生成AIによるIT運用高度化と生産性向上の実証研究

藤田勝貫

1. はじめに

近年，生成AI（Generative AI）は急速に存在感を高めており，さまざまな業界でその有用性が注目されている．生成AIは，人間のように文章や画像を生成し，多岐にわたるタスクを自律的にこなすことができる革新的な技術であり[1]，業務効率化や自動化において大きな役割を果たしつつある[2]．IT運用の分野においても，生成AIを導入することで，システムの複雑化に伴う運用負担の軽減に寄与することが期待されている[3]．

ユニアデックス株式会社（以降，当社）では，生成AIの潜在能力を活用し，ITオペレーションの効率化とエンジニアの負担軽減を目指した取り組みを進めている[4]．

本稿では，当社の取り組みと，生成AI適用時に検討した具体的な設計ポイントや留意すべき点を整理して紹介する．また実証結果をもとに，IT運用における効率化と生産性向上への寄与について考察し，今後生成AIの導入を検討している企業に向けて，その有効性を示す．まず2章でIT運用の課題と，保守サポートに生成AIを適用した当社の取り組み（以降，本研究）を説明する．3章で生成AI適用時に検討した七つの項目を説明する．4章で生成AIの検証内容，5章で検証結果，6章で今後の展望を述べる．

2. 本研究の概要

本章では現状のIT運用における課題と，保守サポートに生成AIを適用した本研究の全体像を説明する．

2.1 現状のIT運用における課題と本研究

システムの複雑化や多種多様なデバイス，クラウドサービスの普及によるITインフラの多様化に伴い，IT運用の難易度は高くなっている[5]．従来の手法による監視や対応には限界があり，効率性を維持することが難しくなってきている．

インシデントが発生した際は，専門知識を持つエンジニアが問題を切り分けて，問題が何に起因するものかを特定して対策を検討する．迅速さが求められる作業の中で，エンジニアは，障害箇所に関係するマニュアルや不具合情報の確認，社内ナレッジや過去事例の検索など，散在している様々な情報から解決策を導かなければならない．そのため，インシデントの対応に要する時間は，エンジニアの経験やスキルに依存している．

当社ではAI技術を活用したIT運用のアプローチであるAIOps（Artificial Intelligence for IT Operations）の評価を2014年より行ってきた[6][7]．AIOpsは，AIを駆使して膨大なシステムデータやログを分析することで，システム障害の予測やトラブルシューティン

グを迅速化し，IT 運用の効率性を高める手法である．AIOps を導入することにより，従来エンジニアが担っていた多くのタスクが自動化され，運用管理の負担軽減が期待できる[3]．

しかし，従来の AIOps では主に数値データの解析が中心であり，自然言語の扱いは限定的であった．本研究では，自然言語を含む情報をターゲットに，従来の AIOps の発展形として生成 AI の活用を試みる．なお，従来型 AIOps と区別するため，生成 AI を活用した AIOps を GAIOps（Generative AIOps）と呼ぶ．

2. 2　GAIOps の特徴と目指す姿

GAIOps の特徴は，大規模言語モデル（Large Language Model, LLM）により，システムへの入出力に自然言語が利用できることにある．ログやマニュアル，エンジニアが行ってきた問い合わせ対応履歴や蓄積されたナレッジベースといった，自然言語で記述された大量の非構造化データを基に，情報整理や文章生成が自動化できるようになる．

例えば，ベンダーによる製品情報（不具合情報や脆弱性情報），スペシャリストのナレッジ，顧客の IT 資産情報を組み合わせて適切な行動を推論し，IT サービスの状態確認や意思決定に関して支援することができる．AI による意思決定支援を受けることで，従来はエンジニアに依存していた判断プロセスが効率化され，より迅速な運用ができるようになる（図 1）．

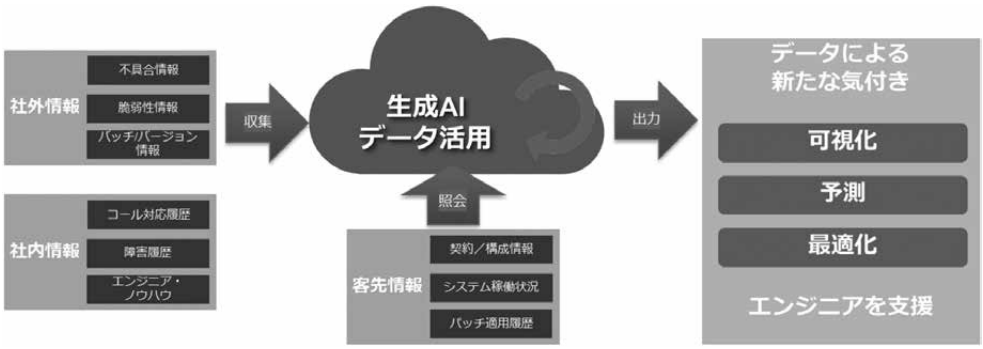

図 1　GAIOps のコンセプト

図 2 は保守サポートに生成 AI を適用した全体像を示している．問い合わせ対応における回答生成では，顧客からの問い合わせ内容を理解し，その意図に基づいた関連情報を自動的に検索・抽出する．その際，情報の出所を明らかにしながら分析し，信頼性の高い応答を提供することができる．また，生成された回答を活用し，顧客対応の効率を向上させエンジニアの負担を軽減する．同時に，迅速な対応ができることで，サービス品質の向上や顧客満足度の向上にも寄与する．

ただし，生成 AI を用いた自動化には課題が残っている．生成 AI は，膨大なデータに基づいて自然な文章を生成する能力があるものの，ハルシネーション（事実に基づかない内容を生成する現象）が発生する可能性がある．そのため，生成された文章をそのまま信

