

令和 5 年度
産業経済研究委託事業
(行政事務の高度化に資するLLM導入可能性検証調査)

調査報告書



Build Beyond As One.

1. 調査事業概要及び政策目的

- 1.1. 概要及び目的
- 1.2. エグゼグティブサマリ

2. 調査事業のアプローチ

- 2.1. 作業全体スケジュール
- 2.2. 成果物

3. ユースケース整理及び検証環境構築

- 3.1. 検証するユースケースの整理
- 3.2. 検証環境構築に向けた要件整理
- 3.3. 検証環境の構築
- 3.4. 検証環境の基本機能
- 3.5. 利用者研修の実施

4. ユースケース検証及び次期課題の整理

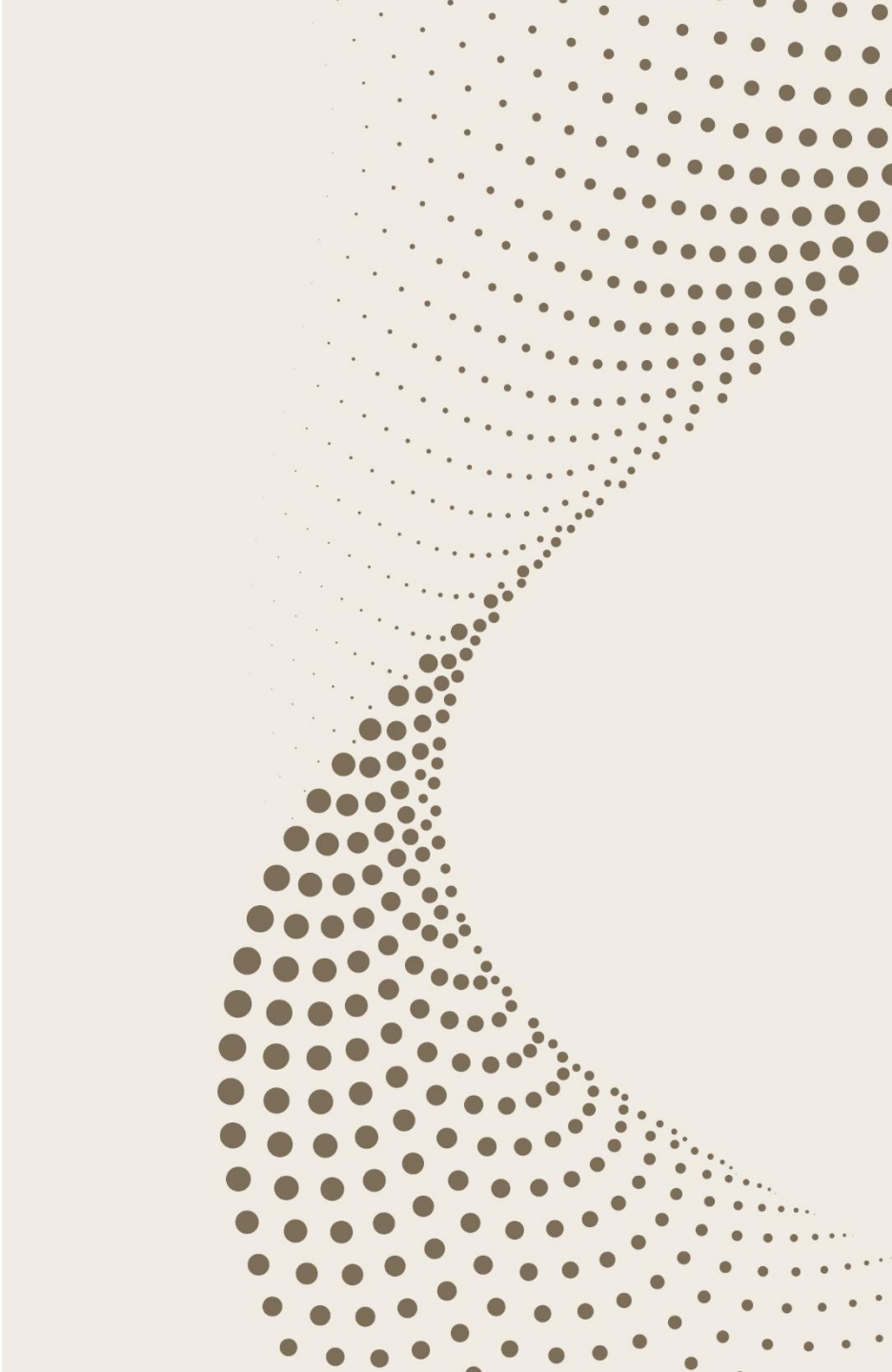
- 4.1. 検証の実施方針
- 4.2. 検証結果の評価
- 4.3. 重要検討課題の整理
- 4.4. 今後のシステムのあるべき姿

5. 生成AI利活用を前提としたデータに関する調査結果

- 5.1. 非構造データの利活用とデータマネジメントの目的
- 5.2. 機械判読性の調査
- 5.3. 権限管理とセキュリティ
- 5.4. 持続可能な運用の仕組み

6. 本調査事業の総括

- 6.1. 総括



1. 調査事業概要及び政策目的

1.1. 概要及び目的

1.2. エグゼグティブサマリ

1. 調査事業概要及び政策目的

1.1. 概要及び目的

実施目的

行政事務の高度化に資する技術であるかの適用可能性を検証するとともに、当該技術を導入するにあたっての具備すべき要件及び制約事項等を洗い出すこと

令和5年度産業経済研究委託事業（行政事務の高度化に資するLLM導入可能性検証調査）仕様書より

実施概要



行政事務における ユースケースの分析調査

概要

- 収集した各課室の検証案に関して、生成AIの適用可否が明らかなケースを仕分けし、構築する検証環境の方針を検討する

主な実施内容

- ユースケースの洗い出し
- 生成AIの活用可否のスクリーニング
- ユースケースのカテゴライズ
- 具備すべき機能の実装検討



生成AIの検証環境構築及び 個別ユースケースの検証

概要

実際に職員が使用できる検証環境を整備し、生成AI技術が行政事務の高度化に資するかどうか100名程度の職員で検証する

主な実施内容

- 生成AIの検証環境の実装
- 検証結果の集計
- 検証結果から課題の抽出
- 検証環境の運用改善



次年度本格導入時の 制約事項の整理

概要

検証から得られた知見から当該技術を本格的に導入するにあたり、行政事務特有の具備すべき要件及び制約事項を整理する

主な実施内容

- 検証で出た課題から要件の洗い出し
- 要件を満たす技術の調査
- 制約事項の整理

1. 調査事業概要及び政策目的

1.2. エグゼグティブサマリ



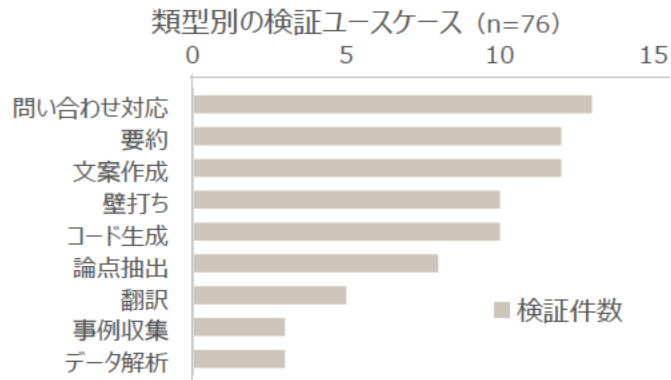
行政事務における ユースケースの分析調査

第3章
第4章
第5章

- 各原課の検証案からユースケースを類型化
- 検証案を基に事務局との協議を重ね、検証環境の要件を検討し、計画・実装

ユースケースの類型化

9種類の類型を定義し、検証対象を明確化



検証環境の要件検討、迅速な計画・実装

検証期間を多く確保する為、迅速に環境構築を推進



検証環境の協議



要件の整理



実装



生成AIの検証環境構築及び 個別ユースケースの検証

第3章
第4章
第5章

- 構築した検証環境を用いて各原課にて検証
- 検証結果から類型別の行政業務への適用可能性を評価

検証環境の構築

必要な機能を整理し、5つの主要機能を具備。継続的に運用も改善

構築した検証環境の画面



実装した主要機能

- Generic Chat: 一般的なチャット機能
- URL Answering: URL先の参照、回答
- PDF Answering: PDFの参照、回答
- AI Search: 大量文書の検索、回答

