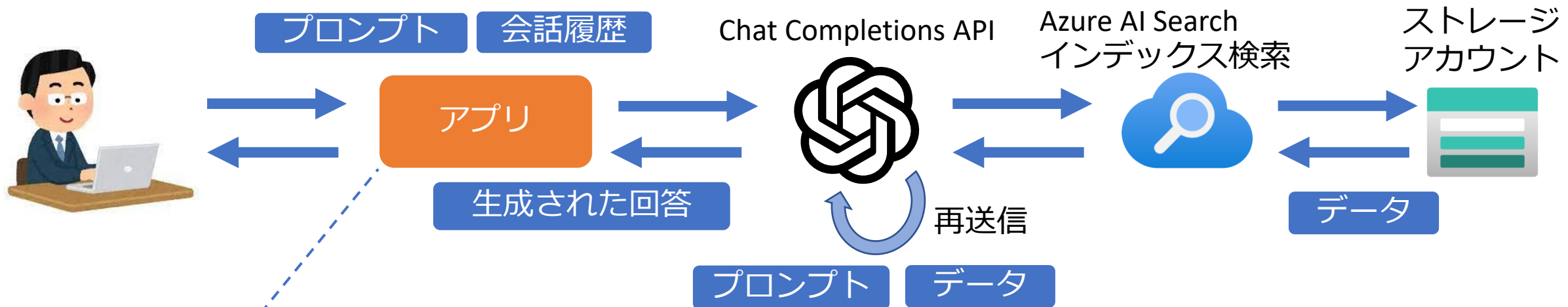
- モデルの**ファインチューニング**をせずに、**ユーザーが提供する独自のデータ**に基づくコンテンツ生成が可能となる
 - ファインチューニングの手間とコストが削減できる
 - ファインチューニングを行うには、トレーニング用のデータの準備とトレーニングが必要。時間と手間とトレーニングコストがかかる
 - ファインチューニングによって作られたモデル（カスタムモデル）の利用には追加の運用コストがかかる
- Azure OpenAI Studioから簡単に設定できる
- gpt-35-turbo / gpt-4（Chat Completions API）で可以使用
 - 使い方は普通のChat Completions API（Chatプレイグラウンド）とほぼ同じ

モジュール6

- Azure OpenAI **on your data**とは？
- Azure OpenAI **on your data**の特徴
- Azure OpenAI **on your data**の動作
- Azure OpenAI **on your data**のコスト
- 知識チェック
- 講師デモ

Azure OpenAI on your dataの動作

- データはストレージアカウントに格納され、Azure AI Searchを使用して、データからインデックス（索引）が作られる
- インデックスを使用して、プロンプトに関係のあるデータが高速に検索され、取得される
- データの検索・取得は Chat Completions API 側で自動的に実行される
- ユーザーの**プロンプト**と**会話履歴**から、適切なデータソースが選択され、データが取得される。



Azure OpenAI StudioのChatプレイグラウンド
または独自のアプリ / システム

- 取得した**データ**が追加された**プロンプト**が生成モデルに**再送信**される

モジュール6

- Azure OpenAI **on your data**とは？
- Azure OpenAI **on your data**の特徴
- Azure OpenAI **on your data**の動作
- Azure OpenAI **on your data**のコスト
- 知識チェック
- 講師デモ

コスト

- **※注意: 現在on your dataはプレビュー段階のため、一般提供開始時にコストについて新情報が出てくる可能性がある。**
- on your data自体の利用にはコストはかからない
- 内部的に、ストレージアカウントとAzure AI Searchを使用するので、その部分のコストは追加で発生する。
- 内部的に、ユーザーのプロンプトに、取得したデータが追加され、モデルに送信されるため、それによる入力・出力トークン数の変動があり、そのためのコストの増減がありうる。

モジュール6

- Azure OpenAI **on your data**とは？
- Azure OpenAI **on your data**の特徴
- Azure OpenAI **on your data**の動作
- Azure OpenAI **on your data**のコスト
- 知識チェック
- 講師デモ

知識チェック 第1問

Azure OpenAI on your data を使用すると、開発者は何を行うことができますか?

