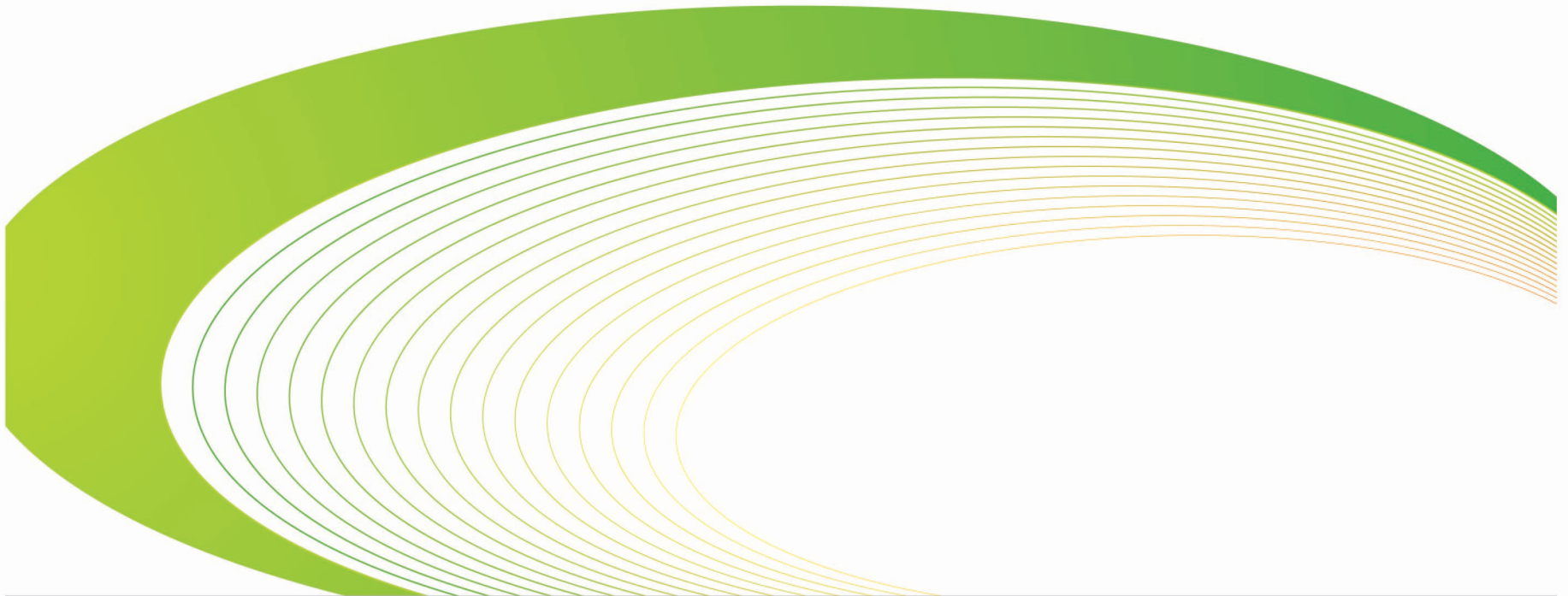


株式会社トップワンテック

Solution Catalogue



## 「Azure OpenAI」の概要

AzureはOpenAIのモデルを展開する唯一のパブリッククラウドサービス

<b>APIでOpenAIの 推論機能を提供</b>	エンドポイントへのリクエストを投げるだけで推論結果が得られます。 <u>AIを動作させるインフラを新たに準備・開発する必要はありません。</u> <u>API仕様やライブラリも基本は共通。</u> (REST API、Python SDKから呼び出しが可能)
<b>SLA・サポート付の サービス提供</b>	<u>99.9%以上の稼働率を保証するSLAを規定し、</u> Azureのサポートサービスが利用可能
<b>Azureのセキュリティ 機能が統合</b>	<u>Microsoft Entra IDと連携した認証、プライベートネットワークでの利用、</u> <u>データの暗号化、コンテンツフィルタリングなど、企業がAIを使うリスクを</u> 軽減するセキュリティをカバーします。
<b>エンタープライズ向けの データ管理</b>	マイクロソフトのモデルトレーニングにお客様データを利用しません。 デフォルトでは入出力データは悪用/誤用の監視目的で30日保持され、承認されたマイクロソフト社員が不正利用時にレビューする可能性はありますが、監視のためのログ保存プロセスはオプトアウト申請が可能で、承認されればログは保持されません。

## 「Azure OpenAI」の概要

### □提供可能なモデルとAPI一覧

モデル	対応API	概要
ChatGPTモデル	Chat completion	ユーザーの入力をチャット形式で返答します。 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ます。
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、文書同志の類似度を計算でき検索などに利用可能です。
DALL-Eモデル	Image generation (preview)	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。
Whisperモデル	(今後リリース予定)	