行政事務の高度化に向けた適用可能性の評価

会話履歴、事後アンケートの定性・定量分析により多面的に評価

適用可能性の評価

類型

業務高度化に資する可能性が高い 翻訳/コード生成/要約

工数削減効果は低いが業務高度化が見込める データ解析/文案作成/壁打ち/事例収集

業務への利活用が限定的となる可能性が高い

問い合わせ対応/論点抽出



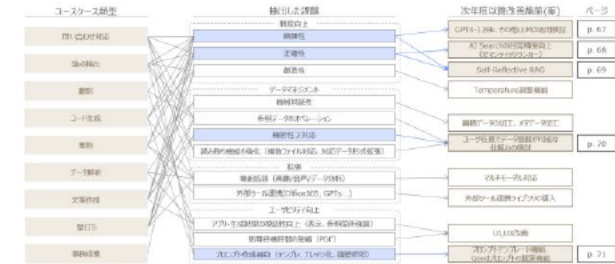
次年度本格導入時の 制約事項の整理

第3章
第4章
第5章

- 抽出した課題及び施策案の検討
- 本格導入する際の制約事項をデータマネジメントと持続可能性の観点に留意し、整理

課題の抽出と施策案の検討

類型別に課題を抽出・分類。優先度の高い課題に対し、施策立案



データマネジメントと持続可能な仕組みの検討

省内に埋もれている資料を政策立案に利活用することを目指す

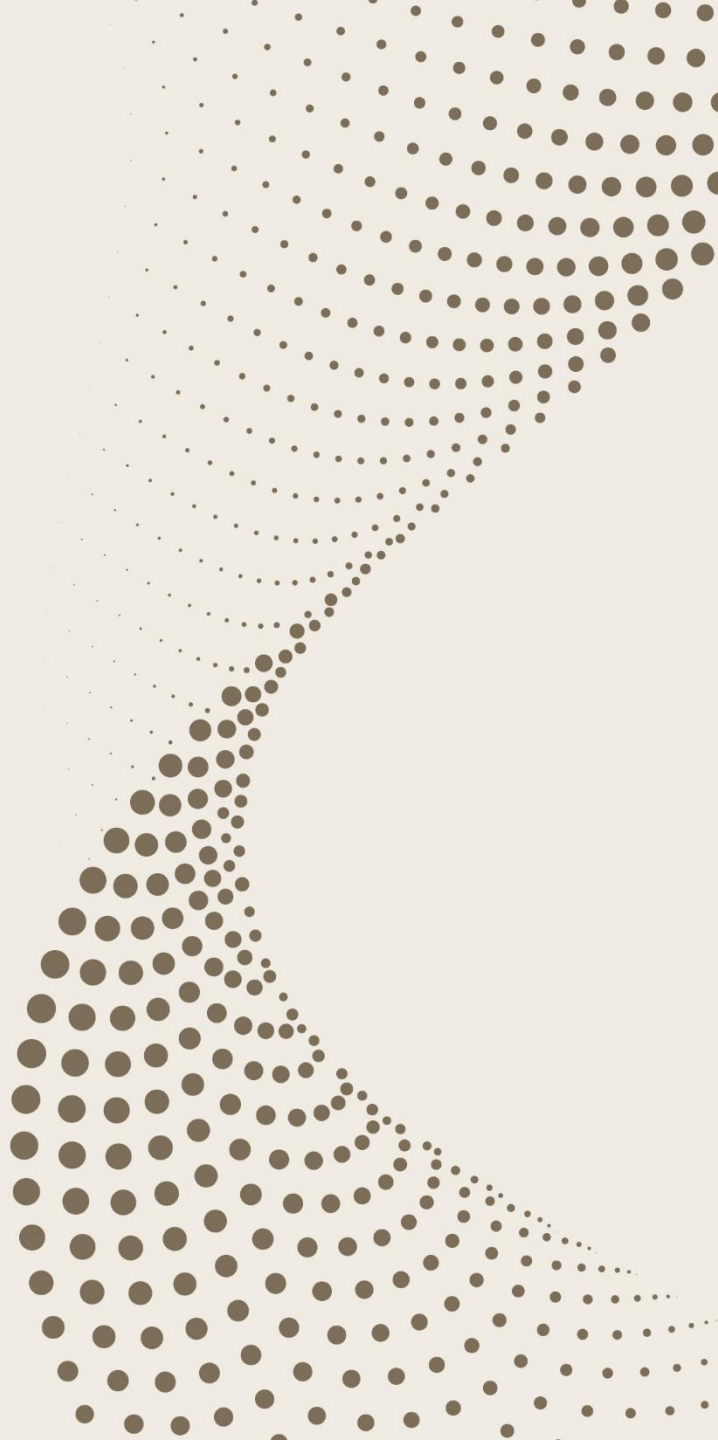
機械判読性

権限管理/セキュリティ

持続可能な運用

- 機械判読性の調査
- 生成AI利活用時の
良し悪しを明確化
- 管理要件の整理
- 権限管理の仕組み
の検討
- 迅速かつ確実な
プロセス自動化を
検討





2. 調査事業のアプローチ

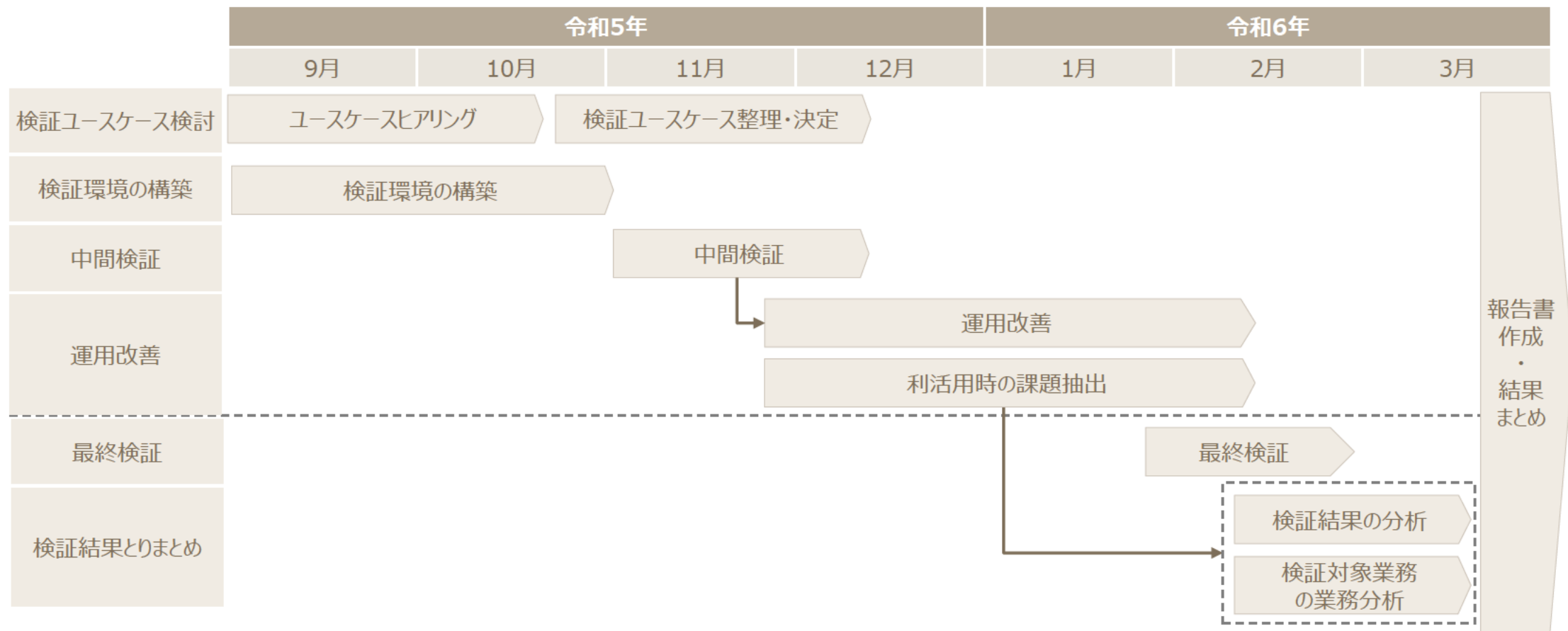
2.1. 作業全体スケジュール

2.2. 成果物

2. 調査事業のアプローチ

2.1. 作業全体スケジュール

検証環境の構築を先行実施し、各課検証ユースケースに即した機能改善、業務分析を実施することで、本事業の検証を効率的に実施した



2. 調査事業のアプローチ

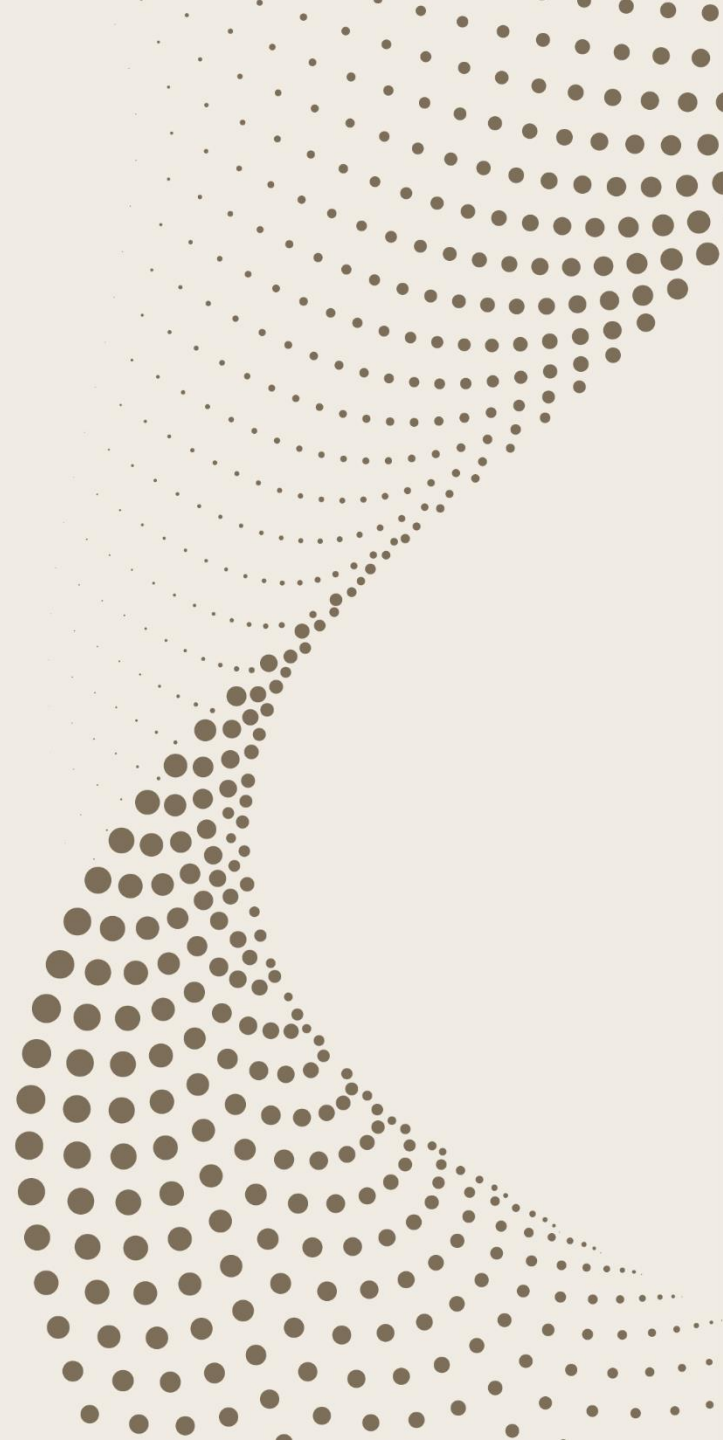
2.2. 成果物

本事業で作成した成果物は以下の通り

	非公表用
業務分析	03_01_各ユースケースの課題整理・分析.xlsx
検証	04_01_集計結果一覧.xlsx 04_02_事後アンケート集計結果.xlsx
利用者研修	07_01_利用者研修資料.pptx 07_02_生成AIガイドライン素案.pptx
調査報告	01_01_調査報告書.pptx 01_02_エグゼクティブサマリ.docx
設計資料・ソースコード	02_01_シーケンス図.xlsx 02_02_画面遷移図.xlsx 02_03_テーブル項目定義-ER図.xlsx 02_04_パラメータシート.xlsx 02_05_環境構成図.xlsx 02_06_プログラム一式.zip 02_07_生成データ.xlsx

3. ユースケース整理及び検証環境構築

- 3.1. 検証するユースケースの整理
- 3.2. 検証環境構築に向けた要件整理
- 3.3. 検証環境の構築
- 3.4. 検証環境の基本機能
- 3.5. 利用者研修の実施