- a. 独自の AI チャット モデルを作成する
- b. 承認されたサブスクリプションなしで Azure OpenAI にアクセスする
- c. ユーザーが与えた任意のデータに基づいて回答を生成できる

知識チェック 第1問

Azure OpenAI on your data を使用すると、開発者は何を行うことができますか?

a. 独自の AI チャット モデルを作成する

不正解。独自のモデル作成には「ファインチューニング」などを利用する。

b. 承認されたサブスクリプションなしで Azure OpenAI にアクセスする

不正解。Azure OpenAI Serviceの利用前に申請を行い、承認を得る必要がある。

c. ユーザーが与えた任意のデータに基づいて回答を生成できる

正解。PDF、テキストファイル、Word、PowerPoint等のデータに基づいて回答を生成できる。

知識チェック 第2問

Azure OpenAI on your data を使用する場合、データを追加するための推奨される方法は何ですか?

a. Azure OpenAI on your data で利用できる任意のデータ ソース オプションを使用する。

b. Azure OpenAI Studio を使用して検索リソースとインデックスを作成する。

c. Azure OpenAI Studio を使用せずにストレージ アカウント内のファイルに接続する。

知識チェック 第2問

Azure OpenAI on your data を使用する場合、データを追加するための推奨される方法は何ですか?

a. Azure OpenAI on your data で利用できる任意のデータ ソース オプションを使用する。 不正解。入力データが分解されない。

b. Azure OpenAI Studio を使用して検索リソースとインデックスを作成する。

正解。Azure OpenAI Studioを使用すると、入力データが適切な単位（最大1024トークンの「チャンク」）に分解されてから、各チャンクごとにインデックスが作成される。

c. Azure OpenAI Studio を使用せずにストレージ アカウント内のファイルに接続する。 不正解。入力データが分解されない。

知識チェック 第3問

Azure OpenAI on your data を使用する場合に推奨されるプロンプト エンジニアリング手法は何ですか?

- a. 短いプロンプトと、必要最小限の会話履歴を含める。
- b. できるだけ多くの会話履歴を含める。
- c. 1つの長いプロンプトを使用して、必要なすべての情報を提供する。

知識チェック 第3問

Azure OpenAI on your data を使用する場合に推奨されるプロンプト エンジニアリング手法は何ですか?

a. 短いプロンプトと、必要最小限の会話履歴を含める。

正解。プロンプトと会話履歴に基づいてデータが取得される

b. できるだけ多くの会話履歴を含める。

効率的ではない

c. 1つの長いプロンプトを使用して、必要なすべての情報を提供する。

効率的ではない

モジュール6

- Azure OpenAI **on your data**とは？
- Azure OpenAI **on your data**の特徴
- Azure OpenAI **on your data**の動作
- Azure OpenAI **on your data**のコスト
- 知識チェック
- 講師デモ

■まず on your data を設定していない状態で、「ニューヨークへの旅行における宿泊先」についてAIアシスタントに質問してみる。AIモデルが学習済みの知識に基づいた回答が得られる（が、具体的な宿泊先情報は得られない）。

チャットセッション

✓ チャットをク
リアする ⚙️ プレイグラウ
ンド設定 📄 コード
の表示 🔍 未加工の JSON
を表示する

I'd like to take a trip to New York. Where should I stay?



New York City offers a wide range of accommodation options to suit different preferences and budgets. Here are a few popular areas to consider when choosing where to stay in New York:

1. Midtown Manhattan: This area is a central location, close to iconic attractions like Times Square, Empire State Building, and Broadway theaters. It offers a variety of hotels, from luxury to more budget-friendly options.
2. Lower Manhattan: If you prefer a more historic and vibrant neighborhood, consider staying in Lower Manhattan. This area includes the Financial District, Wall Street, One World Trade Center, and is close to the Statue of Liberty and Ellis Island.
3. Chelsea: Known for its art galleries, Chelsea is a trendy neighborhood with a vibrant nightlife scene. It offers a mix of boutique hotels and upscale accommodations.
4. Brooklyn: If you're looking for a more relaxed atmosphere, consider staying in one of Brooklyn's neighborhoods, such as Williamsburg or DUMBO. These areas offer a diverse range of hotels, trendy restaurants, and great views of Manhattan.
5. Upper West Side/Upper East Side: These residential neighborhoods offer a more tranquil atmosphere while still being well-connected to the city's attractions. They are home to iconic landmarks like Central Park, the Museum of Natural History, and the Metropolitan Museum of Art.