### 【参考】GPTモデルの種類と用途

GPT (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	ユーザーの問いかけ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
	指示応答ベース GPT-3.5モデル	gpt-3.5-turbo-instruct (Coming Soon)	(今後リリース予定)
	GPT-4モデル	gpt-4 (トークン8k) gpt-4-32k	ユーザーの入力に対し極めて高い解釈性を持ち、 指示に沿った返答が可能。
GPT (Embedding APIで使用)	Text-embedding-ada-002 (トークン8k)	高速で安価。一般のEmbeddingモデルと比較し、かなり大きなトークンサイズでも ベクトル化が実行可能。	

# AI活用



## 「Azure OpenAI Service導入支援サービス」

Azure OpenAIはAzureのセキュリティ機能を使用したうえで、OpenAIと同じモデルを実行できるサービスです。プライベートネットワークやAIコンテンツのフィルター処理を使用し、自社システムとして活用が可能です。トップワンテックでは、お客様のAzure OpenAI 活用に向けた、企画～導入後の運用・改善まで一括でサポートを行います。

### Azure OpenAIサービス導入のSTEP

Azure OpenAI を導入するには、以下のSTEPで導入を検討します。各STEPごとに最適なサービスでお客様のAIサービス導入をご支援します。

#### Step1 課題分析 /AIサービス理解

課題がどこにあるのか？  
AIを活用して何を解決したいのか？  
AIサービスでどのようなことができるのか？  
を整理する必要があります。

#### Step2 導入範囲の明確化 /検証

AIサービスを理解したうえで、  
AIを活用する範囲を明確化します。  
限定的な用途、範囲で実現性検証（PoC）を  
実施し、結果を検証します。

#### Step3 実装 /運用・改善

PoCでの検証結果を受け、AI環境の構築を行います。  
環境構築後は、精度向上や再学習など  
更なる活用に向けて運用、改善のサイクルを回していく  
必要があります。

### トップワンテック Azure OpenAI導入支援サービス

#### Azure OpenAI 簡易PoCサービス

まずAzure OpenAIをお試しいただきたい方向けのサービス  
・お客様Azure環境にPoC用のOpenAI Serviceを構成  
・PoC期間中のQAサポート  
※OpenAI学習済みのデータを使用します。

**¥ 1,100,000**

#### Azure OpenAI PoC支援サービス

Azure OpenAI導入に向けた、PoCの支援を行います。  
・PoCプラン検討  
・お客様Azure環境にPoC用のOpenAI Serviceを構成  
・PoC期間中のQAサポート  
・結果レポート作成  
※検証用のデータはお客様にてご準備いただく必要があります。

**¥ 1,500,000**

#### Azure OpenAI 導入支援サービス

お客様環境にAzure OpenAIの導入をご支援します。  
・お客様Azure環境にOpenAI Serviceを構成  
・OpenAIに関連する各Azureリソースを構成  
・運用に必要な各種手順書のご提供  
・お引き渡し後のQAサポート

**別途お見積り**

※簡易PoCサービスご利用後、PoC支援サービスを利用する場合は90万円（税別）でご利用可能です。

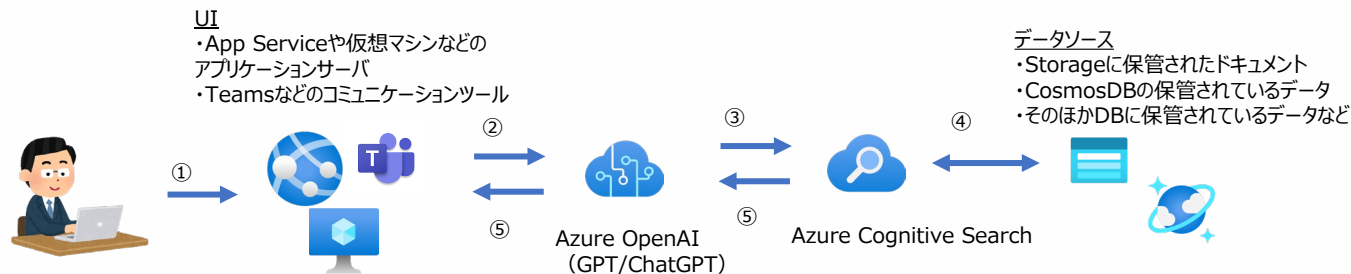
# AI活用

## 「Azure OpenAI Service導入支援サービス」

### Azure OpenAI PoC支援サービスの構成例

#### 【Point】

- 社内限定のチャットサポートを導入したい
- 回答は社内ナレッジを使用して回答を行いたい



#### 【処理の流れ】

- ① ユーザーが質問
- ② Azure Open AI Service (GPT/ChatGPT) が質問の内容を最適化し、キーワード検索用のクエリを作成
- ③ Azure Open AI Serviceからキーワード検索用のクエリをAzure Cognitive Searchに連携する
- ④ Azure Cognitive Searchはデータソースを検索し関連文書を取得
- ⑤ 検索結果やチャット履歴を利用して、Azure OpenAI (ChatGPT) が回答を生成

