Azure OpenAI Service
リファレンスアーキテクチャ

社内文書検索

2023/9/15

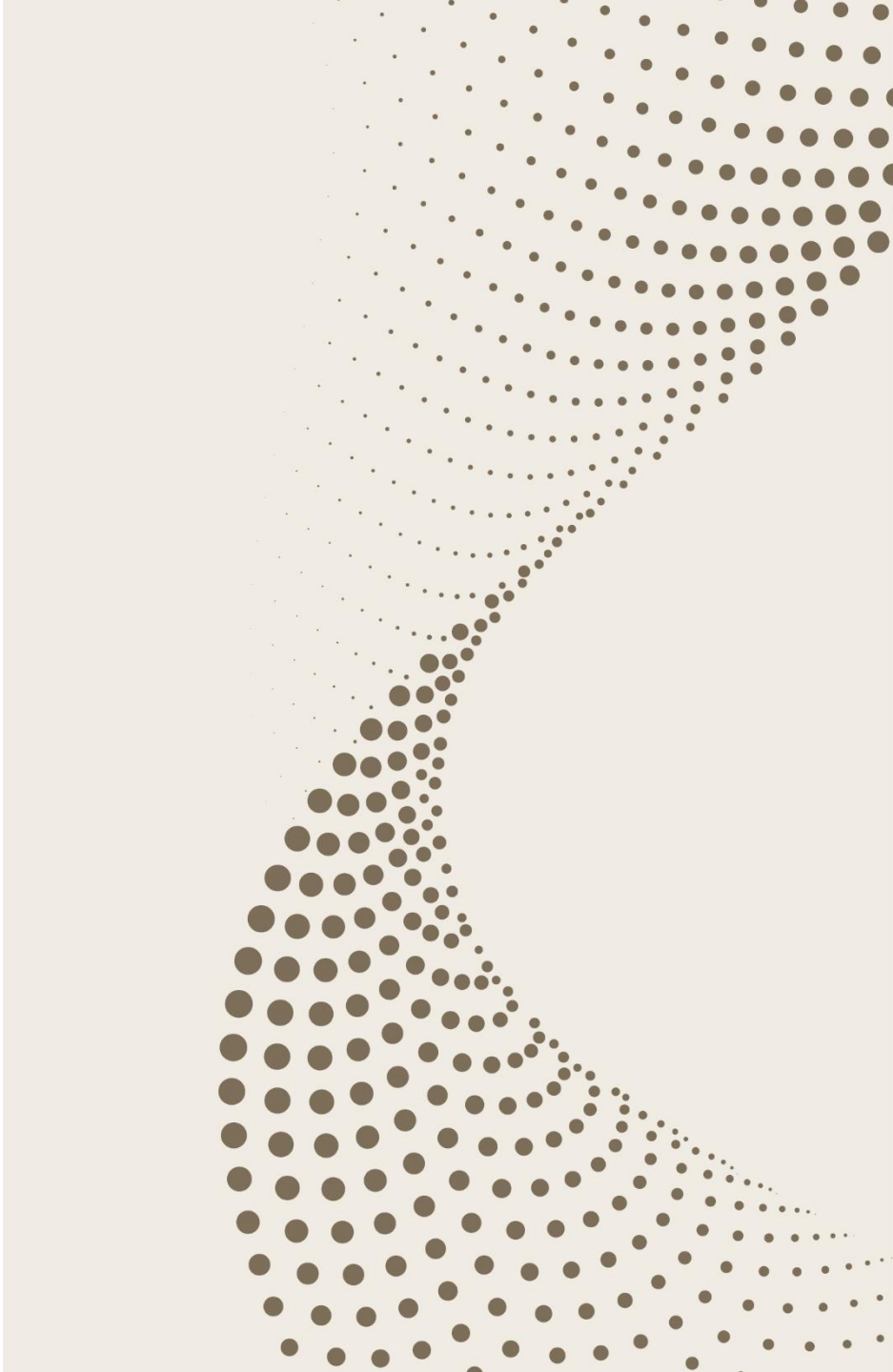
ABeam Consulting®

Build Beyond As One.

コンテンツ

confidential ©2023 ABeam Consulting Ltd.