When choosing your accommodation, consider factors such as proximity to attractions, transportation options, safety, and your personal preferences. It's also recommended to check reviews and compare prices to find the best option for your trip.

ニューヨークへ旅行に行きたいです。どこに泊まればいいですか？

ニューヨーク市では、さまざまな好みや予算に合わせて幅広い宿泊施設のオプションを提供しています。ニューヨークでの滞在先を選ぶ際に考慮すべき人気エリアをいくつかご紹介します。

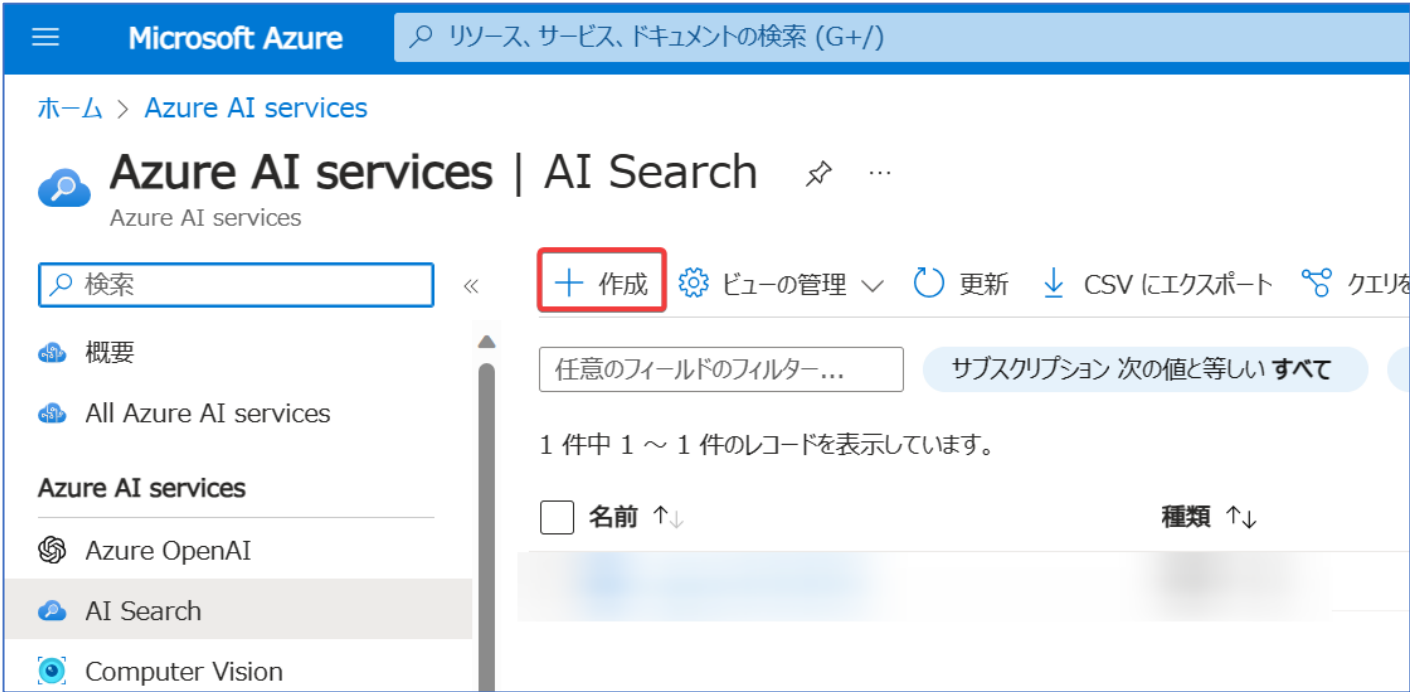
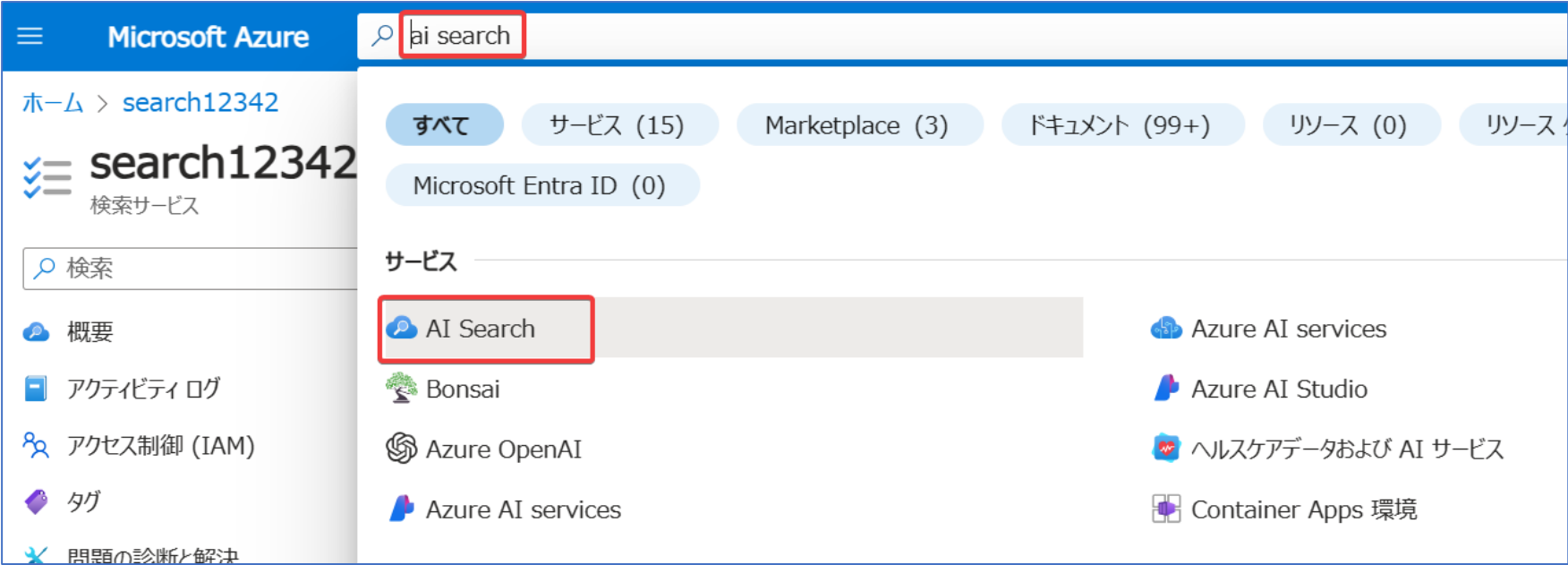
マンハッタンのミッドタウン: このエリアは中心部に位置し、タイムズスクエア、エンパイアステートビルディング、ブロードウェイ劇場などの象徴的な観光スポットに近いです。高級ホテルからリーズナブルなホテルまで、さまざまなホテルが揃っています。