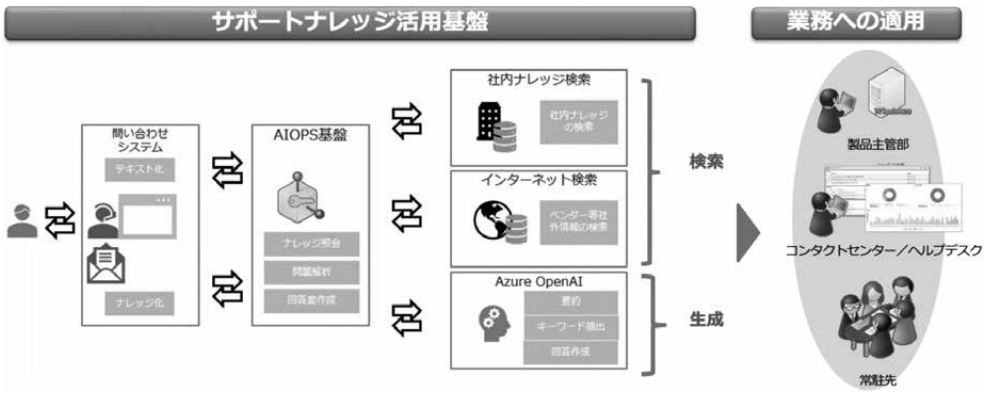

図 2　保守サポートへの生成 AI 活用全体像

用し，回答に使用することはできない．AI が生成した内容に対しては，エンジニアが適切な裏付けを取り，精査するプロセスが不可欠である．

3.　生成 AI 導入時の検討項目

本章では，当社が 2023 年 4 月に生成 AI を社内業務に導入する際に検討した七つの項目について，各節で紹介する（表 1）．

表 1　生成 AI 導入時の検討項目一覧

節番号	検討項目
3.1	適用業務の選定
3.2	内製化と外部製品の選択
3.3	AI モデル選定
3.4	社内データの活用と RAG 採用の背景
3.5	検索手法の選定
3.6	機密データの取り扱い
3.7	開始時期判断と失敗推奨

3.1　検討項目 1　適用業務の選定

生成 AI の導入において，適用対象業務を慎重に検討することが成否の鍵となる．適用業務の選定に失敗すると，十分な成果が得られずコストや機会の損失につながる．選定の基準として，課題の重要性，データ量や形式，実現可能性，定量的な評価指標といった複数の情報を総合的に評価することが重要である．以下 1）～ 4）で説明する．

1）課題の重要性

課題の重要性は，適用業務の選定においてキーとなる指標である．例えば，対応時間の短縮に繋がりやすい業務や，頻繁に問題が発生している業務，多くの組織で同様に行

われている業務を選択することが望ましい．また，その業務に関わる運用者の協力を得ることが肝要である．

2）データ量や形式

AIを活用する際には分析に適したデータが欠かせない．業務に関連するデータがAPI等を介して取得できる状態にあることで分析ができる．当社には，過去の問い合わせ履歴や技術ナレッジベース，ベンダーの不具合情報などのデータが蓄積されており，これらを活用することとした．

3）実現可能性

対象とする課題解決のための技術的実現性とシステムへの適用容易性を評価することが重要である．システム化が困難なものを実現しようとすると，コストが膨大となり，プロジェクトの継続が困難になる可能性がある．ベンダーから提供されているリファレンスアーキテクチャ[8]などを参考に，実現可能性について検討した．

4）定量的な評価指標

課題の重要度と関連し，生成AIによる効果を定量的に評価できる業務を選定することが重要である．生産性向上やコスト削減といった効果を定量的に評価できる業務は，導入の効果を明確に確認しやすい．他の業務への展開を判断するためにも，定量的な評価指標は重要なファクターである．

当社では，本基準に基づき，問い合わせ対応業務や脆弱性対策業務など，既存のナレッジを活用する業務に生成AIの適用を試みている．

3.2　検討項目2　内製化と外部製品の選択

生成AIの導入においては，内製化と外部製品活用の適切なバランスを見極めることが重要である．どの部分を内製化し，どの部分を外部製品に任せるかの判断は，プロジェクトの成功とノウハウ獲得に大きな影響を与える．AI技術は急速に進化しており，すべてを自社で開発・維持することは非現実的であるため，適切な組み合わせを見つけることで効率的な実装ができるようになる．

内製化により，構築を通じて知識とノウハウを蓄積することができる．一方，外部製品を活用することで，時間とコストの節約が期待できる．本研究では，生成AIの導入時に実施する活動を四つの構成要素に分類し，内製すべき部分と外部製品を利用すべき部分を明確に区分した（図3）．

以下，1）～4）で詳細に述べる．

1）データ運用

利用するデータの準備および整形を行い，目的に応じた出力データを得るためのプロンプトエンジニアリングを実施する．また，生成されたデータが業務で実際に利用できるかどうかを評価する．データや業務に関する知見は会社ごとに異なるので，外部に依存することで得られる効果は限定的である．そのため，本研究ではデータ運用を内製対

図 3　生成 AI 導入における活動

象とした．社内データや運用フローに最適化されたシステムを構築することで，運用中の変更や改良に対して迅速な対応ができるようになり，定期的に見直すことで，より業務に適したアウトプットを得られるようになる．

2）アプリ開発

生成 AI を活用して特定の課題を解決するアプリケーションを開発する．この開発には，チャットシステムやその他の業務アプリケーションが含まれ，業務効率化やコミュニケーション改善を目指して設計される．内製化により，ノウハウを蓄積し人材の育成を図りながら顧客サービス向上へ活用していく他，顧客向けのシステム構築提供などのビジネス展開も検討することができる．要件変更に対して柔軟に対応できるメリットもある．本研究では本項目を内製対象とした．

3）AI モデル（LLM）

生成 AI の核となる大規模言語モデル（LLM）は，データ収集やチューニングなど開発に時間とコストがかかるため，本研究では外部製品の利用が効果的だと判断した．LLM は，複雑なテキスト処理タスクを実行する能力を有している．詳細は 3.3 節で述べる．