1.シナリオ概要



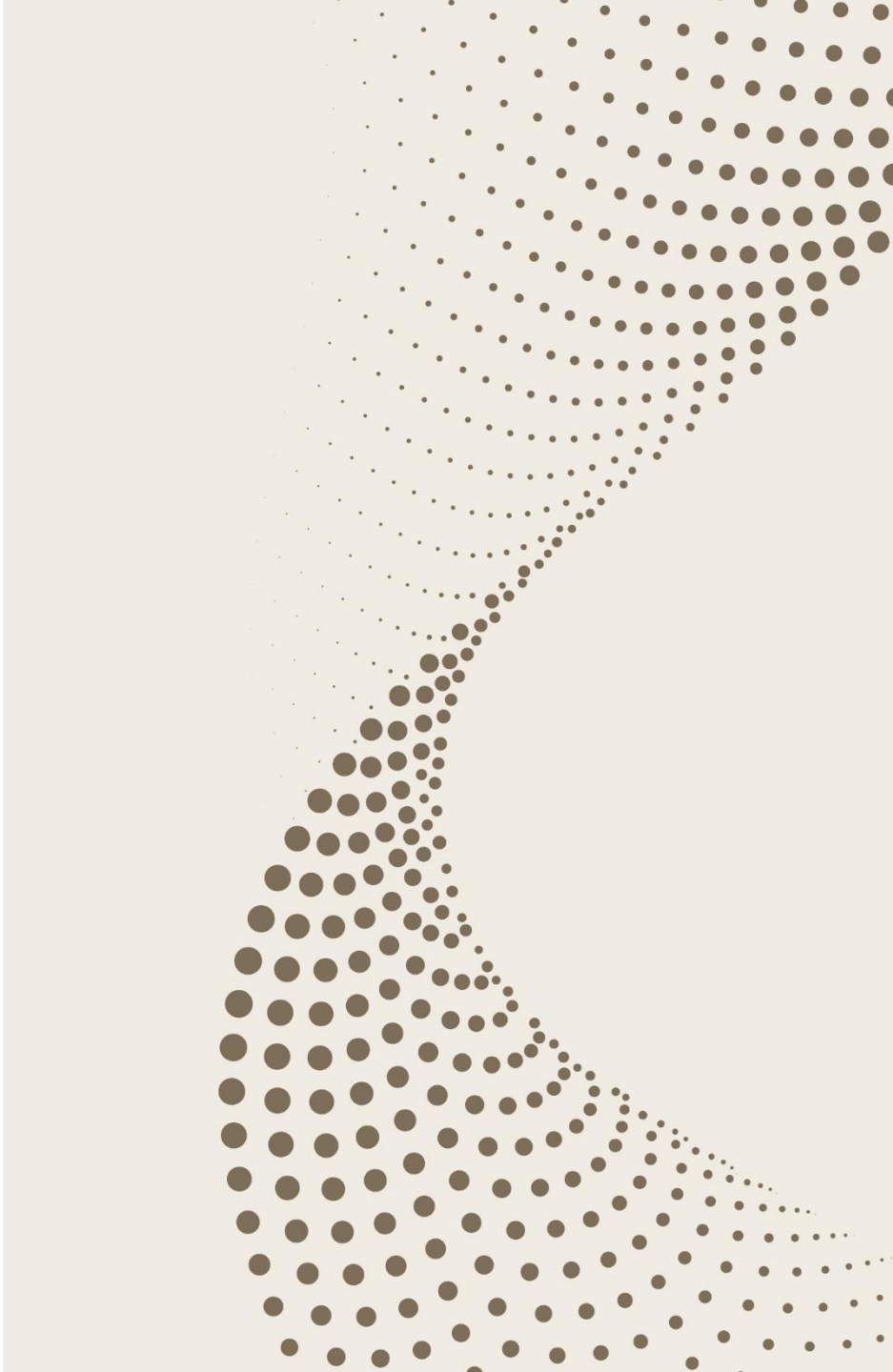
1.シナリオ概要

- ChatGPTは様々な業種業界での活用が想定され、個人利用でも業務効率化や作業品質向上に繋がる可能性が大いにあります。
- 但し、個人の業務利用が進むと機密情報流出の恐れもあるため**セキュアな**社内公式ChatGPTを実装します。
- “社内ナレッジ検索”を先行着手し、順次、”議事録作成”、”リサーチ業務”、”開発業務”等の機能を実装予定です。

