

SyncLect Generative AI

株式会社ヘッドウォータース

2023年x月xx日

証券コード 4011

▶ コンテンツ

- シナリオ概要
- アプリUI
- アーキテクチャ
- デプロイ方法
- 考慮事項

シナリオ概要



▶ シナリオ概要

企業内ドキュメントとAzure OpenAI Serviceを活用した、全社業務支援／効率化チャットボットシステムを、安全な閉域網構成で提供可能です。

活用例：

- ✓ 一人ブレストや壁打ち
- ✓ 提案書雛形作成
- ✓ 社内規定ドキュメントの検索により、管理部の負荷軽減
- ✓ 社内ツールドキュメントの検索により、運用チームの負荷軽減

パッケージテンプレート化されている為、“とりあえず始めてみたい”PoCフェーズでも早期に導入できます。

ドキュメント検索をカテゴリ分け可能な為、利用シーンや特定業務別に、検索ドキュメントを分けたり、プロンプトチューニングが可能です。

アプリUI



▶ アプリUI (SyncLect Generative AI Bot on PowerPlatform)



The screenshot shows a chatbot interface within a Power Apps environment. The title bar reads "Power Apps | AIチャットボット" and "AIチャットボット GPT4". The chat history includes:

- User: 日本一高い山は？
- SyncLect: 日本一高い山は富士山です。標高は3,776メートルです。
- User: 二番目は？
- SyncLect: (Loading spinner)

A text input field at the bottom contains the text: 何でも聞いてください

▶ アプリUI (SyncLect Generative AI Bot on Web)



The screenshot displays the user interface for the AI Chatbot GPT3.5. On the left side, there is a sidebar with a '+ 新規チャット' (New Chat) button, a search bar labeled 'Search...', and a 'No data.' message. At the bottom of the sidebar is a '設定' (Settings) button. The main area is titled 'AIチャットボット GPT3.5' and contains a configuration panel with the following elements:

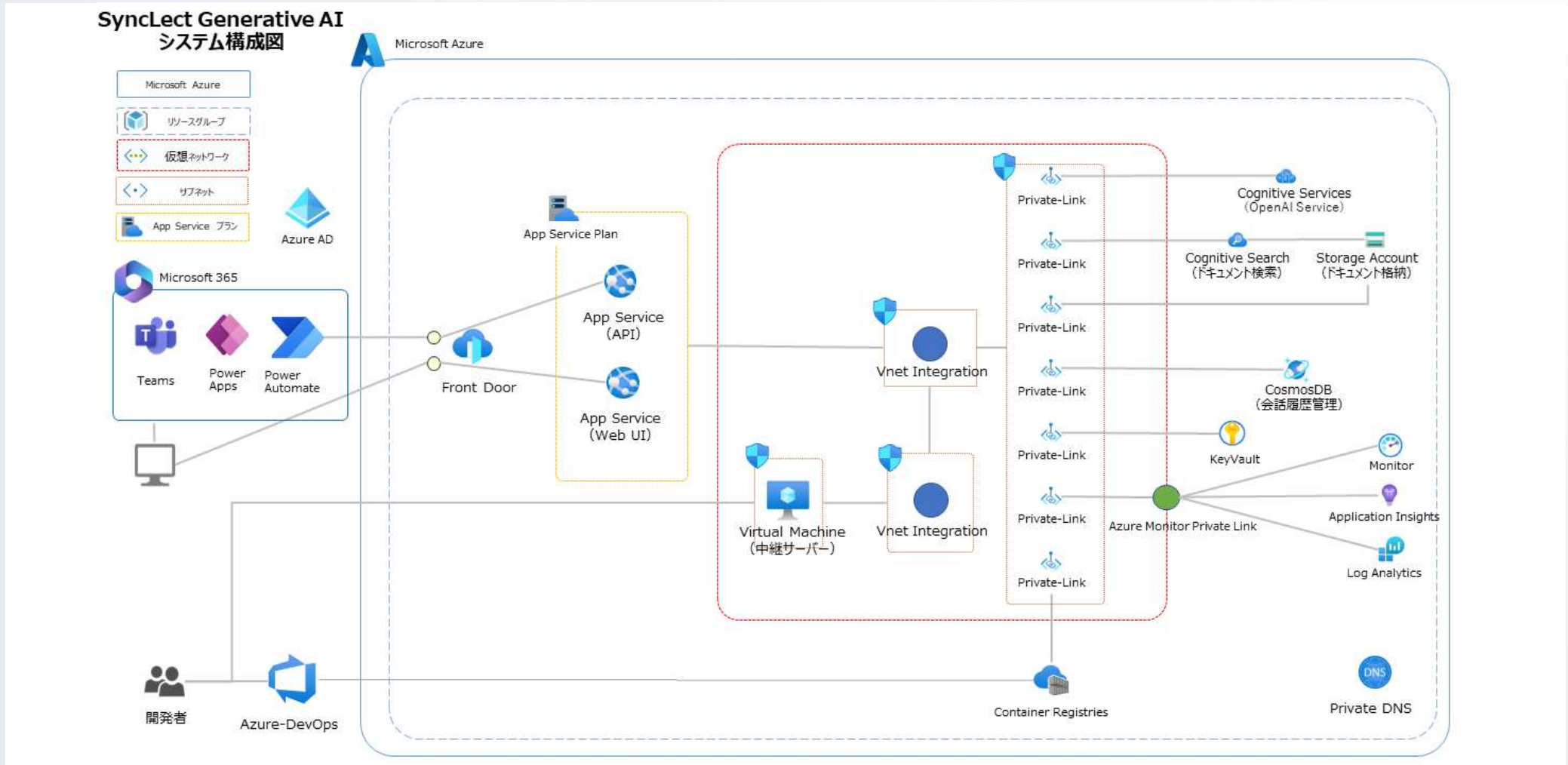
- モデル** (Model): A dropdown menu currently showing 'gpt-3.5-turbo'.
- 機能選択** (Function Selection): A list of three options: '社内文書モード' (Intranet Document Mode), '通常モード' (Normal Mode), and '社内規則モード' (Intranet Rule Mode). The '社内文書モード' option is currently selected and highlighted.