3. ユースケース整理及び検証環境構築

3.1. 検証するユースケースの整理 分析と環境構築の進め方

検証ユースケースの洗い出し、具備すべき機能の仕分けを実施。その上で、検証結果の分析から改善施策検討まで実施した

各作業における実施内容



3. ユースケース整理及び検証環境構築

3.1. 検証するユースケースの整理 ユースケースの類型化

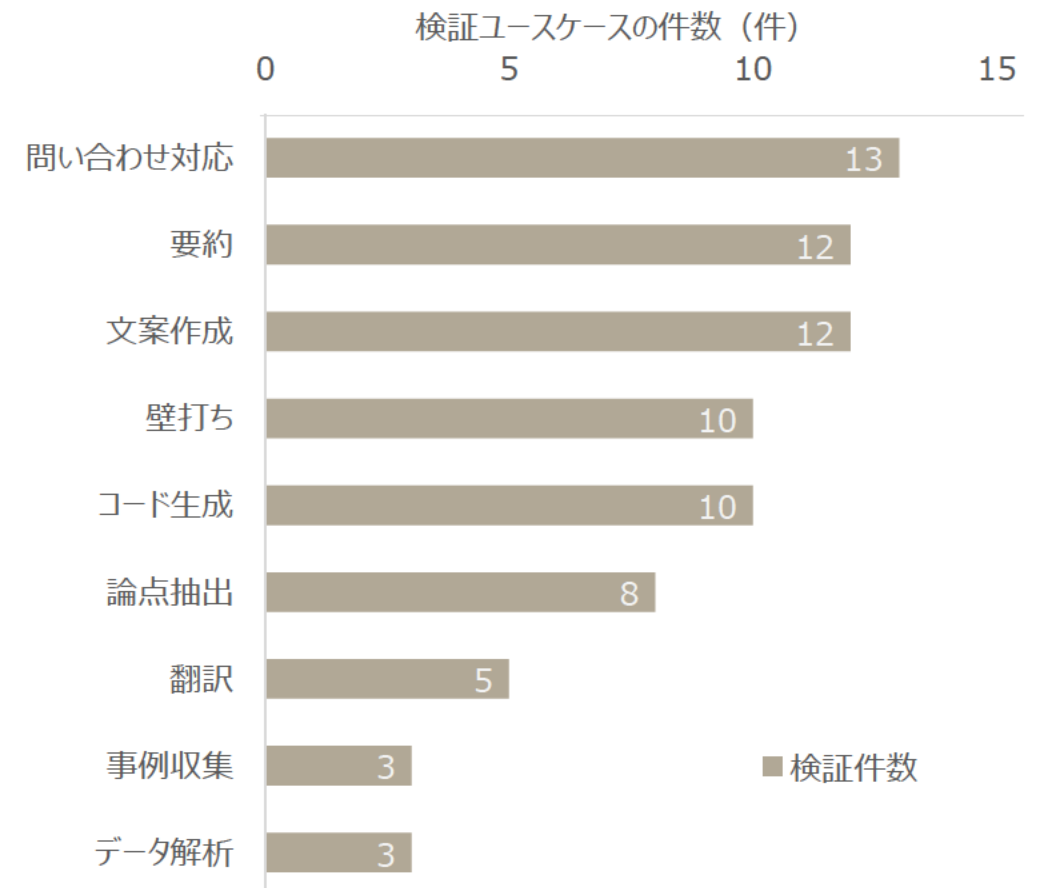
STEP2

検証ユースケースの
類型化

ユースケースを類型化した上で、省内100名程度の職員による検証を実施
対象となるユースケースの件数を類型別に集計した

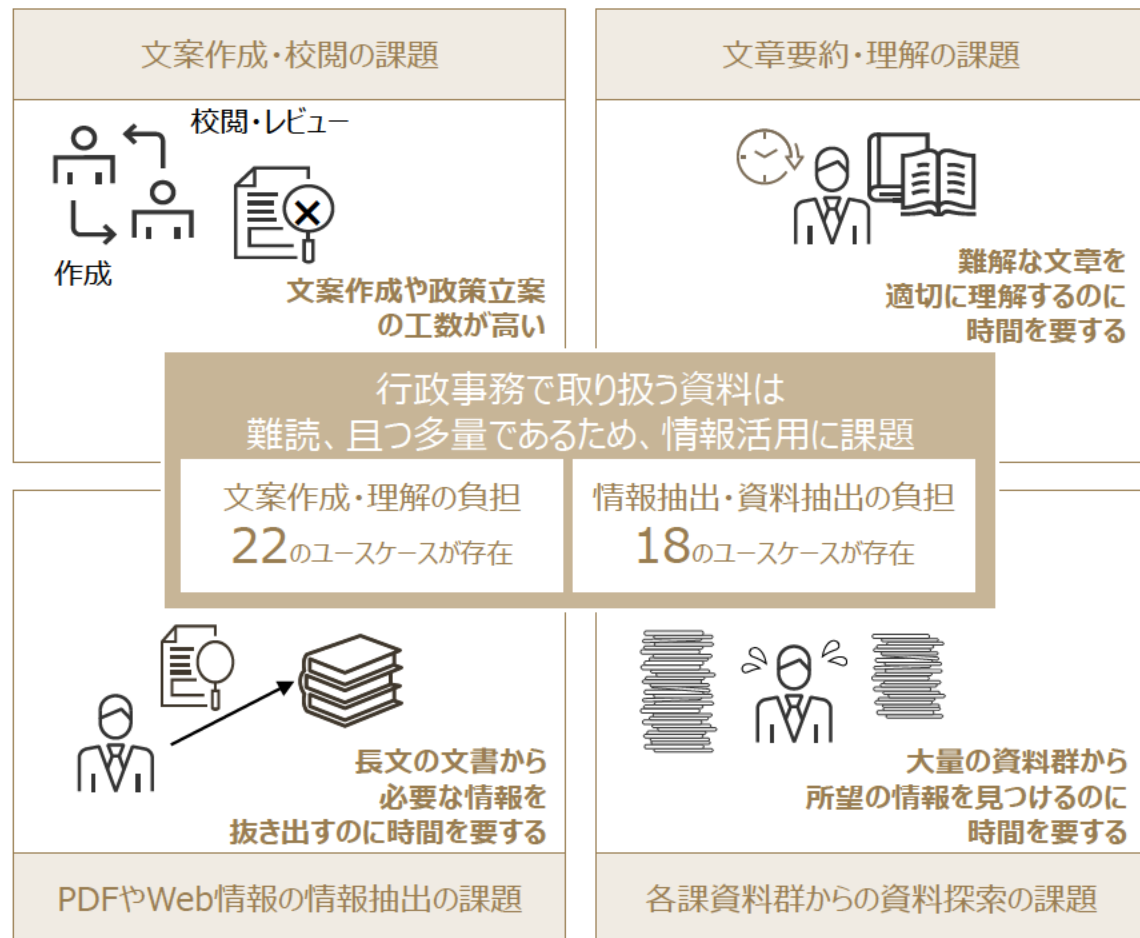
No	ユースケースの類型	
1	問い合わせ対応	外部/内部からの問い合わせ対応として、問い合わせに対する適切な部署や担当者の出力や、回答案の作成を行う
2	文案作成	報告書、プレスリリース、ウェブサイト等各種文書のコンテンツ案を作成する
3	要約	長い文書や報告書を短く、要点を抽出した形で提供する
4	翻訳	外国語の文書を日本語に、または日本語の文書を外国語に翻訳する
5	論点抽出	大量の文書情報から主要な論点や議論の要点を抽出する
6	壁打ち	政策に関する新しいアイデアや提案を複数出す
7	コード生成	特定のタスクを自動化するためのコードを生成する。例えば、Excelマクロ、Power BI等
8	事例収集	特定のテーマや問題に関連する事例を収集し、分析や報告のための資料案とする
9	データ解析	収集したデータを解析し、パターン、トレンド、洞察を抽出する

検証ユースケースの類型 (n=76)

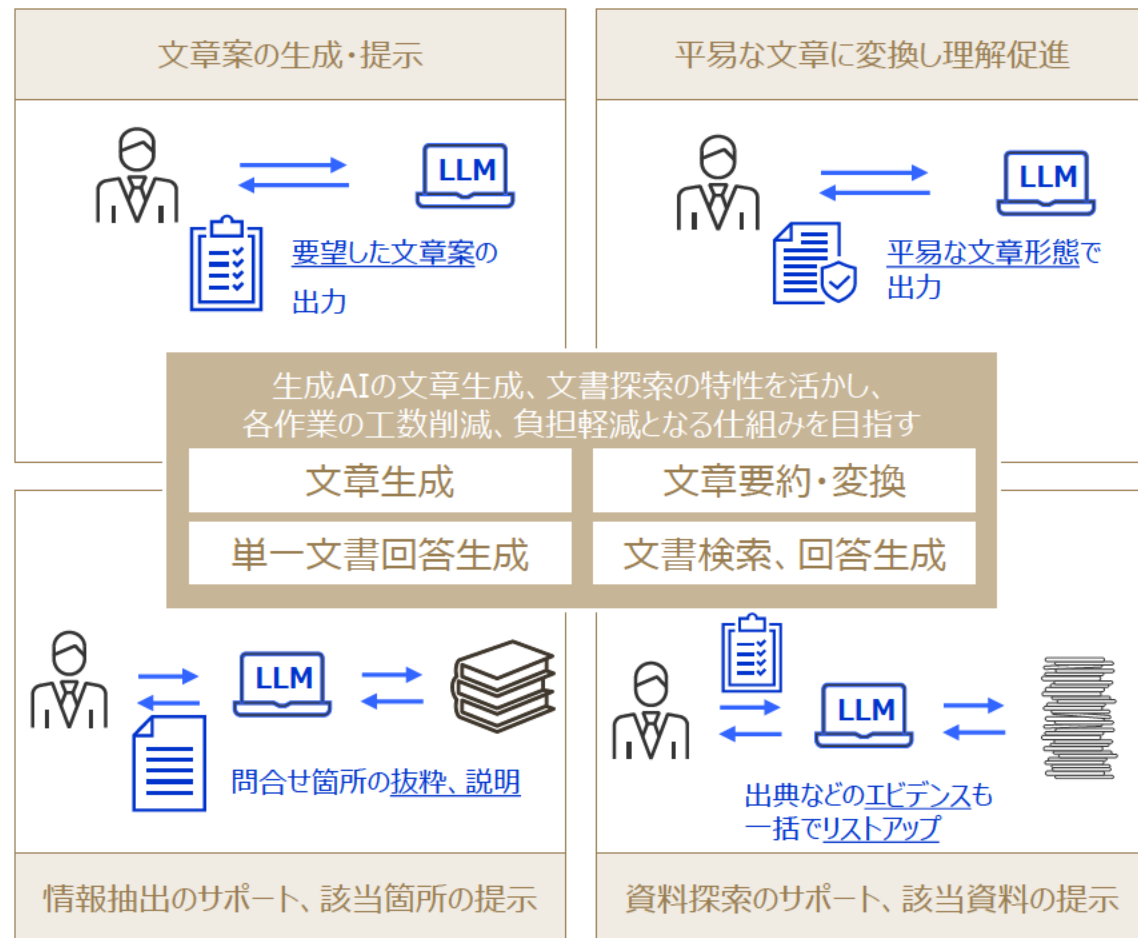


本検証において各課室から集約した、行政事務の高度化が期待されるユースケースに対するAs-Is、To-Beを示す

As-Is（課題となる現状）



To-Be（目指す姿）



検証ユースケースの整理した結果を基に、検証環境に求められるシステム要件を整理した

文章案の生成・提示

Ex. ○○テーマのTwitter投稿文を書いて

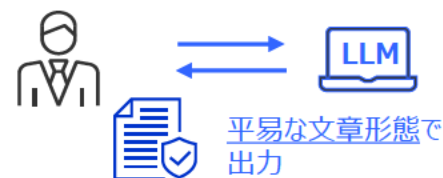
OpenAIの新技术、動画生成AI「Sora」が映画市場にどんな影響を与えるか考えてみました！映画制作のプロセスが一変するかも！
#OpenAI #Sora #映画市場



平易な文章に変換し理解促進

Ex. 衆議院と参議院の違いを整理して、表形式でまとめて

項目	衆議院	参議院
定数	465名	245名
選挙区	小選挙区比例代表並立制	選挙区選出と全国比例代表の並立制
任期	4年	6年



情報抽出のサポート、該当箇所の提示

国会議事録のPDFに対して質問したい
Ex. 岸田首相はどのような発言をされてますか？

少子化対策として発表した「こども未来戦略方針」について言及しており、財源に課題があると発言しています。

ページ番号	文章
3	xxx
4	yyy
...	...