セキュリティ観点	業務効率化観点
個人がOpenAI社のChatGPTを利用すると顧客情報漏洩等の事故が起こる危険性があるため、 ①利用ルールを策定 ②アクセス管理や監視機能のある安全な環境を整備	過去の提案書、成果物、プロジェクト開発標準などのナレッジを読み込ませることで、企業独自の情報に基づく回答作成や、議事録作成支援など社内業務の効率化をはかる

confidential ©2023 ABeam Consulting Ltd.

2.デモUI

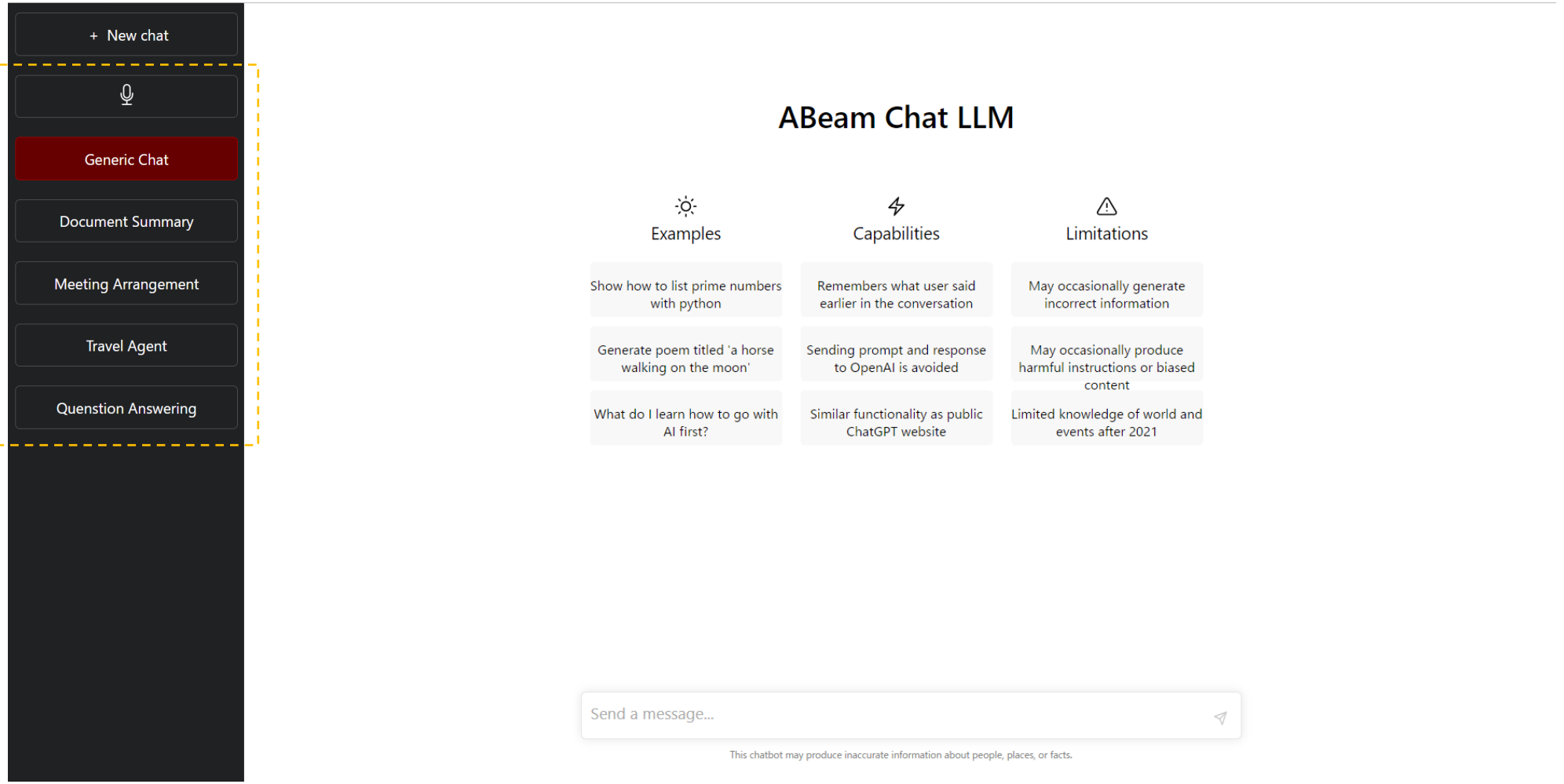


2.デモUI　完成イメージ

■ ChatGPTに似た独自のフロントエンドを用意※1し利用頻度が多いユースケースはメニュー化する想定です。

※1.社員がPlaygroundを直接利用することは難しいため

サイドバーに様々な機能(メニュー)を追加していく想定

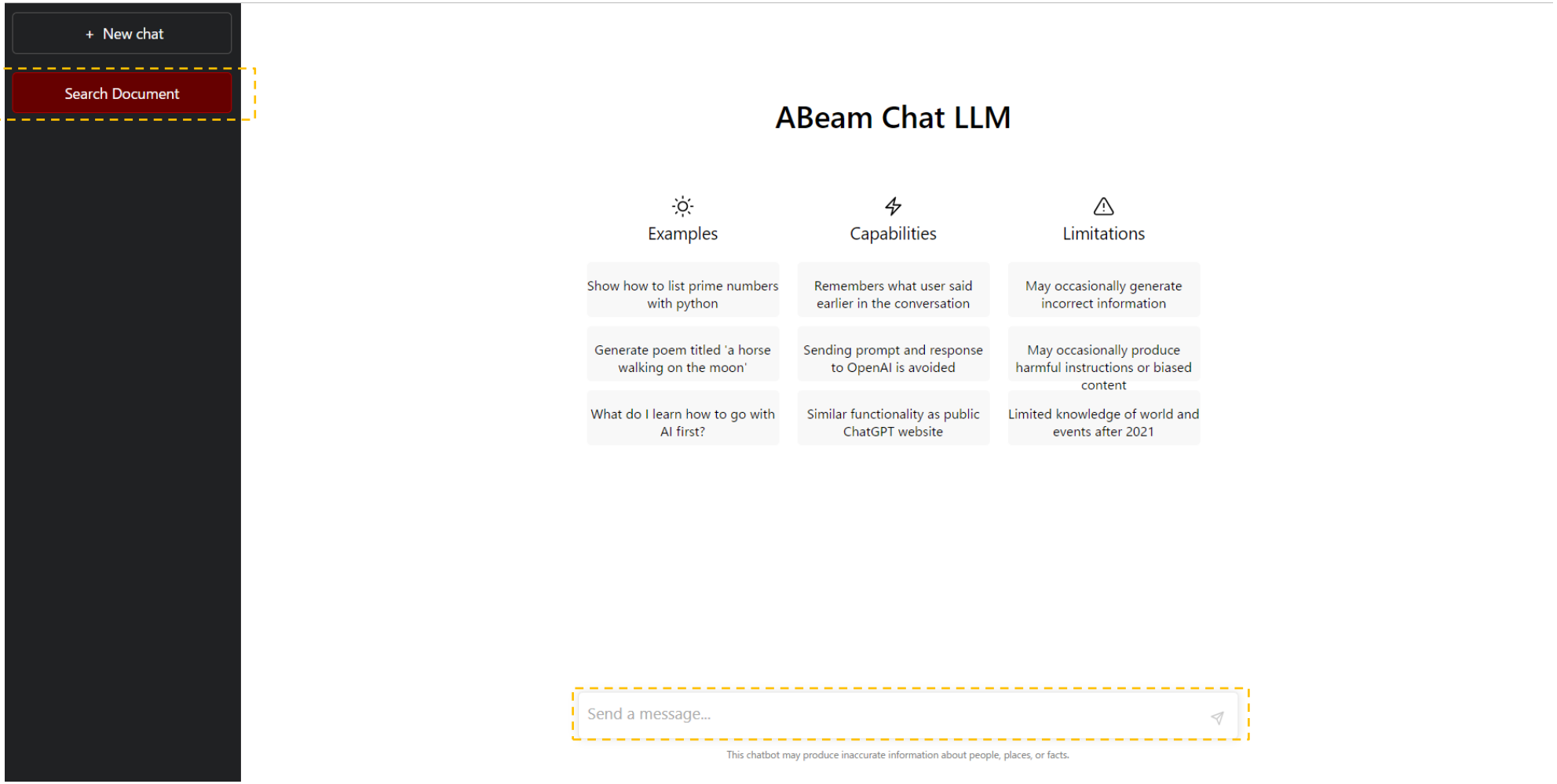