At the bottom of the main area, there is a text input field with the placeholder 'メッセージを送る' (Send message) and a send button with a paper plane icon.

アーキテクチャ



▶ アーキテクチャ



デプロイ方法



▶ デプロイ方法

ソースコードをGit管理して、CI/CDツールにてアプリケーションのデプロイが可能です。
Dockerを前提とした構成となっております。

弊社では、チームでのソースコード管理、シームレスなデプロイの実現にAzure DevOpsを利用しています。

Azure DevOps

- Azure Repos : Gitベースのソースコード管理
- Azure Pipelines : コードビルド、テスト、デプロイを自動化できるCI/CDサービス

Azure DevOps上でDockerイメージをビルド、Azure Container Registryへ配置。Azure App ServiceでDocker Pullすることでデプロイ可能です。

考慮事項



▶ 考慮事項

■ 可用性

- Azureクラウドでは、「可用性ゾーン」や「リージョン」を適切に組み合わせることで、可用性を高められます。
- Azure App Service、Azure OpenAI Service、Azure Cognitive Search、Azure Cosmos DBなどは、ゾーン冗長、リージョン冗長、geoレプリケーション化することで高可用性を実現出来る為、導入を検討下さい。

■ 回復性

- Application Insightsにてアプリケーションの正常性を監視して、パフォーマンス問題時にアラート発砲が可能になります。
- 各データの保持期間、バックアップ場所を設計することで、より回復性の高いシステムになります。

■ セキュリティ

- 本システムでは、企業内データ活用を前提としたシナリオの為、標準で各Azureコンポーネントをプライベートネットワークで構成しています。
- また、各アプリケーションシークレットは、Azure Key Vaultにて管理しています。
- Azure App ServiceのEasy Auth機能によるAzureAD認証に標準で対応しています。オプションで他ソーシャルアカウント認証にも対応可能です。
- より高度なセキュリティとして、WAFやセキュリティ製品の導入を検討下さい。

▶ 考慮事項

■ コスト最適化

- 冗長化やセキュリティ製品の導入はコストに影響します。要件やフェーズに合わせて適切に設計下さい。

■ オペレーショナルエクセレンス

- 適切な監視とログ収集設計を検討下さい。
- API Managementを導入することで、API利用の監視やトレースを行うことが容易になります。
- ソフトウェア/インフラ設計の改修を円滑に進められるよう、DevOpsプロセスを確立してください。

■ パフォーマンス効率

- アプリケーション負荷を予測して、スケーラビリティ設計を検討下さい。Azure App Serviceは負荷に応じた水平スケール設定が可能です。
- アプリ側のキャッシュ戦略、API Managementによるスロットリング導入などもご検討下さい。

■ ドキュメント検索の速度と精度

- Azure Cognitive Searchによるドキュメント検索の速度と精度を向上させる為の施策をご検討下さい。
- コストと要件に合わせて、ベクターストアもご検討下さい。

■ 回答の精度

- LLM（GPTモデル等）の性質上、期待する回答が得られない場合があることをご留意下さい。
- 検索モードごとに、システムプロンプトの変更が可能となっておりますのでご活用下さい。