4）基盤構築

アプリケーションやデータを適切に格納し，効率的に動作させるためのシステム基盤を構築する．この基盤は，全てのアプリケーションが安定して稼働するための要素であり，高い信頼性や拡張性が求められる．利用者のニーズに沿った選択ができるような柔軟な調整が求められるため，本研究では内製対象とした．

3. 3　検討項目 3　AI モデル選定

生成 AI 導入において，AI モデルの選定はアウトプットの精度を上げるために重要なステップである．生成 AI は大量のデータを基にパターンを学習し，その知識を活用して新しいデータを生成する．選定するモデルの性能や適用領域は，生成 AI が実際にどれだけ効果的に運用されるかを大きく左右する．

モデル選定に際しては，回答精度，信頼性，運用性を総合的に評価することが望ましい（図 4）．以下 1）～ 3）で説明する．

図 4　AI モデルの評価観点

1）回答精度

最も重視するものは回答精度であり，納得感のある出力が得られるかどうか，ビジネスで利用できるかどうかについて確認する．LLM ベンダーから公表されているベンチマーク結果は，各社でベンチマーク手法が異なるなど，机上では単純に比較できないことが多い．長文が多い，特殊な用語を使用するなどの，適用業務の特性を考慮して，固有のテストデータを用意しておくと比較が容易となる．また，実際の業務担当者によるアンケート評価が有用である．

2）信頼性

次に，回答の信頼性確認について述べる．LLM の信頼性を評価する観点として，真実性，安全性，公平性，堅牢性，プライバシー，機械倫理[*3]，透明性，説明責任の八つが紹介されている[9]．ハルシネーションなど回答内に虚偽が含まれないようにすることが重要なので，LLM の性能だけでなく，RAG など（3.4 節）他の補完要素も含めて検討すると良い．また，業務で使用するデータの中には機密データが含まれる場合があり，データの機密度を明確にすることが重要である．機密データの取り扱いについては相応のセキュリティ対策が求められる．LLM の機能として，追加学習のデータとされることを除外する設定（オプトアウト）があるかどうかなど，サービス提供元との契約やプライバシーポリシーを確認することが肝要である．

3）運用性

他にも運用費用や稼働率，リリースやテストの容易性，他のシステムとの連携性確保など運用性の観点からの比較も重要である．サービスを利用する場合は，ベンダーの価格体系を確認して費用のシミュレーションを行い，サービス約款などの契約内容を調査しておくと良い．

本研究では AI モデルに Azure OpenAI Service（AOAI）の GPT を採用した[*1]．

3. 4　検討項目 4　社内データの活用と RAG 採用の背景

生成 AI は事前学習済みのデータを用いて回答するため，学習対象でない社内ルールなどの情報については回答できない．社内データなど学習対象でない情報を活用するために

は，追加学習を行って知識を拡張しなければならない．知識拡張にはレベルの異なるいくつかの方法があり，それぞれの特性を理解して使い分ける．代表的なものにプロンプトエンジニアリング，RAG，ファインチューニングがある（図 5）．

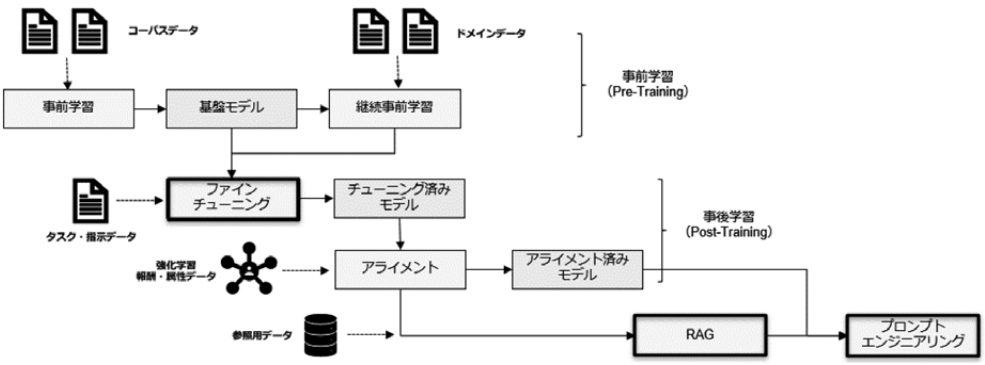

図 5　知識拡張の手法[10]

3. 4. 1　知識拡張の手法について

代表的な知識拡張の手法を，実装コストが少ない順に以下 1）～ 3）で紹介する．適用する環境や要件に応じて選択すると良い．

1）プロンプトエンジニアリング

望む出力結果が得られるように AI に対する入力（プロンプト）を工夫する手法である．利用者は，必要な情報を特定しシステムにプロンプトとして入力する．この方法は，限定された情報を扱う場合に効果的である．広範囲にわたるデータを扱う場合には適さないため，他の手法と組み合わせて使用する．

2）RAG（Retrieval-Augmented Generation；検索拡張生成）

プロンプトに関連する情報を LLM 外部のデータベースから検索して，検索結果を基に新しいコンテンツを生成する手法である．この手法を用いることで LLM に存在しない情報を基に回答を生成することができる，また，生成直前に情報を検索するため，対象のデータをリアルタイムに更新して，生成結果に反映することができる．頻繁に更新される情報を扱う際に特に適している．

3）ファインチューニング

この手法は，事前に学習された AI モデルに新しいデータを追加学習させることで，特定の目的に特化した性能を引き出すものである．大規模なデータ準備とモデルの再トレーニングが不可欠であり，チューニングも難しいため，時間とコストがかかる．特にデータの更新頻度が高い場合，頻繁な再トレーニングを要する．