ロウワー マンハッタン: より歴史的で活気のある地区をお好みの場合は、ロウワー マンハッタンに滞在することを検討してください。このエリアには金融街、ウォール街、ワンワールドトレードセンターが含まれており、自由の女神やエリス島にも近いです。

...

宿泊施設を選ぶときは、観光スポットへの近さ、交通手段、安全性、個人的な好みなどの要素を考慮してください。口コミを確認し、料金を比較して、旅行に最適なオプションを見つけることもお勧めします。

■ Azure AI Searchのリソースの作成



■ Azure AI Searchの名前などを指定して作成

Microsoft Azure

リソース、サービス、ドキュメントの検索 (G+)

2

ホーム > Azure AI services | AI Search >

検索サービスを作成する ...

基本

スケール

ネットワーク

タグ

確認および作成

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 *

instances details

サービス名 * ⓘ

場所 *

価格レベル * ⓘ

aoairg

新規作成

cognitivesearch293747234

West US 2

Standard

25 GB/パーティション*、最大 12 個のレプリカ、最大 12 個のパーティション、最大 36 個の検索ユニット


価格レベルの変更


確認および作成


前へ


次: スケール


■ストレージアカウントの作成





 リソースの作成


 ホーム


 ダッシュボード


 すべてのサービス


 お気に入り


 すべてのリソース


 リソース グループ


 Azure OpenAI


 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**

 仮想ネットワーク

 Microsoft Entra ID


 モニター


 **Microsoft Azure**


ホーム >


ストレージ アカウント


Contoso

 **作成**

 復元

 ビューの管理

 更新

 CSV にエクスポート

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**

4 件中 1 ~ 4 件のレコードを表示しています。

■ストレージアカウントの名前などを指定

ホーム > ストレージ アカウント >

ストレージ アカウントを作成する ...

基本

詳細設定

ネットワーク

データ保護

暗号化

タグ

レビュー

Azure Storage は、高可用性、セキュリティ、耐久性、スケーラビリティ、冗長性を備えたクラウド ストレージを提供する Microsoft が管理するサービスです。Azure Storage には、Azure BLOB (オブジェクト)、Azure Data Lake Storage Gen2、Azure Files、Azure Queues、Azure Tables が含まれます。ストレージ アカウントのコストは、使用量と、下で選ぶオプションに応じて決まります。[Azure ストレージ アカウントの詳細](#)

プロジェクトの詳細

新しいストレージ アカウントを作成するサブスクリプションを選択します。ストレージ アカウントを他のリソースと一緒に整理して管理するには、新規または既存のリソース グループを選択します。

サブスクリプション *

リソース グループ *

aoairg

[新規作成](#)

インスタンスの詳細

ストレージ アカウント名 ⓘ *

st2938742242

地域 ⓘ *

(Canada) Canada East

[エッジ ゾーンにデプロイ](#)

パフォーマンス ⓘ *

☒ Standard: ほとんどのシナリオに対して推奨される (汎用 v2 アカウント)

☐ Premium: 低遅延が必要なシナリオにお勧めします。

レビュー

< 前へ

次へ: 詳細設定 >

■ ストレージアカウント内にBlobコンテナを作成

Microsoft Azure

リソース、サービス、ドキュメントの検索 (G+/)

ホーム > ストレージ アカウント > st293874297342

st293874297342 | コンテナ

☆ ☆ ...

ストレージ アカウント

検索

「+」 コンテナ

アクセス レベルを変更します

コンテナを復元する

更新

概要

アクティビティ ログ

タグ

問題の診断と解決

アクセス制御 (IAM)

データ移行

イベント

ストレージ ブラウザー

Storage Mover

データ ストレージ

コンテナ

ファイル共有

プレフィックスによるコンテナの検索

名前	最終変更日時
<input type="checkbox"/> \$logs	2023/11/14 21:55:03
<input type="checkbox"/> margiestravel	2023/11/14 22:00:18

この例では「マーギーズトラベル」社（架空の会社）の旅行パンフレットを格納することを想定

■ Chatプレイグラウンドの「アシスタントの追加」で「データの追加」、「+データソースの追加」をクリック



● データソース

○ データ管理

○ レビューして終了

データソースを選択または追加する

データソースは、生成された結果をデータに基にするために使用されます。基になるデータの構築元となるソースとして、既存のデータソースを選択するか、Azure Blob Storage、データベース、またはローカルファイルを使用して新しいデータ接続を作成します。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。

[Azure AI のデータプライバシーとセキュリティの詳細をご覧ください。](#)

データソースを選択する *

Upload files

サブスクリプション *

ME-MngEnvMCAP347065-hiryamada-1

Azure Blob ストレージ リソースの選択 ⓘ *

st2983742

[新しい Azure Blob ストレージ リソースを作成する](#)

✔ このリソースでは、クロス オリジン リソース共有 (CORS) が有効になっています。

Azure AI Search リソースを選択する ⓘ *

search12342

[新しい Azure AI Search リソースを作成する](#)

インデックス名を入力してください ⓘ *

index1

☒ ベクトル検索をこの検索リソースに追加します。

モデルの埋め込み

データの一部としてベクトル モデルを使用するには、以下のいずれかを選択してください。

埋め込みモデルを選択する ⓘ *

Azure OpenAI - text-embedding-ada-002

☒ Azure AI Search アカウントに接続すると、アカウントが使用されるようになることに同意します。 [価格の表示](#)