資料探索のサポート、該当資料の提示

中小企業白書全体に対し質問したい
Ex. 中小企業白書の最新版を連携して

最新版は以下URLにあります。
[中小企業白書](#)
こちらご確認ください。

ファイル名	文章
中小企業白書xx	xxx
中小企業白書yy	yyy
...	...



ユースケース検証に満たす機能の要件を整理し、複数機能に切り分け、開発の方針を決定した

No	取り組みの方向性	機能名	検討すべき ユースケース数
1	GPT-4と同等のチャット機能	Generic Chat	42
2	URL先の内容を参照し、GPT-4で回答させる	URL Answering	9
3	PDFファイルの内容を参照し、GPT-4で回答させる	PDF Answering	28
4	事前に大量の文章の登録、検索・参照し、GPT-4で回答させる	AI Search	20

構築した検証環境のシステムアーキテクチャの概要を示す

環境構成

1 ユーザー認証

クラウドサービスの認証システムを利用し、ユーザー認証を行う

2 回答生成

ユーザーが投入したプロンプト/データに対して、生成AIがウェブアプリ上で回答を生成する

3 データ検索

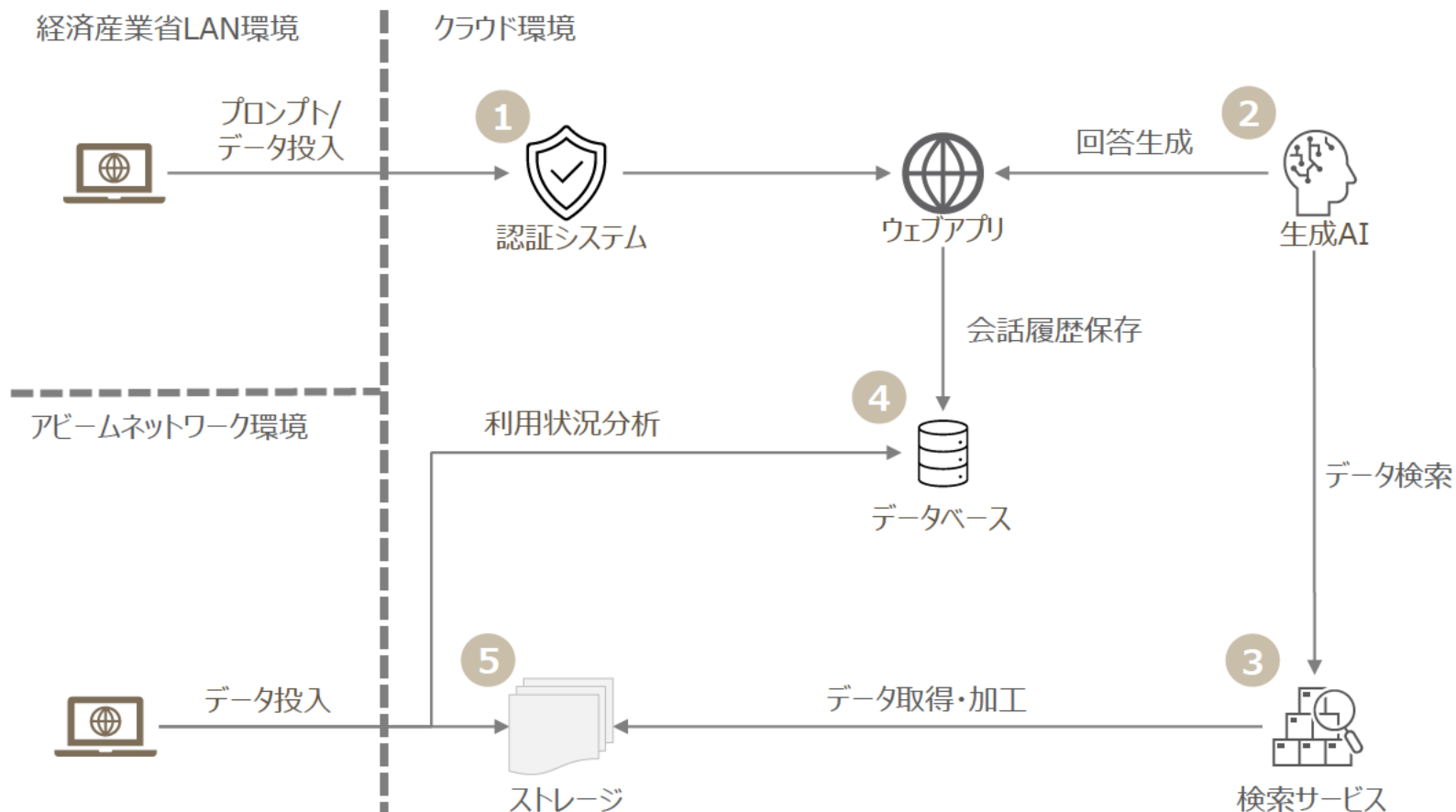
生成AIと検索サービスを利用してデータを検索する。クラウドストレージ内のデータを予め検索用に加工しておくことで効率的な検索が可能となる

4 利用状況分析

ユーザーの会話履歴をデータベースに蓄積する。管理者がデータベースにアクセスし、利用状況を分析する

5 データ管理

容量の大きいファイルは事前にクラウドストレージに格納し、検索可能な状態に加工する



3.3. 検証環境の構築 (1) Generic Chat / URL Answering / PDF Answering機能概要

汎用的なチャットUI画面を意識したUI設計を実施し、検証に向けて下記環境を準備した

- 1 用途に応じて使用する機能を選択できる
 - **Generic Chat** : 汎用的な質問を生成AIにすることが可能
 - **URL Answering** : 調べたいURLの範囲に限定した質問が可能
 - **PDF Answering** : 調べたいPDFの範囲に限定した質問が可能

- 2 質問内容を質問入力欄に記載 →  をクリック
 - 質問内容がチャットエリアに表示され、AIから返答
 - 続けて質問を入力し、会話を継続することが可能

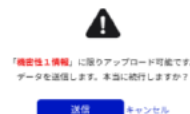
URL Answering/PDF Answeringの場合 :

- 1) 会話を開始する前に各々URLとPDFファイルをアップロード

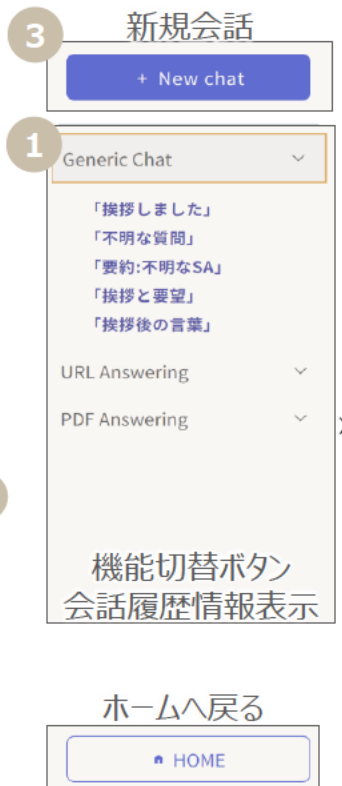
URLの場合 :

PDFの場合 :