3. 4. 2　RAG vs ファインチューニング

RAG とファインチューニングのどちらが知識拡張手法として適しているかを比較し，

本研究ではRAGを採用した．実装上のコスト効率が良い点が，RAGの大きな利点である．また，データの更新が頻繁に行われる業務でも，迅速に生成AIに反映させることができる点が選択の決め手となった．

Microsoftが2024年に発表した論文「Fine-Tuning or Retrieval? Comparing Knowledge Injection in LLMs」[11]の結果も，RAGが多くのテストでファインチューニングを上回る結果を示しており，RAGが優れていることが示されている．

3.4.3　社内データの種類と活用範囲

当研究がGAIOpsを実装する上で活用した社内データは，以下1）2）の二種類である．

1）問い合わせ履歴データ

過去の問い合わせ内容とその対応履歴を蓄積し分析に活用することで，類似の問題が発生した際に自動で対応策を提示することができる．本データの活用によりエンジニアが毎回ゼロから対応方針や回答を考えずに済むため，工数低減が期待できる．

2）技術ナレッジベース

社内外の技術ドキュメントやFAQを基に，生成AIが適切な情報を検索・提示する．複雑な技術的問題にも対応でき，特定の技術に関する深い知識を持たないエンジニアでも適切な対応ができるよう支援する．

これらのデータは生成AIが適切な応答や判断を行うため，日々の業務の中で追加更新され続けるものである．現時点では実証段階であるためデータは固定して取り扱っている．本データを自動的に取り込むことで．AIが常に最新のデータにアクセスできるようになり，リアルタイムで適切な問題解決支援を行うことを目指している．

3.5　検討項目5　検索手法の選定

RAGを活用している場合，回答精度はデータ検索の性能に影響される．LLMがいくら高性能であっても，対象の情報が検索結果の中に入っていなければ適切な回答生成はできないためである．社内のデータベースから最適な情報を検索し，生成AIのパフォーマンスを最大化するために，複数の検索手法を比較・検討し，最適な方法を選択することが望ましい．

3.5.1　検索手法の種類

生成AIの運用において，検索手法にはいくつかのアプローチが存在している．以下に代表的な検索手法として，テキスト検索，ベクトル検索，セマンティック検索を紹介する．

1）テキスト検索

テキスト検索は，伝統的でシンプルな手法である．キーワードを基にデータベース内の文書やデータから一致する内容を検索する．構造化されたデータに対して効果的であり，データが整理されている場合には迅速かつ正確な結果を得ることができる．しかし，

キーワードが正しくないと，意図した情報をうまく引き出せない．また，構造化されていないデータや，長文データに対しては精度が低下する可能性がある．

2）ベクトル検索

ベクトル検索は，テキストデータをベクトル量に変換し，意味的に関連するテキストを検索する．単なるキーワードの一致ではなく，文脈や意味に基づいた検索ができる．特に，曖昧な問い合わせや複雑な質問に対して有効性が高い．ただし，ベクトル検索は計算リソースを多く消費し，実装と運用にコストがかかることが課題となる．

3）セマンティック検索

セマンティック検索は，文章全体の意味を理解して検索を行う手法である．利用者が明確に示さなかった内容についても，文脈から推測し，関連性の高い情報を検索することができる．生成 AI が行う情報検索においては，精度の高い結果を得ることができる．しかし，ベクトル検索と同様，リソースを多く消費し，運用コストが高くなることが多い．

3.5.2　その他検索について検討すべきこと

その他検索について検討すべきことを三つ紹介する．生成 AI を適用する業務の特性によって検討すると良い．

1）チャンク分割

チャンク分割とは大規模な文書を適切なサイズ（チャンク）に分割する手法である．検索対象のドキュメントサイズが大きい場合，生成 AI で許容されるトークン長を超えることがある．検索対象にするために，文書の中で関連する一部分を抽出する．データ量が多い場合は回答精度も低下するため，チャンクに分割して扱うことが多く，ベクトル検索と合わせてよく用いられる．

2）データクローリング

データクローリング（データ修正）の要否について検討する．RAG 対象のデータがログや履歴などのように継続的に増え続ける性質のものであった場合には，対象の Web サイトやデータベースからデータ取得・整形をし続ける（データクローリング）．こうした場合にはデータ取得・整形の自動化が望まれることがある．

3）アクセスコントロール

利用者ごとにデータへのアクセス権限が異なる場合には，個人ごとのデータ参照制御を実装する．合わせて利用者を識別するための認証機能も実装する．

3.5.3　本研究における検索手法の選定

本節で述べてきた検索手法を検討するため，使用するデータを分析した．その結果，以下の特徴があることがわかった．

1）構造化データが多い

2）データがサイロ化されている

3）データロケーションが変化する

今回の実証で対象としたデータは，ナレッジベースやFAQなど，すべて構造化された情報であり，チャンク分割は不要であった．またデータがサイロ化されており，最初にすべてを洗い出すことができず，利用の拡大に伴って対象データが増えていく可能性がある．これらの特徴から，テキスト検索を中心として，あいまい検索機能を持つ全文検索エンジン（以降，インテリジェントサーチ）を利用することとした．

インテリジェントサーチのクローリング機能（データ収集機能）を使用することで，複数のシステムやデータベースにまたがる情報を迅速かつ柔軟に集約し，生成AIがリアルタイムで利用できるようになった．

3.6　検討項目6　機密データの取り扱い

生成AIをIT運用に導入する際，特に慎重に検討すべき要素の一つが機密データの取り扱いである．IT運用に関わるデータには，顧客情報やシステムの詳細など機密性の高いデータが含まれることが多く，どのようにAIに提供し，活用するかが，重要なポイントとなる．データの機密性に応じて対策を講じて，AI導入のリスクを最小化しつつ，生成AIを効果的に活用するための運用ルールを構築した（図6）．