confidential ©2023 ABeam Consulting Ltd.

2.デモUI　現在の実装 メインページ

■ 現在は社内ナレッジを自然文で検索できるシステムを構築済みです。

現在はナレッジ
検索機能のみ実装



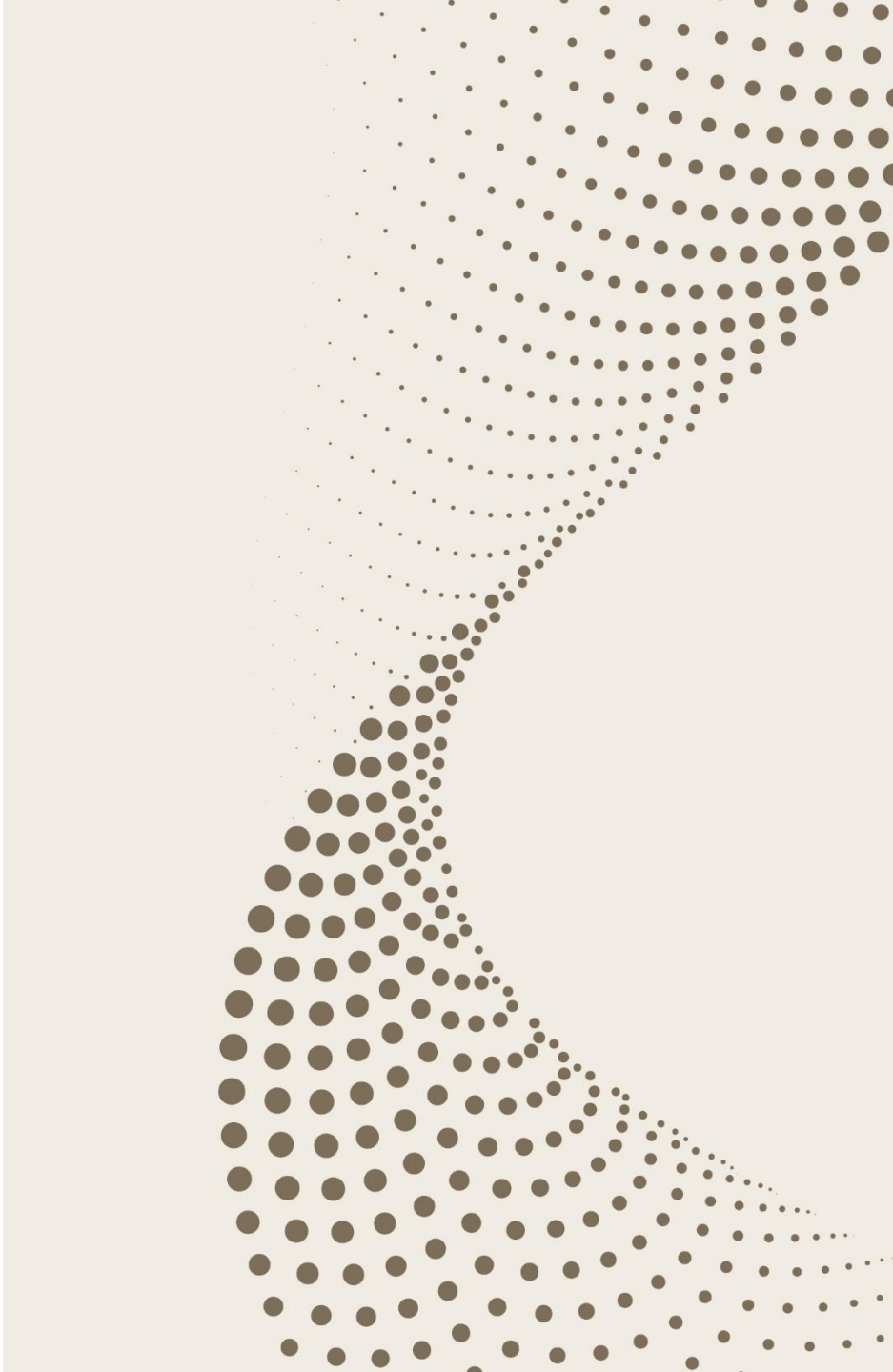
confidential ©2023 ABeam Consulting Ltd.