- 2) アップロード時は下記ポップアップ画面が出るので、機密性がないデータであることを確認し、データを送信

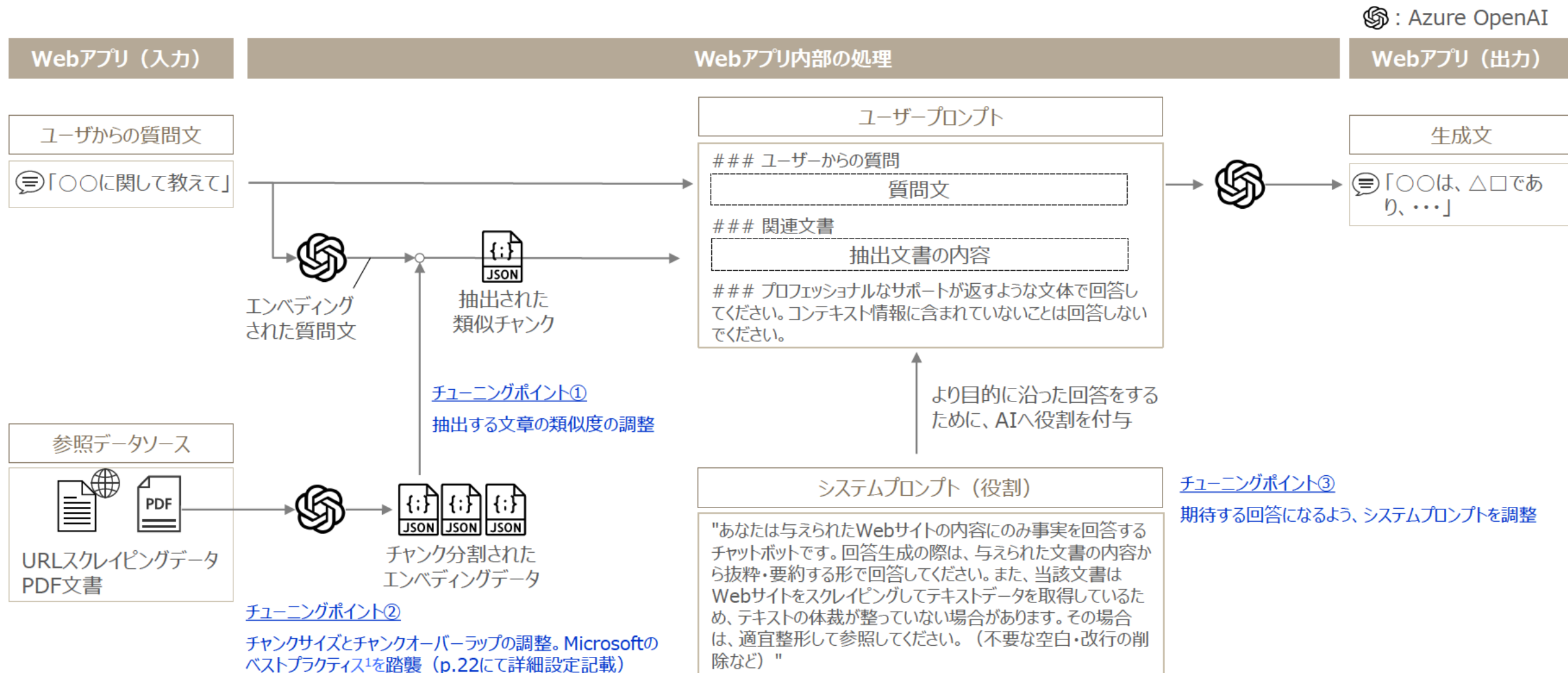


- 3 新しい会話を開始する場合は、 をクリック
 - チャット表示エリアの会話がクリアされ、新規に会話を開始可能



3.3. 検証環境の構築 (2) URL Answering / PDF Answeringの回答生成の処理フロー

URL / PDF Answeringにおける回答生成プロセスの処理フローを以下に示す

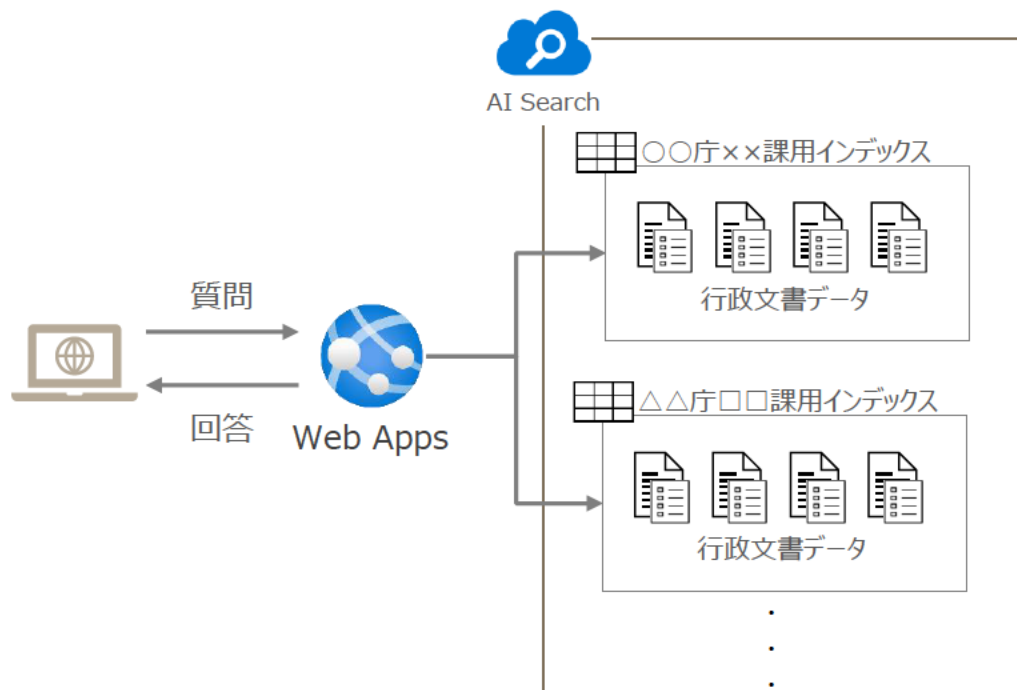


1. <https://speakerdeck.com/hirosatogamo/chatgpt-azure-openai-da-quan?slide=121>

Azure AI Searchに事前登録した大量のデータから、生成AIによる関連情報の検索及び回答を生成させる機能を導入した

概要説明

- 省内データ事前に解析し、インデックス化（目次化）することで、素早い検索を可能にする
- 検索精度を高めるための複数のオプションも用意されている
- Word、ExcelやPPT等、幅広いドキュメント形式に対応可能（pptx / ppt, pdf, xlsx / xls, docx / doc, txt, csv）



提供画面

+ New chat

検索対象の選択 決定

会話履歴



注意事項

1. データのアップロードに関して

ー ルールとして、「機密性1の情報」に限りアップロード可能としていますので、厳守してください。

2. データ回答生成サービスの利用に関して

ー 利用終了時、アップロードしたデータをクリアするために、「+New chat」または「HOME」ボタンをクリックしてください。

※ アップロードしたデータが残ると、次回利用時に予期しない回答が生成される可能性があります。必ずクリアを行ってから次のサービスを利用してください。

HOME

メッセージを入力して [Enter] で送信、[Shift] + [Enter] で改行します。

chatbot may produce inaccurate information about people, places, or facts.

GPT-4

3.3. 検証環境の構築 (4) Azure AI Search (ベクトル検索のサービス比較)

Azure AI Searchはフルマネージド、高セキュリティ、メタデータ保持・管理が容易なため、実証時点で最も有効と判断した

分類	システム	Azure AI Search	FAISS	Qdrant	ChromaDB
基本情報	概要	クラウドベースのフルマネージド検索サービス	Facebookが開発した効率的な近似最近傍検索ライブラリ	オープンソースのベクトル類似検索エンジン	オープンソースのベクトル類似検索エンジン
	特徴	フルマネージド、高いセキュリティとAzureインフラによるスケーラビリティ	高速なベクトル検索、大規模データに適している	性能を重視したい場合のニーズに対応	機能の全体像は非常にコンパクト。ベクトルDBを試したいというニーズに対応
技術	メタデータ保持可否	可能	不可	可能	可能
	検索オプション	・ベクトル検索/ハイブリッド検索 ・AIエンリッチメント、フィルタ等	・ベクトル検索 ・多様なフィルタ処理が可能（必須/除外キーワード/フィルタ等）	・ベクトル検索	・ベクトル検索 ・キーワード検索(1単語のみ)
運用・管理		マネージドサービスであるため、容易。ポータル等でユーザーフレンドリー	プログラミングとカスタマイズが多く必要	プログラミングとカスタマイズが多く必要	プログラミングとカスタマイズが多く必要
セキュリティ		Private Endpoint利用による、高いセキュリティ	高い（AppServiceデプロイ前提）	高い（AppServiceデプロイ前提）	高い（AppServiceデプロイ前提）
コスト・価格		有料（月額利用料が発生）	無料（オープンソース）	無料（オープンソース）	無料（オープンソース）
ベンダー リレーション	サポート	Microsoftのサポートを利用可能	コミュニティサポート	コミュニティサポート	コミュニティサポート
	ベンダーロックイン	高い（Azureエコシステムに依存）	低い（オープンソース）	低い（オープンソース）	低い（オープンソース）



採用において重視した観点

※本事業ではMicrosoft Azure環境で使用可能なOSSをピックアップし比較。本格導入検討の際は、他ベクトル検索サービスの適用も視野に入れ、費用や相互運用性等の観点を踏まえた再精査、評価が必要と考える

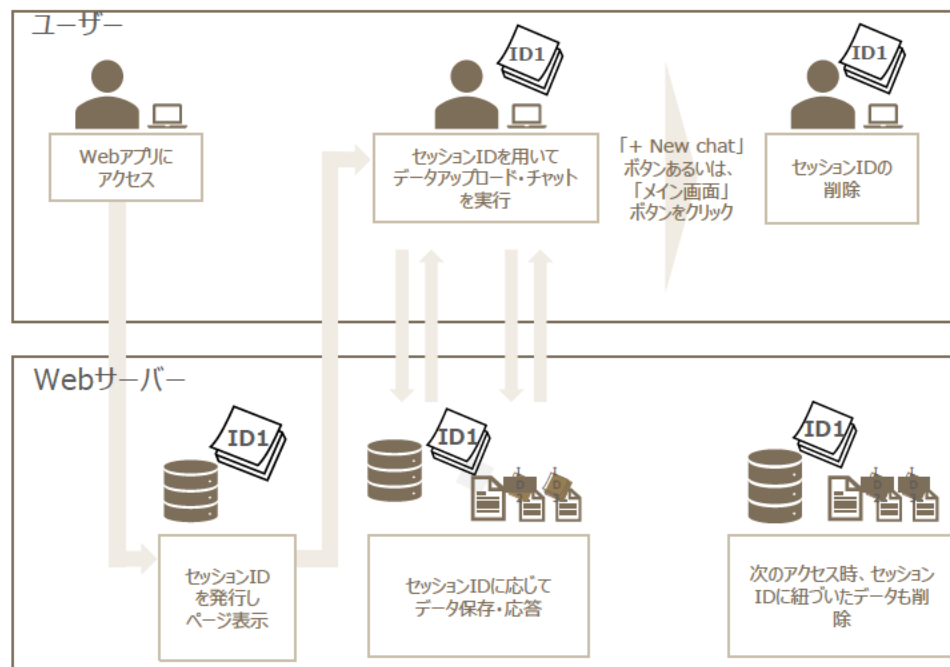
3.3. 検証環境の構築 (5) 機能要件・非機能要件

本検証環境で具備した機能要件・非機能要件を一覧に整理した

機能要件	詳細	実現方法
ログイン機能	予め設定されたアカウントでのみ、本サービスを利用可能となるよう、主体認証を備えたログイン機能を設けること	Azure AD B2CによるAzure認証及びアプリ側のユーザー登録
コンテンツ生成機能	ユーザー職員がプロンプトを入力することにより、最新のLLM及び独自データを用いたコンテンツ生成を行う機能とすること	GPT-4-32kモデルを提供
参照又は学習に供する独自データ登録および学習機能	ユーザー職員が、コンテンツ生成の高度化を行うため、当省が独自に参照又は学習させるデータを登録できるようにWebベースでの画面インターフェイスを構築し、その結果を環境内に蓄積できる機能を設けること。ただし、当省事務局職員においては全ての登録データのアクセス権限を保有できるようにすること	省内データをBlobへ登録し、Azure AI Searchを使用し、データ参照可能にした
利用状況分析機能	ユーザー職員が、本サービスの利用状況を確認、分析が可能となる機能を提供し、トレーサビリティ確保を目的に、アクセスログがダウンロードできるようにすること	PostgreSQLのDBテーブルをCSVファイル形式で出力、手動連携
非機能要件	詳細	実現方法
キャパシティ要件	ユーザー職員のユーザー数100人程度の登録及び利用を想定すること。入出力トークン数の合計は3億トークン程度を想定すること。上記トークン数は、事業期間中に職員が可能な限り本環境を利用できるよう、消費トークン数を低減する工夫も入れ込んで実装すること	ユーザー数100人以上の登録、及び、利用できる形で運用。消費トークン数の低減については改善案を提案
利用環境要件	本サービスの利用環境は、経産省基盤情報システムにて提供されるセキュアPC、および同システムネットワークからの接続を想定すること。ブラウザはHP Sure Click Secure Browserを想定	職員PCからアプリへのアクセス、アプリ利用可能
ユーザビリティ要件	デジタル庁デザインシステム対応。操作方法、視認性、状態の分かりやすさ、ヘルプ	デザインシステムを踏襲したUI開発実施
アクセシビリティ要件	「みんなの公共サイト運用ガイドライン」（総務省）、経済産業省ウェブアクセシビリティ方針を参考に、アクセシビリティを確保した設計・開発。指示や状態の分かりやすさ	貴省担当職員と協議しながらUI開発実施
可用性・継続性要件	メンテナンス時間を除く24時間365日の稼働が前提であることとする	本事業安定稼働（リリース時除く）
性能要件	本事業では要件として提示しないが、ユーザー職員が最低限利用可能となるようにすること	貴省職員利用可能な環境を提供
セキュリティ要件	本事業で扱う情報は、機密性1に限る。ユーザー職員が誤って機密性1以外のデータを登録しないよう、Webベースの画面インターフェイスには注意喚起を促すメッセージを表示する等、対策を講じること	注意喚起メッセージおよびアップロード直前のポップアップでの確認表示
教育要件	プロンプトの入力方法や制限事項等の内容を提供。研修テキストの準備	利用者研修実施（2023年11月6日）
移行要件	経産省が保有するデータを参照データ、学習データとして組み込むこと。組み込み時期については担当職員と調整のうえ、実施すること	データ受領時、弊社作業でデータ登録実施
運用要件	受託者は職員の利用において消費したキャパシティを日次で監視し、翌営業日にキャパシティ使用状況を担当職員に報告すること	週次、データ手動連携。使用状況の監視実施
保守要件	サービス提供不可となった場合は迅速な復旧に努めること。また、利用する製品・サービスに脆弱性を検出した場合は、速やかに対処すること	随時トラブルシューティング実施