図6　機密データの取り扱い

3.6.1　機密データの分類と取り扱い方針

生成AIの取り扱いに関しては，BIPROGYグループ内でガイドラインが制定されており，従業員はそのガイドラインに従って業務利用している．取り扱うデータを機密レベルに基づいて分類しており，この分類により，どのデータがAIに提供されるべきか，どのデータがさらに厳密な管理の下で扱われるべきかを判断している．機密レベルに応じて，AIがアクセスできるデータ範囲を制限し，データの格納場所はセキュアな領域に限定している．

3.6.2　データ匿名化の適用

機密データを生成AIで活用する際には，データ匿名化などの実施が不可欠となる．用

途に応じて自動匿名化なども検討しているものの，必ずしも万能ではない．最終的には人間が目で確認しなければならない．このように，モデル開発や適用時などでシステムの中に人間を取り込むことはよく行われており，ヒューマン・イン・ザ・ループ（HITL：Human-in-the-Loop）と呼ばれている．

3. 6. 3　ローカル環境の活用

特定の機密データについては，生成 AI がインターネット経由で処理せずローカル環境のみで処理を実施することが望ましい．例えば RAG の検索フェーズ（Retrieval）と生成フェーズ（Generation）をシステム的に分割して，検索までをローカルネットワークで動作するインテリジェントサーチの仕組みで実装し，検索結果を人間が確認して，生成アプリを利用するなどの方法が考えられる．またローカル LLM を活用することで，社内に閉じた環境で生成 AI を運用し，機密データの流出リスクを低減することが賢明である．

ローカル LLM とは完全にローカルネットワーク上でオンプレミスの環境のみで利用する LLM を指しており，オープンソースを含め様々な LLM が存在し，当社で評価を進めている．

3. 7　検討項目 7　開始時期判断と失敗推奨

生成 AI のような新技術の導入においては，計画に時間をかけすぎることで技術の進化に遅れをとるリスクがあり，早期に取り掛かることが重要である．最新技術[*2]をいち早く試していくことにより，現在の課題解決にどの程度有効なのかを継続的にキャッチアップできるようになる．

当社では，アジャイル手法を採用して小さなスコープから素早く導入を開始し，その結果をもとに機能拡張と改善を進めている．その結果，早い段階でフィードバックを得て，技術の精度を向上させることができた．初期段階では想定外の問題が発生することもあり，それを学びとして次に活かすことができるため．失敗もまた貴重な知見として組織に蓄積することができる．以下，当社が実践した際に得られた「失敗知見」の一部を記す．

1）計画に時間をかけ過ぎる

事前計画が重要とはいえ，新技術の実装では不測の事態が生じる可能性がある．柔軟なスケジュール管理と，リスクを吸収できる体制が重要である．

2）どの予算で活動すべきか悩む

組織横断的な業務において，予算の割り当てが難しい場合がある．目的に合わせて予算を分割するなどの対応を検討する．

3）関係者が多く意思決定に時間を要する

関係者が多い場合，合意や報告のプロセスが複雑化し，プロジェクトの実行スピードが低下する可能性がある．計画に余裕を持ち，いたずらに関係者を増やさないことも考慮する．

4）開発重視で実証に時間を割けない

開発に注力しすぎると，実証段階への移行が遅れる可能性がある．開発と実証のバランスを取ることが重要である．

5）途中で追加作業が発生し要員不足になる

開発の過程で追加のタスクが見つかることがあり，要員配置のバランスが崩れる場合がある．特にデータサイエンティストと開発要員が分かれている場合の分担の取り決めが重要である．

4. 生成 AI 導入の検証内容

本章では，生成 AI 適用実証の内容と，業務改善の成果イメージを紹介する．

4. 1 実証内容の紹介

生成 AI を活用したチャットサポートの有効性を検証するため，サポート業務への生成 AI 導入（サポートチャットボット）の実証実験を行った．その概要を表 2 に示す．サポート業務とは，社内外からの問い合わせ対応を主な業務とし，技術支援やトラブルシューティングを含む．具体的には，社内ナレッジを適切に活用し，問い合わせ内容の理解，必要な情報の抽出，対応方法の検討および回答文の生成を行うものである．

当社のサポート業務は，各種サービスの主管部門（スペシャリスト）が該当サービスに関する問い合わせを受けて業務支援を行うほか，顧客向けヘルプデスクのサポートなど，広範囲にわたるため，生成 AI を導入することで業務改善の効果が高いと考えた．

表 2　実証実験の概要

項　目	内　　容
準備期間	3 ヶ月間/実証計画・環境構築及び開発
検証/評価期間	3 ヶ月間/利用部門による評価を実施
使用モデル	Azure Open AI GPT（GPT-3.5Turbo/GPT-4o）
評価目的	チャットサポートの有効性の確認とノウハウ蓄積
評価方法	過去に対応したインシデントを本システム利用と比較
評価項目	1）生成された回答の有用性 2）導入による業務効果 応答時間（TAT; Turn Around Time）の比較 3）利用者満足度
評価者	5 部門 17 チーム
テストケース数	120 種類

4. 2 アプリケーション動作イメージ

アプリケーションの動作イメージを図 7 に示す．アプリケーションは，問い合わせから回答生成までを一連の流れで実施する．与えられた文章を生成 AI へ送付し，結果を出力

しているだけではなく，精度をあげるために工夫を施した．

問い合わせ内容を分析すると，記述が冗長であったり，必要な情報が省略されていたりする場合がある．このため，一度生成 AI を呼び出して①「問い合わせ内容の要約」を行うことで，意図や背景を踏まえた問い合わせ文章に近づけることができる．続いて，要約された文章から検索のための②「キーワードを抽出」する．検索キーワードは，検索対象のナレッジから適切にデータを取り出すために重要であり，特徴的な語であることが望ましい．次に，キーワードを用いて③「社内ナレッジを検索」し，関連文書を複数抽出する．検索で得た文章を問い合わせ文章と合わせて，回答要素になりうるかどうか，生成 AI を用いて定量評価した．最後に，採点結果のスコアを用いて，ランキングの高い方から数件取り出して，④「回答を生成」する．何件取り出すのか，スコアの足切りを行うのか，については精度向上のチューニングの対象となる．