2.デモUI　現在の実装 チャットページ

■ 質問を行うとチャット画面に遷移し、検索結果が表示されます。

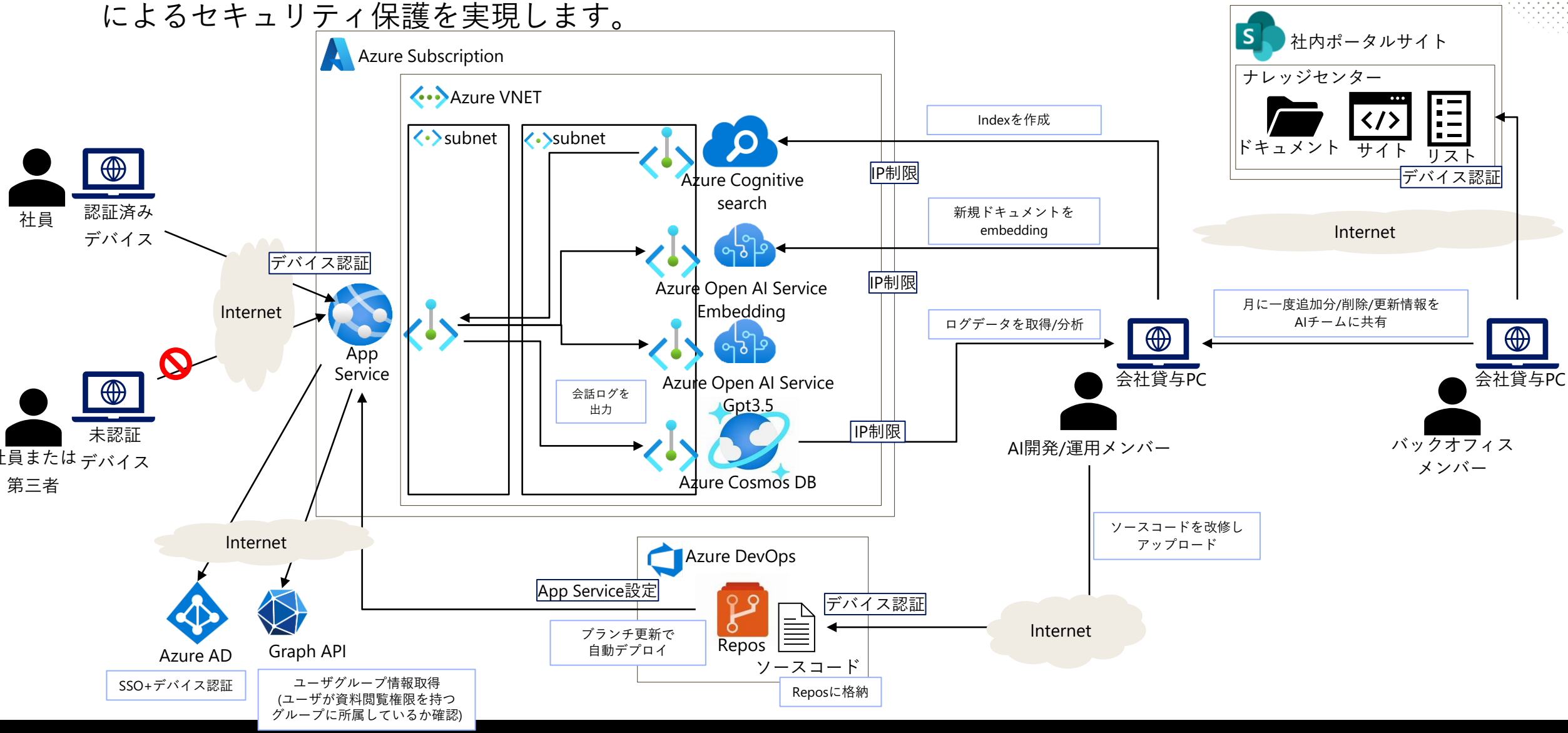
confidential ©2023 ABeam Consulting Ltd.