複数職員の同時検証を実現するためのセッション管理機能、回答の根拠確認のため参照情報の出力、暗号化された行政文書ファイルに対応した

セッションIDを用いたセッション管理イメージ



※セッション：アプリの使用開始～終了までの単位

参照情報の出力¹

一例：「意匠審査基準 - 特許庁」のPDF(406ページ)

③ 特許庁 が検索対象に設定されました。回答生成が可能です。

④ どのような時に新規性の拒絶理由とならないか

⑤ 新規性の拒絶理由とならないのは、出願された意匠が意匠法第3条第1項各号、同第3条の2、同第9条第1項の規定に該当し、新規性、先願等の要件を満たす場合です。具体的には、意匠が具体的なものであり、工業上利用することができる意匠に該当する場合、また、出願された意匠が部品の意匠又は物品等の部分について意匠登録を受けようとする意匠である場合などが該当します。さらに、意匠の新規性喪失の例外規定の適用を受けることも可能で、これはあくまでも意匠登録出願より前に公開された意匠は意匠登録を受けることができないという原則に対する例外規定です。

ファイル名	ページ番号	参照文型
isho-reigai-qa.pdf	2	意匠の新規性喪失の例外規定は、あくまでも意匠登録出願より前に公開された意匠は意匠登録を受けることができないという原則に対する例外規定であることに留意する必要があります。
isho-shinsakijun-all.pdf	37	審査官は、出願された意匠が意匠法第3条第1項各号、同第3条の2、同第9条第1項の規定に該当し、新規性、先願等の要件を満たさない場合は、拒絶理由通知書に審査判断の理由を具体的に示す。また、引用意匠の特定にあたっては、引用意匠の出典がわかる情報（文獻名、発行日、号、巻、掲載頁、掲載位置等）を記載する。その際、出願された意匠が部品の意匠又は物品等の部分について意匠登録を受けようとする意匠である場合は、必要に応じて対比、判断をするための引用部分を明示する。

暗号化ファイル対応²

pycryptodomeライブラリを利用し、暗号化されたファイルの取扱いを可能にした

ライブラリ 記載名称	正式名称	暗号方式
AES	Advanced Encryption Standard	ブロック暗号
DES	Data Encryption Standard	ブロック暗号
DES3	Triple DES	ブロック暗号
Blowfish	- (左記同様)	ブロック暗号
ARC4	Alleged Rivest Cipher 4	ストリーム暗号
PKCS1_OAEP	Public Key Cryptography Standards #1 Optimal Asymmetric Encryption Padding	非対称(公開鍵)暗号

1. 意匠審査基準 | 経済産業省 特許庁 (ipo.go.jp)
2. <https://pycryptodome.readthedocs.io/en/latest/index.html>

3. ユースケース整理及び検証環境構築

3.4. 検証環境の基本機能 会話ログの機能概要（蓄積/表示/出力/回答生成評価）

STEP4

検証環境の構築

職員、機能別の会話ログ蓄積、履歴表示。利用状況の可視化のためのcsv出力及び、回答の評価機能を具備した

会話ログの呼び出し、表示の流れ

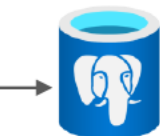
①チャットUIのサイドメニューに、機能モード切替用アコーディオンリストを表示

②選択した機能別に会話履歴を指定件数呼出し。対象会話履歴データを抽出



※会話継続不可。ユーザ入力エリア無効化

会話ログデータベース
(PostgreSQL¹)



⑥CSVファイル出力



利用者状況分析
(Power BI/経産省環境)

会話ログテーブルのDBテーブル設計

項番	PK	カラム名	項目概要	データ型	実データ例
1	●	conversation_id	会話ID	INT	0~
2		email_address	E-mailアドレス	CHAR	xxx@xx.xx
3		conversation_datetime	会話ログ日時	DATETIME	YYYY-MM-DD HH:MM:SS
4		content	会話文	TEXT	こんにちは。
5		role	発言者	TEXT	system / user / assistant
6		conversation_summary	会話の簡易サマリ	TEXT	xxx
7		first_conversation_id	会話キー	TEXT	xxx
8		function_type	機能キー	INT	Generic Chat(0) / URL(1) / PDF(2) / AI Search(4)
9		length	文字数	INT	1~
10		token_count	トークン数	INT	1~

PowerBIによる利用状況可視化画面（経産省環境）



1. Azure Database for PostgreSQLの選定理由

- ハードウェア、OS、データベースエンジン等の管理が不要なフルマネージドサービス
- MySQLと比較して集計機能が豊富
- ユーザー管理を行うことも想定し、一貫性や正確性が保証されているRDBを選定

参照した単一文書の回答生成において発生した課題に対し実施した施策とその効果を示す

発生した課題と検討した施策

施策による効果

① PDF読取ライブラリの選定

Document Loaders

採用手段 pyMuPDF Loader

当初、pyPDF Loaderを用いていたが、PDFをテキスト化する際、文字化けが発生する現象を検知し、PDF読取ライブラリを見直した

また別途調査により、pyPDF Loaderは比較的日本語の読取が不得手であることが判明し、代替手段として、日本語の解析に優位性があるPyMuPDF Loaderへ変更した

③ Robots.txtパースブロックしているサイトの対処

Webスクレイピング時、Robots.txtをパースする際に利用しているurllib.robotparseライブラリにおいて、デフォルトのユーザーエージェント（ウェブサーバにアクセス元を伝えるための識別子）をブロックしているサイトがあるため、デフォルトのユーザーエージェントを変更することで、対処した

② チャンク分割におけるライブラリ活用

Text Splitters

採用手段 RecursiveCharacterTextLoader

本検証環境は、セッションデータでベクトルデータを管理する方針としており、インメモリ管理であるベクターストアと親和性が低いため、VectorStore, Retrieversのメソッド群は利用しない方針とした

chain_typeによる精度向上は別事業調査により、効果が薄いと判断した

④ チャンク分割のパラメータ検討

施策②による等間隔の文字列分割で、GPT4-32kのコンテキストウィンドウの限界値の設定を実施

- チャンクサイズ(文字数) : 512
- チャンクオーバーラップ(前後の文字を重ね合わせる文字数) : 128(チャンクサイズの25%)
- PDF, URLのチャンク参照数 : 60件(512 × 60 = 30,720 ; gpt4-32k限界値設定)

読取可能になったPDF文書、Webサイトが増え、柔軟な検証が可能になった

一例 : 「意匠審査基準 - 特許庁」のPDF(406ページ)

PDF (isho-shinsakijun-all.pdf) を解析中です...少々お待ちください...

PDFの解析が完了しました。回答可能です。

要約してください。

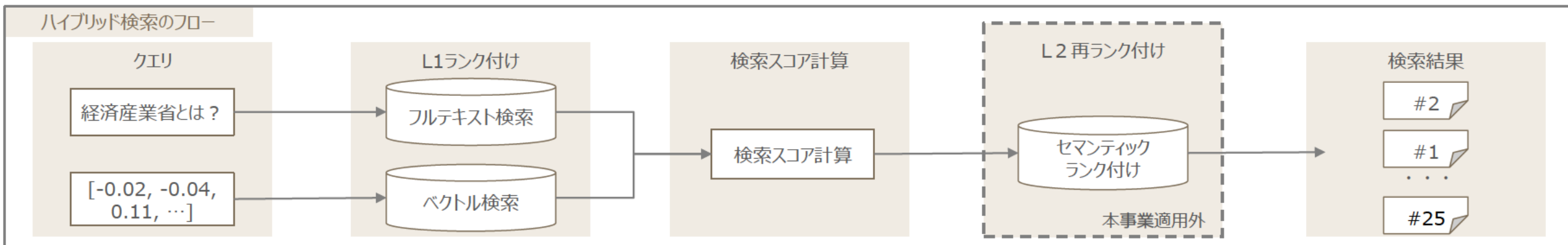
この文書は、特許庁が提供する意匠審査基準に関する情報を含んでいます。意匠登録出願の手続き、意匠の認定、意匠ごとの出願、個別の意匠登録出願、関連意匠、パリ条約による優先権などについての詳細なガイドラインが記載されています。また、意匠登録出願の際に提出する書類や証明書の提出期間、意匠の新規性や創作性、意匠の具体性などについても触れられています。さらに、意匠登録出願の際に必要な図面や写真の提出についての規定も含まれています。

ページ番号	参照文章
178	3 前項の規定の適用を受けようとする者は、その旨を記載した書面を意匠登録出願と同時に特許庁長官に提出し、かつ、第三条第一項第一号又は第二号に該当するに至つた意匠が前項の規定の適用を受けることができる意匠であることを証明する書面(次項及び第六十条の七において「証明書」という。)を意匠登録出願の日から三十日以内に特許庁長官に提出しなければならない。
353	パリ条約による優先権の主張を伴う我が国への意匠登録出願ができる期間(優先期間)は、第一国への最初の出願日から6か月である。
322	ようとする意匠を記載した図面を添付して特許庁長官に提出しなければならない。
280	出願当初の願書の記載及び願書に添付した図面等を総合的に判断しても、当然に導き出すことができる同一の範囲を超えるものであり、要旨を変更するものと判断する。
354	我が国においては、特許法と意匠法での法域相互間の出願の変更が可能である。
29	審査官は、願書又は図面等に対してなされた補正が、意匠の要旨を変更するものであるときは、その補正を決定をもって却下する。
57	本事例では、「意匠に係る物品」の欄の記載が不明確であり、図面の記載を考慮しても、何を装飾するものであるのか等、用途及び機能が明らかでなく、この意匠の意匠に係る物品等を明確に認定することができない。
196	4.1.1 「意匠に係る物品」の欄の記載 画像意匠について意匠登録出願する場合は、「意匠に係る物品」の欄に、画像の具体的な用途を明確に記載する。