RAG の対象データとしては社内ナレッジの公開可能データのみを対象とし，機密とすべき顧客情報を含むナレッジは使用しなかった．導入当初のモデルは GPT-3.5turbo を使用し，実証途中で GPT-4o が利用できるようになったため，アルゴリズムを変更した．次章では，GPT-3.5turbo と GPT-4o を用いた結果を併記してその違いにも言及する．

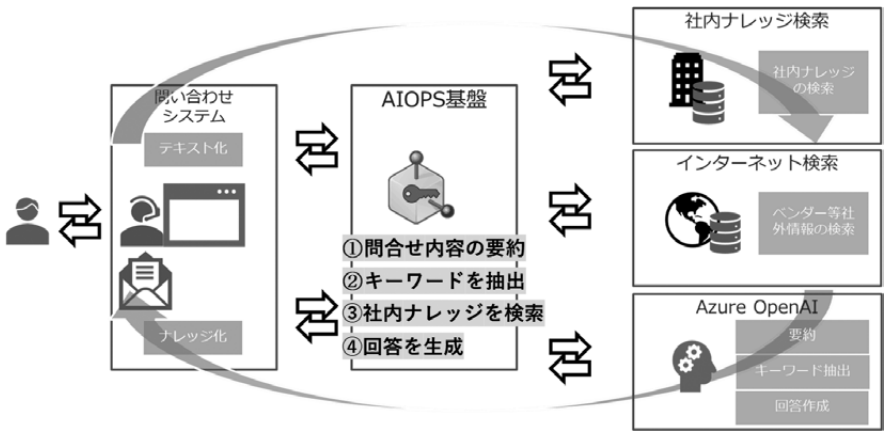

図 7　実証実験アーキテクチャ概要

4. 3　実際の業務改善イメージ

利用者は，顧客や社内からの技術問い合わせ先である当社の社員である．二次サポート（窓口からのエスカレーション先）と呼ばれる担当エンジニアを想定している．彼らは問題の切り分けから解決までを担当する．

図 8 のアプリケーション画面に示すように，画面左上には顧客や社内からの問い合わせ入力欄があり，状況に応じてインシデント管理ツールと連携し問い合わせ内容を入力することができる．入力された文章は，送信ボタンを押すと生成 AI に送信され，回答ドラフトが生成されて表示される仕組みである．

回答生成には通常数分を要するため，進捗表示があり，例えば「ナレッジベース検索中」

や「検索結果を用いた回答生成中」など，システムが内部でどの段階にあるかが表示される．また，回答作成後は途中経過の入出力をいつでも確認でき，回答精度が思わしくなかった場合に，どこに問題があったのかを振り返ることができる．

図 8　入力画面例

回答欄には問い合わせ内容を要約した文章とともに，回答案が表示される（図 9）．回答案はドラフトとして扱われ，エンジニアが実際に回答として使えるかどうかを確認する．さらに，回答案の下には，検索して見つかった社内ナレッジやインターネット検索結果のリンクが表示される．回答に利用したデータをリンクから確認し，それが正しい情報であるかどうか，適切な情報であるかどうかを判断するための材料とすることができる．

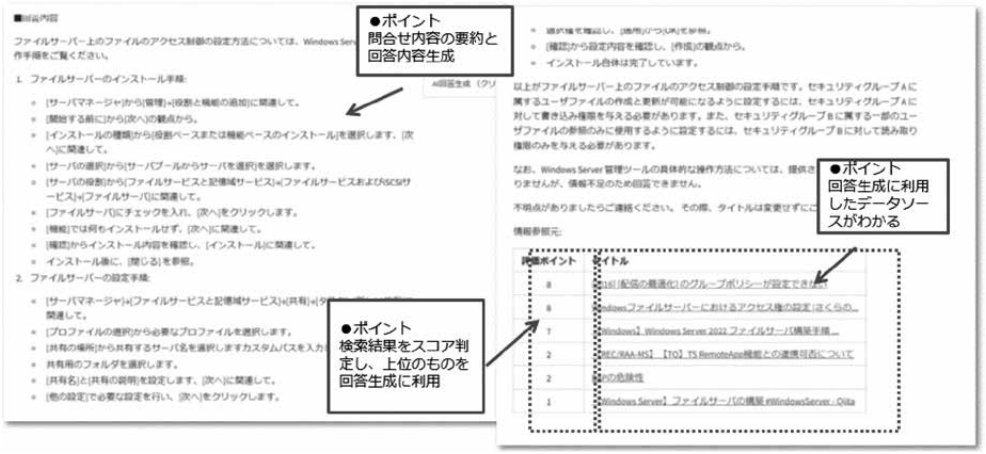

図 9　出力画面例

5. 生成 AI 導入の検証結果

本章では実証の成果を以下 1）～ 3）の評価指標に沿って各節で述べる.

1）生成された回答の有用性（回答が業務に使えるかどうか）

2）導入による業務効果（対応時間がどの程度短縮できるか）

3）利用者満足度（今後も業務で利用したいかどうか）

5.1 生成された回答の有用性（回答が業務に使えるかどうか）

生成 AI により作成された回答がどの程度業務に活用できるかについて，テストケースごとに 5 段階で評価を行った（表 3）. 期待を超えて詳細な原因や調査のために実施すべき項目やコマンドなど，追加の情報が得られた場合を 5 点とした. 回答として利用できる内容が揃っている場合を 4 点，一般的な情報が得られた場合を 3 点，回答として十分な内容が得られなかった場合は 2 点. 殆ど使用できる部分がない場合を 1 点とした.