3.アーキテクチャ



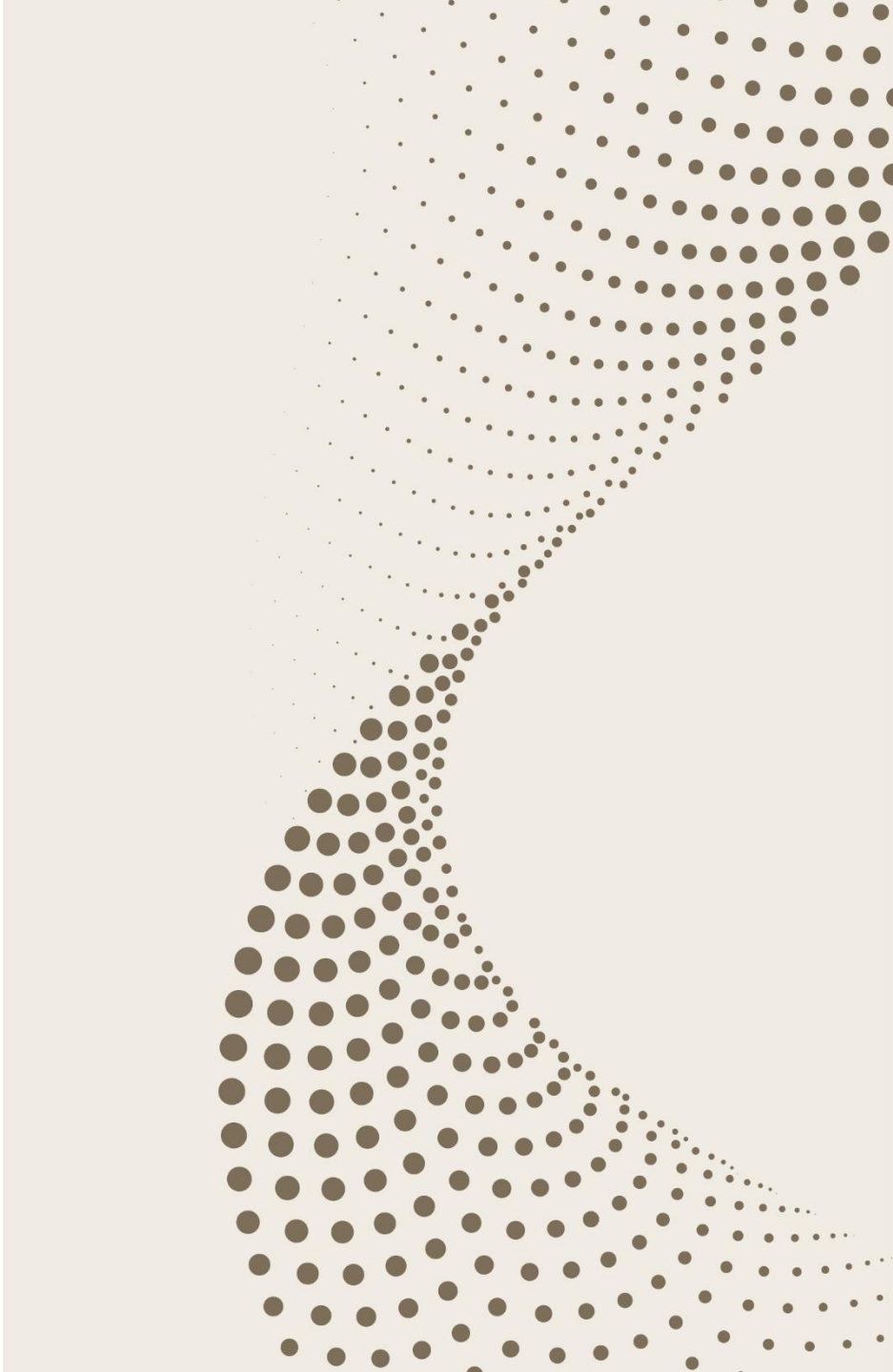
アーキテクチャ図

■ デバイス認証でユーザーアクセスを制御し、ナレッジをSharepointサイトに格納することで、権限管理によるセキュリティ保護を実現します。



confidential ©2023 ABeam Consulting Ltd.