この例では
ファイルのアップロードを
使用

データ格納先の
ストレージアカウント

Azure AI Search リソース

Azure AI Search の
インデックス名

ベクトル検索を
使用する場合はチェック

ベクトルを計算するための
埋め込みモデルを選択

次へ

キャンセル

✓ データソース

✓ データ管理

✓ レビューして終了

ファイルのアップロード

追加するファイルを選択します。ファイルは Azure Blob Storage に格納され、前の手順で作成または選択した Cognitive Search リソースによってインデックスが作成されます。

[Azure AI のデータプライバシーとセキュリティの詳細をご覧ください。](#)

ファイル名	種類	サイズ	状態
-------	----	-----	----



ドラッグアンドドロップ。

または

[ファイルを参照する](#)

(.txt, .md, .html, .pdf, .docx, .pptx)

16 MB のサイズ制限

ファイルのアップロード

戻る

次へ

キャンセル

PDFファイルを
アップロード


■アップロードするPDFの例（Azure AI Searchは内容を自動的にいくつかの「チャンク」（塊）に分割する）


Margie's Travel Presents...

New York

The City of New York, often called New York City or simply New York, is the most populous city in the United States. With an estimated 2015 population of 8,550,405 distributed over a land area of about 302.6 square miles, New York City is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, one of the most populous urban agglomerations in the world. A global power city, New York City exerts a significant impact upon commerce, finance, media, art, fashion, research, technology, education, and entertainment, its fast pace defining the term New York minute. Home to the headquarters of the United Nations, New York is an important center for international diplomacy and has been described as the cultural and financial capital of the world.

The Statue of Liberty





The Empire State Building

New York Hotels

Margie's Travel offers the following accommodation options in New York:

The Manhattan Hotel
Stay in the heart of New York City, within an easy walk to Times Square and Broadway.

The Grand Central Hotel
Comfortable mid-town hotel. Close to Grand Central Station, the Chrysler Building, and the Empire State Building.

The Park Hotel
Luxurious accommodation in upper Manhattan, with views of Central Park.

ニューヨーク市は、米国で最も人口の多い都市です。ニューヨーク市は、2015 年の推定人口が 8,550,405 人で、約 302.6 平方マイルの土地に分布しており、米国で最も人口密度の高い主要都市でもあります。ニューヨーク州の南端に位置するこの都市は、世界で最も人口の多い都市集積地の 1 つであるニューヨーク都市圏の中心です・・・

ニューヨークのホテル
マーギーズ トラベルでは、ニューヨークで以下の宿泊施設のオプションを提供しています：

ザ マンハッタン ホテル
ニューヨーク市の中心部に位置し、タイムズスクエアやブロードウェイまで徒歩圏内にあります。

ザ グランド セントラル ホテル
快適なミッドタウンホテル。グランドセントラル駅、クライスラービル、・・・

■ 設定内容を確認して、「保存して閉じる」をクリック。

データの追加

データソース

レビューして終了

レビューして終了

レビューして終了

データに設定した構成を確認する

[Azure AI のデータプライバシーとセキュリティの詳細をご覧ください。](#)