3.4. 検証環境の基本機能 Azure AI Searchの回答精度向上

Azure AI Search機能であるハイブリッド検索を利用することで、検索精度の向上を図った

検索オプション	詳細	フィルター/メタデータによる絞り込みの併用
Single vector search（単一ベクトル検索）	ベクトル検索のみで検索するオプション。現在は、ベクトル検索するフィールドは文書データの内容（チャンク区切り）をfieldに設定している	○
▼		
検索オプション	詳細	フィルター/メタデータによる絞り込みの併用
Hybrid search（ハイブリッド検索） ¹	キーワード検索とベクトル検索を組み合わせで検索するオプション ①全文検索：BM25を用いた単語ベースでのランク付け ②ベクトル検索：cos類似度によるベクトル情報に基づいたランク付け ③ ①②の相互スコアを評価し（RRF）、候補をランク付けする	○



1. [Hybrid search scoring \(RRF\) - Azure AI Search | Microsoft Learn](#)

3.4. 検証環境の基本機能 LangChainを用いたChatGPT Plugin機能の活用

外部ツール連携による回答生成を目的に、ChatGPTプラグイン機能の実現性を検証した

採用手段と調査結果

LLMユースケースにおいて強力なAIオーケストレーションライブラリである「LangChain」を活用し、公開されているChatGPTプラグインを呼び出す方針とした。当該機能により外部情報収集、その情報に基づく回答生成が可能となる。具体的には、Agents機能を用いてユーザーからの要望をタスク分解し、問合せるツールを判断、実行する（下記図）



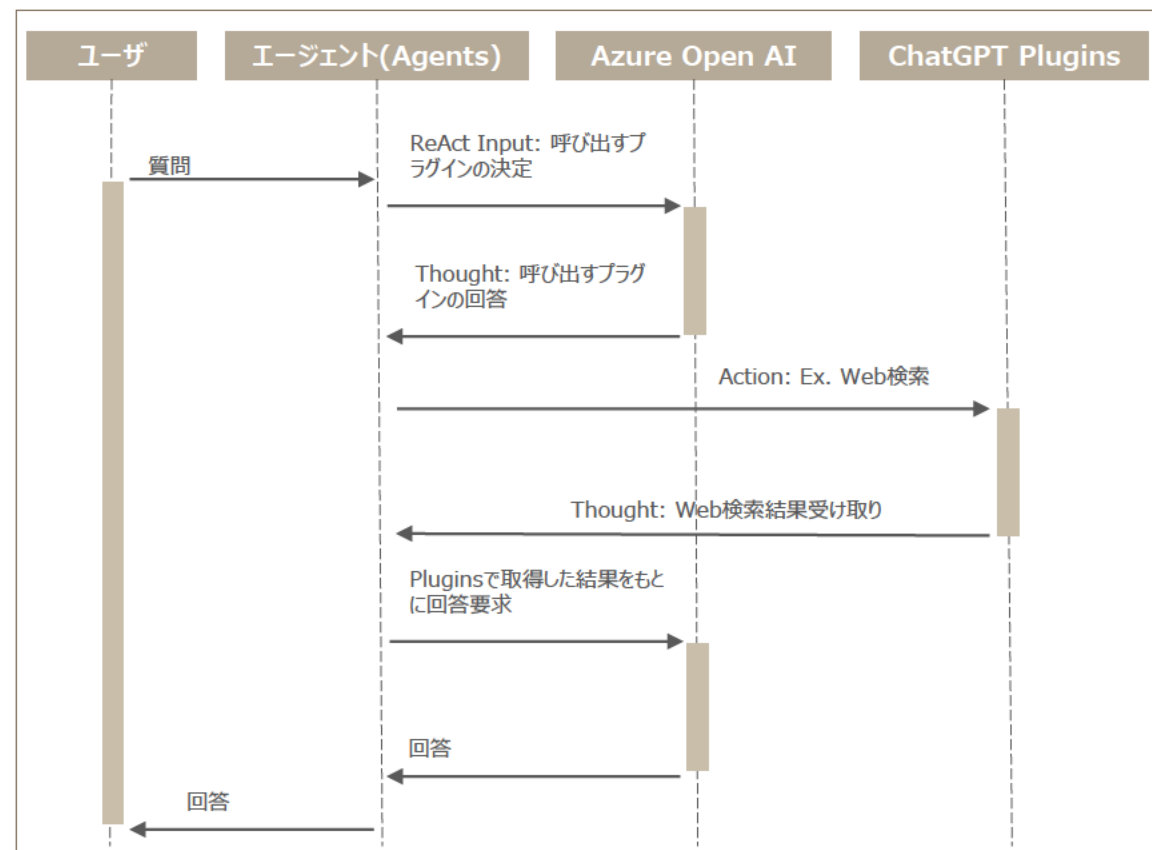
例) 代表的なAgentsで利用できるツール群

#	LangChain Agents	概要
1	Bing Search	Bingサービスと連携し、Web検索からの回答生成機能を付与する
2	ChatGPT Plugins	ChatGPTプラグイン機能を付与する

弊社調査による実現見通し(2023年11月時点)

- Azure OpenAIモデルにおいても、プラグイン利用は可能
- プラグイン活用による回答出力はStream出力が可能

シーケンス図



3.4. 検証環境の基本機能 再現性パラメータの調査

出力の再現性に関わるモデルパラメータを調査した。出力回答を意図的に再現可能となる

パラメータ名	目的	用途	使い方/ふるまい
temperature (top_p)	ランダム性の指定	出力トークンをサンプリングする際に、どの程度ランダム性を許容するかを決める。0.0～2.0の実数値。top_pの場合は、上位どれくらいを取るかを指定する。問合せ時に指定する。	ランダム性を許容しない値(temperature=0.0)を指定することで、ほぼ同じ出力を繰り返し得ることが期待できる
Seed	疑似乱数初期値の指定	出力トークンをサンプリングする際に使用する乱数を初期化するパラメータ。同じseed値を指定することで、同じ疑似乱数が発生する。問合せ時に指定する。	同じseed値を指定することで、ランダム性を許容しても(temperature>0.0)、ほぼ同じ出力を繰り返し得ることが期待できる
system_fingerprint	システム同一性の確認	バックエンドシステムの構成やモデルデプロイでユニークとなる文字列を返す。応答値に含まれる。	バックエンドシステムの構成やモデルデプロイが同一であるかどうかを確認する ※プロンプトおよび他のパラメータが同一でもモデルが変わると応答が変わるため

検証		質問文：“雪を題材にした俳句を考えてください”	
Temperature = 0.0（低ランダム性）		Temperature = 1.5（高ランダム性）	
Seed: 1	試行 1：雪解け水 春待つ心 音を立て 試行 2：雪解け水 春待つ心 音を立て	試行 1：初雪や 世界を覆う 白の紅 試行 2：初雪や 世界を覆う 白の紅	
	 同一、あるいは類似		 類似しない
Seed: 42	試行 1：雪解け水 春待つ芽吹き 静かなり 試行 2：雪解け水 春待つ心 音を立て	試行 1：雪解ける 静かに耳に 春の足音 試行 2：雪解ける 静かに耳に 春の足音	
	再現性は保証されない	 類似しない	

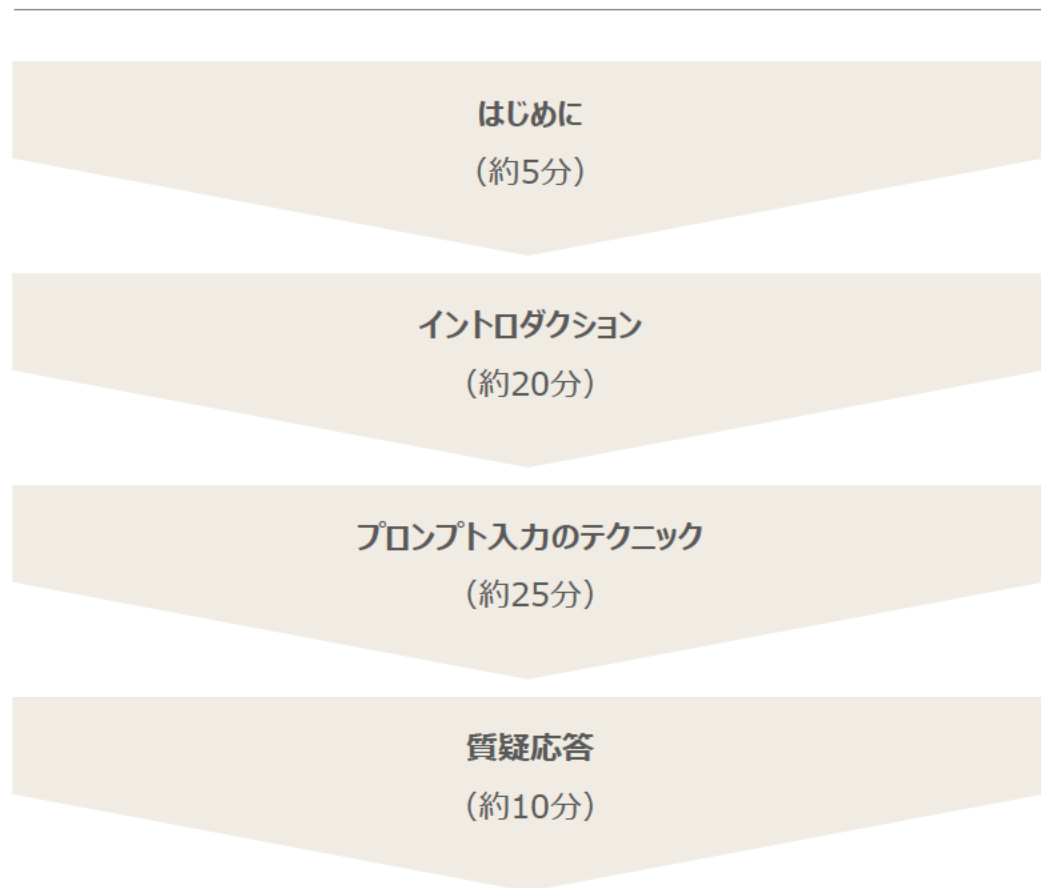
調査結果

TemperatureとSeedが同一の場合、同一または類似の出力
Temperatureが0.0の場合は、Seedに関わりなく同一または類似の出力
よって、Seedはランダム性と再現性を両立させたいときに活用するパラメータと
考えられる