4.考慮事項



4.Azure Well-Architected Framework観点での考慮事項 (1/2)

Microsoft Azure Well-Architected Frameworkの観点に則った以下の点を考慮した実装・運用は、非機能要件の充足に有効です。

信頼性	可用性	• Azureのサービスでは「可用性ゾーン」や「リージョン」といった単位で可用性を設計しており、 これらを適切に組み合わせることで、ビジネス クリティカルなワークロードの信頼性を実現するように設計することが可能です。詳細は「Azure リージョンと可用性ゾーンとは」をご参照ください。 • このシナリオで用いられているAzure Open AI、Azure Cosmos DB for PostgreSQL 等のコンポーネントはそれぞれゾーン冗長、リージョン冗長、geoレプリケーションなど高可用性のオプションや構成を利用可能です。必要となる可用性に応じて導入を検討してください。複数リージョン間/複数ゾーン間でAct-Act構成を取る場合にはAzure Front Door、Azure Application Gatewayなどの利用をご推奨します。
	回復性	• Azure Cosmos DB は、定期的にデータのバックアップを自動的に取ります。 自動バックアップは、データベース操作のパフォーマンスや可用性に影響を与えずに取得されます。 すべてのバックアップは別々のストレージ サービス内に個別に保存されます。これらのバックアップは、リージョンの障害からの回復性を確保するためにグローバルにレプリケートされます。 詳細については、「Azure Cosmos DB と信頼性」を参照してください。
セキュリティ		• セキュリティは、重要なデータやシステムの意図的な攻撃や悪用に対する保証を提供します。 詳細については、「セキュリティの重要な要素の概要」を参照してください。 • このシナリオでは、Azure ADを使用してユーザーを認証します。セキュリティで保護されたソリューションの設計に関する一般的なガイダンスについては、「Azure のセキュリティのドキュメント」を参照してください。 • このシナリオにおけるWebAppsとAzure Open AI Serviceの間の通信ではPrivate Linkを使用し、内部からアクセスしています。

confidential ©2023 ABeam Consulting Ltd.

4.Azure Well-Architected Framework観点での考慮事項 (2/2)

コスト最適化	• 不要な費用を削減し、運用効率を向上させる方法を検討することです。 詳しくは、コスト最適化の柱の概要に関する記事をご覧ください。
オペレーション エクセレンス	• システムの健全性の担保、トラブルの解決、利用動向の監視を行うためには適切な監視とログ収集が必要となります。詳細は「ワークロードの監視」をご参照ください。 • ソフトウェアのアップデートや脆弱性への対応など、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。詳細は「リリース エンジニアリングの継続的インテグレーション」をご参照ください。
パフォーマンス 効率	• アプリケーションの負荷が高まることを見越し、スケーラビリティの確保をあらかじめ検討することは重要です。詳細は「スケーリング用のアプリケーションを設計する」をご参照ください • App Serviceは負荷に応じて水平にスケールさせることが可能です。詳細については「自動スケーリングを有効にする方法」をご参照ください。 • 特定のユーザーにAzure Open AIの利用が集中することを避けたい場合には API Management によるクォータ導入などをご検討ください。設定の詳細は「Azure API Management を使用した高度な要求スロットル 」をご参照ください。

confidential ©2023 ABeam Consulting Ltd.

ABeam Consulting®

Build Beyond As One.

アビーム、ABeam及びそのロゴは、アビームコンサルティング株式会社の日本その他の国における登録商標です。
本文に記載されている会社名及び製品名は各社の商号、商標又は登録商標です。 ©2023 ABeam Consulting Ltd.