データソース

Azure Blob Storage

ストレージリソース

st293874297342

ストレージコンテナー

margiestravel

Azure Cognitive Search のリソース

cogsearch9238742

インデックス名

margiestravel3

検索の種類

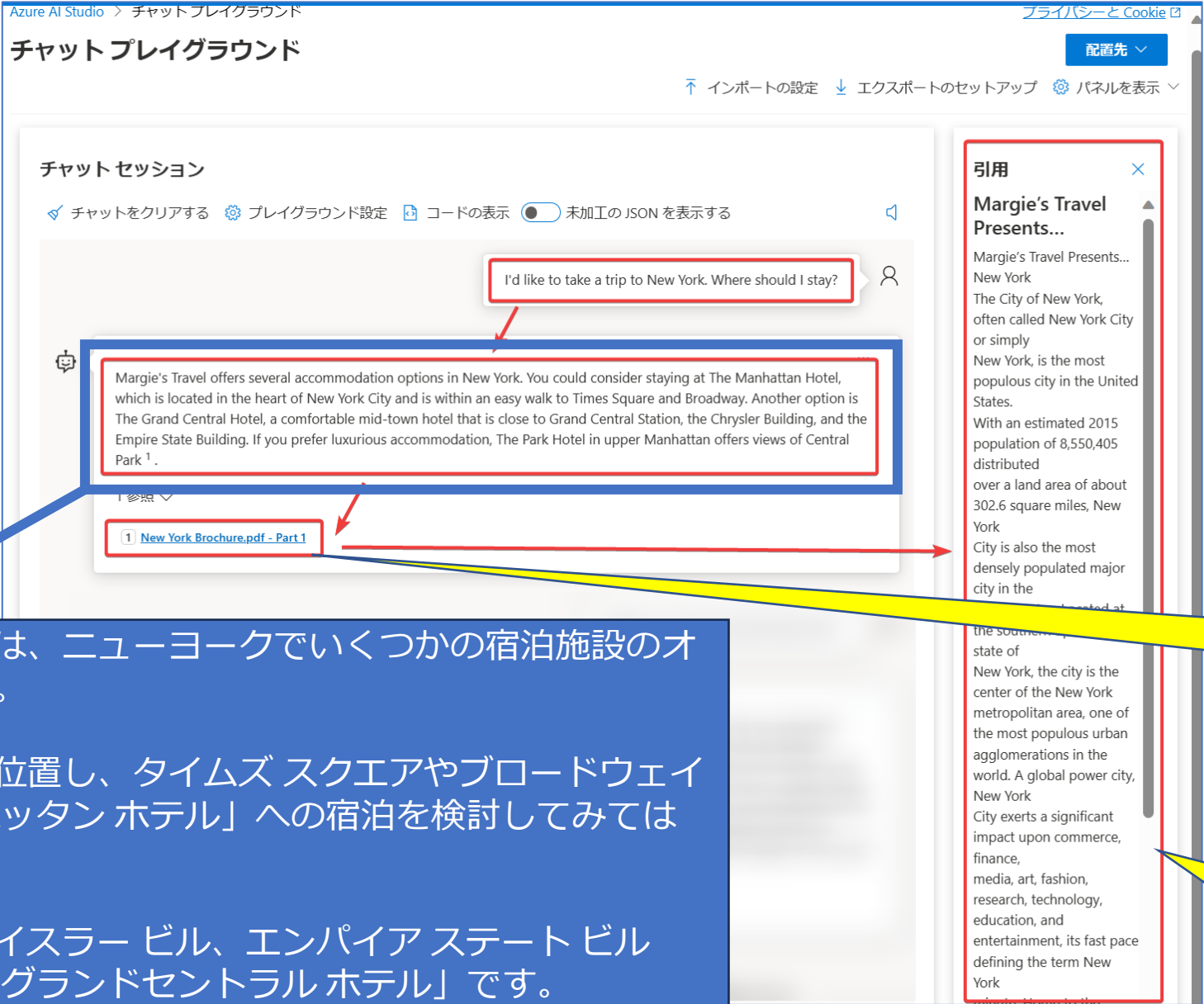
キーワード

戻る

保存して閉じる

キャンセル

■再度、「ニューヨークへの旅行における宿泊先」についてAIアシスタントに質問すると、on your dataにより、Azure AI Searchを使用した検索が行われ、PDF資料の中から、質問に対応するデータが取得される。そのデータを使って、より適切な回答が生成される。また、この際に使用されたデータを確認できる。



「マーギーズトラベル」では、ニューヨークでいくつかの宿泊施設のオプションを提供しています。

ニューヨーク市の中心部に位置し、タイムズスクエアやブロードウェイまで徒歩圏内の「ザマンハッタンホテル」への宿泊を検討してみてください。

もう1つの選択肢は、クライスラービル、エンパイアステートビルディングに近い快適な「ザグランドセントラルホテル」です。

引用元のデータを表示するリンク

引用元のデータ

モジュール6は以上です。

- 休憩しましょう。