表 3　生成された回答の有用性評価

採点	内　　容
5 点	期待した回答に加えて，詳細の情報が得られた
4 点	期待した回答内容が得られた
3 点	一般的な回答は得られた
2 点	回答として不十分
1 点	回答として使用できる部分がない

業務で有用とされる水準を 3 以上とし，評価結果を確認した. GPT-3.5 では業務水準を下回る結果が多く，GPT-4o では業務水準を上回る結果が多かった. 集計すると GPT-3.5 は業務で利用可能なテストケースが約 30％に留まるのに対して，GPT-4o では 55％のテストケースに対して有効な回答が得られた（図 10）.

評価が低かったケースを確認すると，50％程度は情報不足が原因であり，今回対象外とした機密情報を含むデータが有用であることがわかった. 今後検索対象を増やすことで解決される可能性がある.

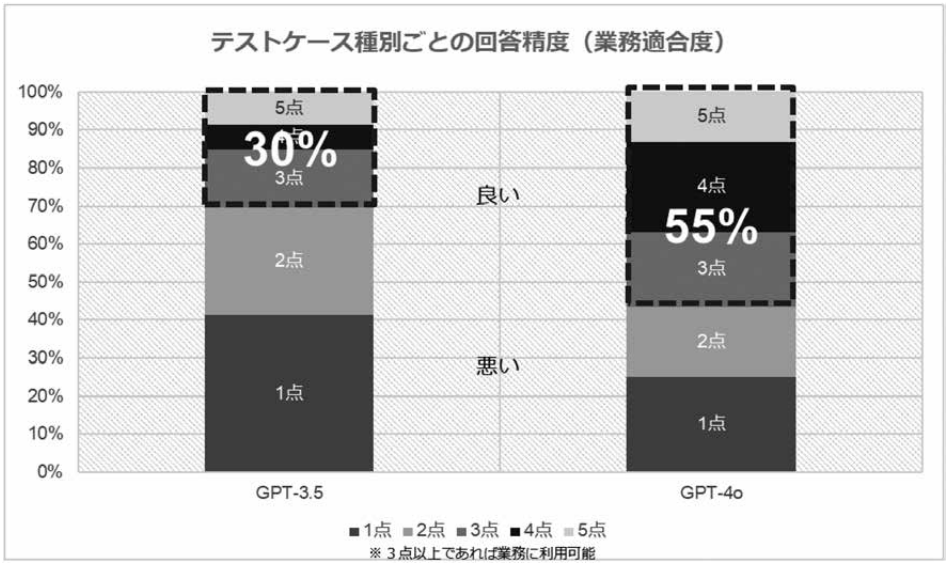

図 10　回答精度の業務適用度

5. 2　導入による業務効果（対応時間がどの程度短縮できるか）

生成 AI の活用によって，業務プロセスにおける工数削減を確認できた．従来の手法と比較して回答作成時間の改善を検証した結果，半数以上のテストケースで生産性向上が見られ，回答ドラフト作成までの対応時間は従来の 1/10 から 1/100 程度に短縮された．

ただし，生成 AI が作成した回答は，内容の正確性の確認や顧客向けに適した文章表現への修正に時間を要するものがあった．回答精度が低いケース（前節の評価点 1 ～ 2 点とされたもの）は，回答時間の短縮効果が見られなかったものとして評価した．

5. 3　利用者満足度（今後も業務で利用したいかどうか）

利用者が実際に使い続けたいかどうかの評価を実施した．本研究の結果は図 11 に示すように，2/3 の利用者が続けて利用したいと回答している一方で，現状では利用したくないと回答した利用者は 1/3 であった．利用したくない意見としては，良い結果が得られない，文章が作成されたが待ち時間が長い，UI が使いづらい，といったものが含まれていた．今後，利用者からの改善要望を確認して，アプリケーションを改善していく予定である．

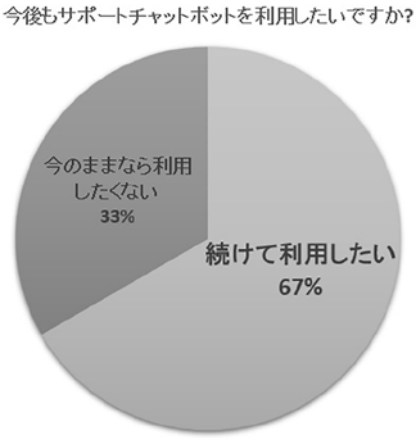

図 11　利用者満足度

6.　今後の展望

本実証を通じて，生成 AI の導入が IT 運用に対して変化をもたらすことを確認できた．また，AI 領域での技術進化は今後も継続する見通しである．当社では，生成 AI のさらなる活用に向けた計画を進めており，業務の高度化と効率化を図ることを目指している．

6. 1　問題の解析と回答精度改善

今回の実証において，生成 AI が十分な精度で回答を生成できなかった半数のテストケースに対して，詳細な解析を実施した．今後は，解析結果に基づいて，生成 AI のアルゴリズムやプロンプトエンジニアリングの最適化を図り，回答精度の向上を目指す．また，利用者から寄せられたフィードバックや機能改善要望を収集し，それらを整理した上で，新たな機能を実装する計画である．例えば，生成 AI の回答プロセスに対するリアルタイムのフィードバックループを設け，利用者ニーズに応じた改善をシステムに反映させる仕組みを検討している．これを導入することで，より実用性の高いシステムに進化させることができると考えている．

6. 2　機密データを取り扱った RAG の仕組みの提供とローカル LLM

機密データを安全に活用するための仕組みを構築し，業務改善をさらに進めることが今後の重要課題となる．現在，その多くがクラウド環境で提供されている生成 AI について，セキュリティ上の理由から，ローカル環境での運用を検討している．ローカル LLM の活用は，機密情報を含むデータの取り扱いにおいて大きな利点を持つ．オンプレミス環境で LLM を運用することで，機密データが外部に流出するリスクを最小限に抑えつつ，生成 AI の能力を引き出すことができる．