#### 【使用するAzureリソース (例)】

#	Service名	機能/役割	参考SKU	単価	コスト	コスト試算例
1	App Service	フロントUI・Cognitive SearchおよびopenAIの連携	Basic:B1	従量課金：¥10.538/時間	¥1,743.54	月730時間稼働想定
2	Azure Cognitive Search	Datasourceのインデックス保持・Chat使用時の検索機能を提供	Basic	従量課金：1SUあたり¥14.19/時間	¥10,358.70	月730時間稼働想定
3	Azure Open AI	Chat本文の検索キーワード抽出 (Text-Davinci)・検索結果を踏まえた回答の生成(ChatGPT)	Standard:S0	使用モデル：Text-Davinci,ChatGPT (gpt-3.5-turbo) Text-Davinci：¥2.81/1000トークン ChatGPT (gpt-3.5-turbo)：¥0.28/1000トークン	Text-Davinci：¥56,200.00 ChatGPT：¥42,000.00	Text-Davinci:20,000,000token使用想定 ChatGPT:150,000,000token使用想定
4	Datasource (BLOB)	Chatの回答内容になるデータを格納 (.txt)	Standard LRS/ホット	従量課金：¥2.9504/GB	¥1.48	ナレッジ総容量500MB想定

**月額想定ランニングコスト：¥110,303.72 (税抜)**

※Azure料金計算ツールで算出 (1 USD = 140.50 JPY)

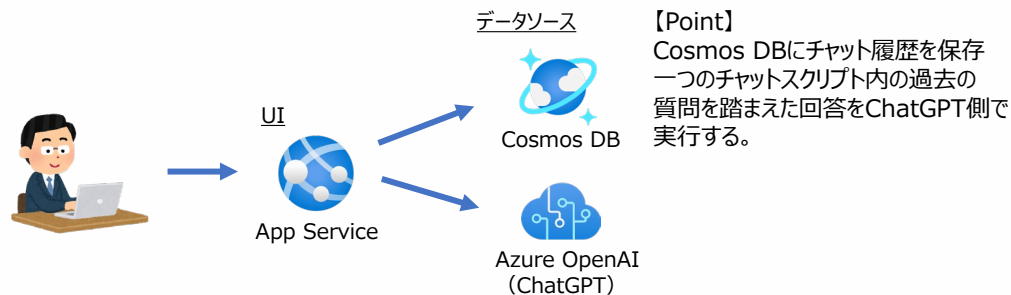
# AI活用

## 「Azure OpenAI Service導入支援サービス」

### Azure OpenAI 簡易PoCサービスの構成例

#### 【Point】

- Chat GPTを試してから、今後の活用方法を考えたい
- Azure OpenAIでのChat GPT環境を構成したい。



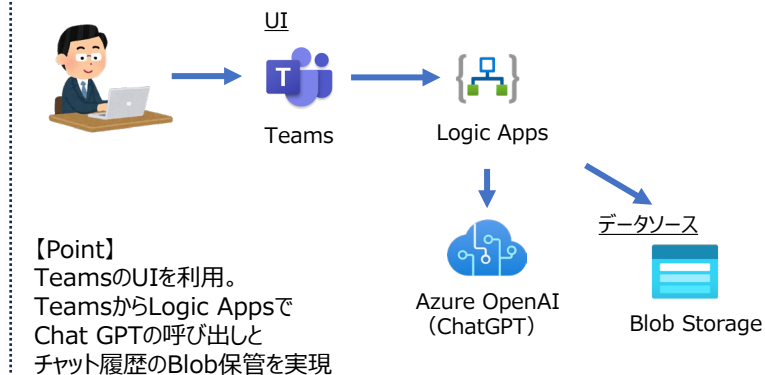
#### 【使用するAzureリソース（例）】

#	Service名	機能/役割	参考SKU	単価	コスト	コスト試算例
1	App Service	フロントUI・Cognitive SearchおよびopenAIの連携	Basic:B1	従量課金：¥10.538/時間	¥1,743.54	月730時間稼働想定
2	Azure Open AI	Chat本文の検索キーワード抽出 (Text-Davinci)・検索結果を踏まえた回答の生成(ChatGPT)	Standard:S0	使用モデル：ChatGPT (gpt-3.5-turbo) ChatGPT (gpt-3.5-turbo) : ¥0.28/1000トークン	ChatGPT: ¥14,000.00	ChatGPT:50,000,000tokens使用想定
3	Cosmos DB	Chatの質問履歴を保存	Free	100 RU/秒：¥1.264/時間 1GB：¥39.69	¥0	無料枠使用想定 1000RU/秒と25GBのストレージ使用 ※サブスクリプション毎に1つのみ有効