ローカル LLM は，特に機密情報を含むデータの取り扱いにおいて，優れたセキュリティを提供することが期待されている．今後，オンプレミス環境での LLM 運用により，クラウドベースのリスクを回避しつつ，AI の恩恵を享受できるようになる．

6. 3　ベンチマーク用のテストケースの充実

アプリケーションのバージョンアップ時に毎回異なるテストケースが使用されるため，効率が悪い．この課題に対処するため，現在運用されているテストケースを拡張し，標準化されたベンチマークとして整備する計画を検討している．具体的には，適切に変換されたサンプルを整理し，生成 AI のパフォーマンスを評価するための統一基準を設ける．

これにより，アプリケーションやアルゴリズムのバージョンアップのたびに同じ基準で生成 AI の能力を測定でき，改善点を迅速に特定することができる．さらに，ベンチマーク結果をもとに，生成 AI の精度や応答速度を向上させることで，より高度な IT 運用支援ツールとしての完成度を高めることが期待される．

7. お　わ　り　に

本稿では，IT 運用の高度化と生産性向上を目的に生成 AI を活用し，導入時の検討ポイントと実証成果を示した．サポート業務におけるユーザーからの問い合わせ対応に生成 AI を取り入れることで，10 倍～最大 100 倍程度の生産性向上が期待できる結果を得た．今後は，さらに多くのナレッジ活用を推進するため，ローカル LLM の活用を含め，機密データの取り扱いに関する仕組みを強化する予定である．また，継続的に新しい技術を取り入れ，生成 AI の性能向上や新たなテストケースの発見に努め，より一層の業務効率化とコスト削減を目指す．

AI 技術の進化に伴い，将来的には 2. 2 節（図 1）に示すように，内外の多様なデータを AI が統合的に分析し，深い洞察を導き出すことで，ユーザーシステムの安全性と最適性を一層高める支援を実現することを目指している．

実証実験を通じて得られた知見が，生成 AI の導入を検討している他の企業にとっても有益な指針になれば幸いである．生成 AI の可能性は，今後さらに拡大して多くの業務分野で利便性の活用が進むと想定される．当社においても生成 AI を活用し，顧客の業務を支える新たな価値創造を継続していく．

本稿の執筆にあたり，実証協力や技術支援をいただいた多くの関係者に深く感謝申し上げる．

＊ 1　2023 年 4 月の実証研究当時は，2022 年 11 月 30 日に OpenAI 社が ChatGPT を発表し，続いて 2023 年 3 月に OpenAI 社の AI モデルが Azure に実装されて Azure Open AI Service（AOAI）が発表されたところであり，まだ他の有力な選択候補が存在しなかった．現在であれば，他の選択肢も候補として検討すべきである．

＊ 2　SOTA（State-of-the-Art）モデルと呼ばれる最先端のモデルなど．

＊ 3　機械倫理（Machine Ethics）とは，機械や人工知能（AI）の開発や使用において，倫理的観点を尊重するための指針や原則．機械や AI による回答や回答を導く判断に，倫理的な問題が生じる可能性に対応し，機械や AI の行動を制限するもの．（Brundage, M. (2014). Limitations and risks of machine ethics. Journal of Experimental & Theoretical Artificial Intelligence, 26(3), 355-372. https://doi.org/10.1080/0952813X.2014.895108）

参考文献
［1］ 総務省，令和 6 年版 情報通信白書，2024 年 7 月
［2］ 経済産業省，AI 事業者ガイドライン（第 1.0 版），2024 年 4 月
［3］ IDC，Generative AI Drives Significant Impact to IT Operations in Asia/Pacific, According to IDC，2023 年 11 月，https://www.idc.com/getdoc.jsp?containerId=prAP51365723
［4］ 杉山義治，社内業務への生成 AI 適用事例，BIPROGY 技報，BIPROGY，Vol.43 No.4 通巻 159 号，2024 年 3 月，https://www.biprogy.com/pdf/15903.pdf
［5］ GovTech，Master IT Complexity with Observability Solutions，2024 年 10 月，https://www.govtech.com/sponsored/master-it-complexity-with-observability-solutions
［6］ 日経クロステック，「Spark」と機械学習で匠の技を再現，12 日前に機器障害を検知し予知保全に活用，株式会社日経 BP，2015 年 11 月，https://xtech.nikkei.com/it/atcl/column/14/090100053/112000106/
［7］ 藤田勝貫，ICT インフラの高度化に向けた取り組みと AI 活用のポイント，ユニシス技報，日本ユニシス，Vol.39 No.1 通巻 140 号，2019 年 6 月，https://www.biprogy.com/pdf/tec_info/14004.pdf
［8］ マイクロソフト，Azure OpenAI Service リファレンスアーキテクチャ，2023 年 6 月，https://www.microsoft.com/ja-jp/events/azurebase/contents/
［9］ Yue Huang 他，「TrustLLM: Trustworthiness in Large Language Models」，Cornell University，2024 年 8 月，https://arxiv.org/abs/2401.05561
［10］ Data Analytics Lab，LLM のファインチューニングを他手法との違いから理解する（Part 1），2024 年 3 月 7 日，https://dalab.jp/archives/journal/llm-finetuning-part1/
［11］ Oded Ovadia 他，Fine-Tuning or Retrieval? Comparing Knowledge Injection in LLMs，Cornell University，2024 年 1 月，https://arxiv.org/abs/2312.05934

※ 上記注釈および参考文献に含まれる URL のリンク先は 2024 年 12 月 9 日時点での存在を確認

執筆者紹介 藤　田　勝　貫（Masatsugu Fujita）
2005 年にユニアデックス株式会社に入社．ソフトウエアプロダクトの企画・開発および顧客システムの開発に従事．その後，2014 年より AI のビジネス活用に関する研究開発に従事し，現在は社内外の AI 適用プロジェクトやビジネス化に関する業務を担当している．