**月額想定ランニングコスト：¥15,743.54（税抜）**

※Azure料金計算ツールで算出（1 USD = 140.50 JPY）

#### （参考） Azure OpenAI for Teams 簡易PoCの構成例





## 「Azure OpenAI Service導入支援サービス」

【参考】GPTの課金単位トークンとは  
 トークンカウントについてはtiktokenと呼ばれるライブラリで確認が可能。  
 新しいモデルになるほどトークン効率の良いcl100k\_baseがトークン化に使用される。

例 「こんにちはOpenAI」 → トークン化 → ["こんにちは", "Open", "AI"] (3トークン)

モデル	エンコーディングツール	文字当たりのトークン数 (参考値)	トークンあたりの文字数 (参考値)
gpt-4 gpt-3.5-turbo 他	cl100k_base	1.0905	0.9170
text-davinci-003 text-davinci-002 他	p50k_base	1.4013	0.7136
text-curie-001 text-babbage-001 text-ada-001 他	r50k_base	1.4015	0.7135

### □コストの概算をシュミレート

GPTは入力と出力の両方のトークンに対して課金される。

- 従業員 1 万人の会社でChatGPTを業務利用するケース (利用が多いことを想定したケース)

ユーザ数	[人]	10000
1日のチャット回数	[回]	10
1チャットあたりのトークン数	[トークン]	1100
月の稼働日	[日]	20
1KトークンあたりのGPT3.5の料金	[¥]	0.265
1KトークンあたりのGPT4 (32K) 料金	入力	[¥] 7.95
	出力	[¥] 15.9



#### GPT-3.5

$$10000 \times 10 \times 1100 \times 20 \times 0.265 \div 1000$$

**583,000円/月**

#### GPT-4 (入力:出力=6:4と仮定)

$$10000 \times 10 \times 1100 \times 20 \times (7.95 \times 0.6 + 15.9 \times 0.4) \div 1000$$

**24,486,000円/月**

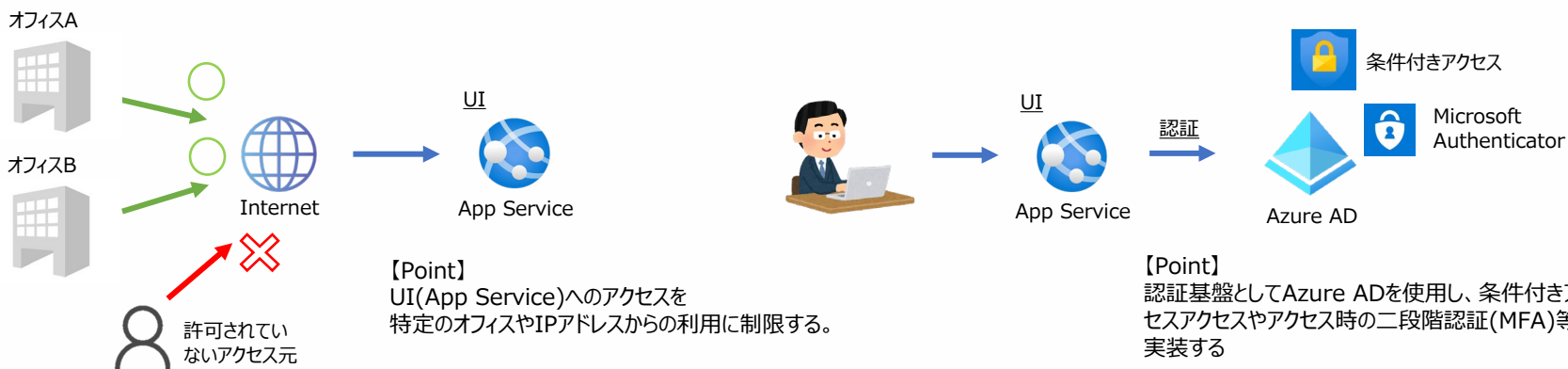
# AI活用

## 「Azure OpenAI Service」をセキュアに利用したいといった要望にも対応

Azure OpenAI を利用する際のセキュリティを向上させる ※一部、今後展開予定の機能も含んでおります。

### 【Point】

- グローバル環境からのアクセスに対して、アクセス制限や二段階認証等の認証セキュリティを導入する



### 【Point】

- Azure OpenAIをオンプレミスなどのプライベートネットワーク(閉域環境)から利用する