3.5. 利用者研修の実施 研修の流れ

検証にあたり生成AIを初めて取り扱う職員も参画するため、リテラシー向上を目的に利用者研修を実施した

利用者研修の流れ



利用者研修の概要

項 目	概 要
対象者	これからLLMを業務で活用したいと考える経産省職員（Python等の知識は不問）
ゴール	- LLMを業務で活用するシーンを想定し、その利便性や業務効率化の効果を実感できる - 実証用システムの利用方法や制限事項などを理解できる
利用環境	Azure OpenAI（本検証環境）
利用バージョン	GPT4
実施環境	オンライン形式
実施日時	2023/11/6（月） 12:00～13:00

利用者研修の構成を示す

プロンプトのテクニック及び具体例の説明（50分）

質疑応答（10分）

はじめに

イントロダクション

プロンプト入力の
基礎テクニック

質疑応答

✓ 本研修の位置付け

- 本研修の位置づけ
- 目的とゴール

✓ アプリ環境の使用方法

- ログイン方法
- アプリの操作方法

✓ ChatGPTの概要

- LLMとは
- ChatGPTとは
- Promptとは
- Prompt Designとは
- プロンプトの要素
- プロンプトの種類
- GPT-3.5とGPT-4.0の違い

✓ 代表的なChatGPTの処理方法

- テキスト要約
- 情報抽出
- 質問応答
- テキスト分類
- 会話
- 推論

✓ 基礎テクニック

- 指示とコンテキストを区切る
- 一貫性のある言葉を使う
- 自分の理解度・立場を説明する
- ロールを付与する
- 追加情報をリクエストする
- 出力形式を規定する
- AIに質問させる
- 英語で説明する

（コラム：思考の木）

研修資料の詳細はAppendixに掲載

4. ユースケース検証及び次期課題の整理

- 4.1. 検証の実施方針
- 4.2. 検証結果の評価
- 4.3. 重要検討課題の整理
- 4.4. 今後のシステムのあるべき姿

各課室のユースケースに対し、実装した環境・機能で職員が実際に検証し、行政事務に資するかを定性、定量的に判断することを本検証の目的とした

※今年度の検証では、機密性のない情報に限って入力し、出力結果は内部用の素案として必ず職員が確認している

精度検証における前提

本事業の目的	行政事務の高度化に資する技術であるかの適用可能性を検証するとともに、当該技術を導入するにあたっての具備すべき要件及び制約事項等を洗い出すこと 令和5年度産業経済研究委託事業（行政事務の高度化に資するLLM導入可能性検証調査）仕様書より
検証目的	以下検証目的を満たすため、定性的、定量的観点で検証を進める ✓ LLM技術を活用した本検証環境を利用して、職員が業務導入できる可能性があるかを検証する ✓ 多数ある行政業務のユースケースのどれがLLM技術に有効になり得る業務が整理する ✓ 行政業務において一定の効果があり、導入するにあたり、どのような制約条件が発生するか見極める
検証における評価観点	各課ごとに着目している行政事務のユースケースそれぞれに対し、LLMの特性と検証環境の品質を踏まえながら、以下観点で検証を進める LLMモデルの網羅的評価を実施した論文 ¹ を参考にしつつ、各課検証職員の負担にならない評価項目の設計を実施した ✓ 生成結果の妥当性（回答の正確性、網羅性、ばらつき） ✓ 生成結果は行政事務において効果的、有効であるか（問題解決効果、業務改革性、活用満足度） ✓ 検証環境が十分足り得るか（環境満足度）

1. [Guo, Zishan, et al. "Evaluating Large Language Models: A Comprehensive Survey." arXiv preprint arXiv:2310.19736 \(2023\)](https://arxiv.org/abs/2310.19736)

4. ユースケース検証及び次期課題の整理

4.1. 検証の実施方針 進め方

各課室の検証担当者による検証環境利用と意見収集を行い、検証結果の取りまとめを実施した

検証の進め方

作業内容		1. 評価項目、指標決定	2. 検証環境利用、評価	3. 評価結果収集	4. 検証結果のとりまとめ
役割	共通	- 評価項目、指標の設定 - 評価項目、指標の調査	- 各課で想定したユースケースを検証環境を用いて各課検証担当者が検証、評価する - 検証環境で提供している機能の有効性を確認するための評価項目と指標を各課検証担当者が検証結果記入フォーマットをもとに記入する	- 各課の利用を踏まえて、検証環境を用いた検証・評価結果の収集を事務局が実施する - 各課検証・評価結果の収集	- 評価項目に沿って各課の検証結果を集計し、評価指標を算出、整理を実施する - 各検証観点(具体性/効果)ごとの検証結果を本報告書に記載する
	経済産業省				
	委託事業者				
作成物		✓ 検証ユースケース一覧 ✓ 評価項目、指標の方針 ✓ 検証結果記入フォーマット	✓ 会話履歴データ (一部生成内容に対する評価含)	✓ 各課それぞれの検証結果記入フォーマット	✓ 検証結果の評価及び分析 ✓ 課題の抽出及び施策の検討結果

4. ユースケース検証及び次期課題の整理

4.1. 検証の実施方針 評価項目一覧

各課検証メンバーが評価することを想定し、統一的な検証作業の評価観点の一覧を定めた

前提		1. 定量的な評価については納得感が得られるように、なるべく一貫した結果となるような方法を検討 2. 検証メンバーに無用な負担を掛けないようにしたい(各課の検証パターンは3～4ケース、各課1人あたり1ケースを対応する、と仮定を置く) 3. できる範囲でのプロンプト改善は各位実施してから、検証に入る(有効なプロンプトは再利用可能で、そこに到達するまではラーニングコストと捉える)		
No	評価項目	評価項目概要	評価方法	評価結果 (割合、多段階評価、理由)
1	回答の正確性	生成結果の内容がどの程度正しいか	・ 質的にどの程度の手直しが必要か (担当職員の模範解答を100とする)	〇/100(点)
2	回答の網羅性	生成結果は十分な情報を提供できているか	・ 期待した結果と比較して、何割程度の抜け漏れが発生しているか ・ 逆に、冗長すぎる回答になっていないか	〇/100(点)
3	回答の再現性 (ばらつき)	同一ユースケース内で、質的・量的なばらつきがあるか	手直しにかかる時間や正確性に関して、同一ユースケース内でばらつきが大きい場合に記載	定性評価可
4	問題解決効果	生成結果は、課題に対し効果的、あるいは実現可能な解決策を提示、提案しているか	業務を進めるにあたり、妥当な解決策を提供できているかを定性的に判断する。回答内容が納得できるか否かで判断する	100 (最高) 50 (人間のアイデアと同様) 0 (最低)
A	業務改革性 (工数削減効果)	LLMを活用することで、タスク単位の工数(時間)削減、月間の工数(時間)削減が期待できるか	検証したタスクについて、ゼロの状態から期待する結果を得るまでの時間を計算 ※クリアプロセスにかかる時間は除く 結果は、事務局が業務改善効果として数値化、集計して共有する	①定型業務 or 非定型業務 ②生成AIを活用した際の工数 (〇人×〇時間) ③職員のみで対応した際の工数 (〇人×〇時間) ④当該タスクの発生頻度 (週に〇回程度) ⑤タスクの担当レベル (係員・係長・補佐・専門非)
B1	活用満足度	生成AI活用の可能性	検証したタスクについて、生成AI活用に対する満足度を5段階で評価する	5 (最高) 3 (人間の手作業と変わらない) 1 (最低)
B2	環境満足度	省内環境への満足度	今回検証したシステムに関する満足度を5段階で評価する	5 (省内で導入する価値がある) 3 (無料版のChatGPTと同様) 1 (著しい欠陥がある)

4. ユースケース検証及び次期課題の整理

4.1. 検証の実施方針 業務改革性の評価方法

STEP5

検証方針の検討
及び実施

タスク単位の時間削減効果、月間の業務時間削減インパクトによる定量評価を実施した

前提

1. 毎回プロンプトを考える必要がある業務(非定型業務)と、過去作成したプロンプトを再利用できる業務(定型業務)を二分化する想定
2. 定型業務の場合、削減効果を評価しやすくするため、プロンプト調整は一通り終わっている前提でLLM回答時間(h)のみ計算する
3. 非定型業務の場合、やりとりしながら業務を進める特性があるため、プロンプト調整時間(h)、LLM回答時間(h)を含めて計算する
4. 3において、削減効果が検証担当者のプロンプト習熟度に影響するため、今後、プロンプト習熟度が向上されることを想定する

評価手法

i) タスク単位の時間削減効果

$$\text{LLM適用による削減効果 (\%/回)} = \left[1 - \frac{\text{LLMを適用しない場合のタスク時間(h)}}{\text{LLMとやり取りする時間(h) + 手直し時間(h) + 補足時間(h)}} \right] \times 100$$

LLMとやり取りする時間(h) = プロンプト調整時間(h) + LLM回答時間(h)
プロンプト調整時間(h) ≒ プロンプトの習熟度



正の数値(%) : 削減効果あり
負の数値(%) : 削減効果なし

例) 要約業務をLLMを使って30分から24分になった場合、
(1 - 0.4(h) / 0.5(h)) × 100 = 20(%)の削減効果あり、
となる

ii) 月間の業務時間削減インパクト

$$\text{業務時間削減インパクト(h/月)} = \text{LLMを適用しない場合のタスク時間} \times \text{LLM適用による削減効果 (\%/回)} \times \text{タスク発生頻度 (回/月)}$$



削減時間の数値に対する最終評価は
各原課で再評価

例) i)のタスク単位の時間削減効果の値を利用し、
0.5(h) * 0.2 × (1(回) * 20(/営業日/月)) = 2.0(h/月)
の削減となる

※サンプルが多く採取でき、再現性の影響（ばらつき具合）を考慮可能であれば実施

検証環境の現状と課題を明らかにし、業務活用に向けた課題改善施策の考案に繋げることを目的とした



4.1. 検証の実施方針 対象ユースケースの類型一覧（再掲）

下記ユースケースの類型に関して、定性分析、定量分析を実施した

No	ユースケースの類型	
1	問い合わせ対応	外部/内部からの問い合わせ対応として、問い合わせに対する適切な部署や担当者の出力や、回答案の作成を行う
2	論点抽出	大量の文書情報から主要な論点や議論の要点を抽出する
3	翻訳	外国語の文書を日本語に、または日本語の文書を外国語に翻訳する
4	要約	長い文書や報告書を短く、要点を抽出した形で提供する
5	コード生成	特定のタスクを自動化するためのコードを生成する。例えば、Excelマクロ、Power BI等
6	データ解析	収集したデータを解析し、パターン、トレンド、洞察を抽出する
7	文案作成	報告書、プレスリリース、ウェブサイト等各種文書のコンテンツ案を作成する
8	壁打ち	政策に関する新しいアイデアや提案を複数出す
9	事例収集	特定のテーマや問題に関連する事例を収集し、分析や報告のための資料案とする

4. ユースケース検証及び次期課題の整理

4.1. 検証の実施方針 ユースケースの厳密性

各ユースケースについて、1つの評価軸として業務要求の高さ（厳密性）を考慮して分類した

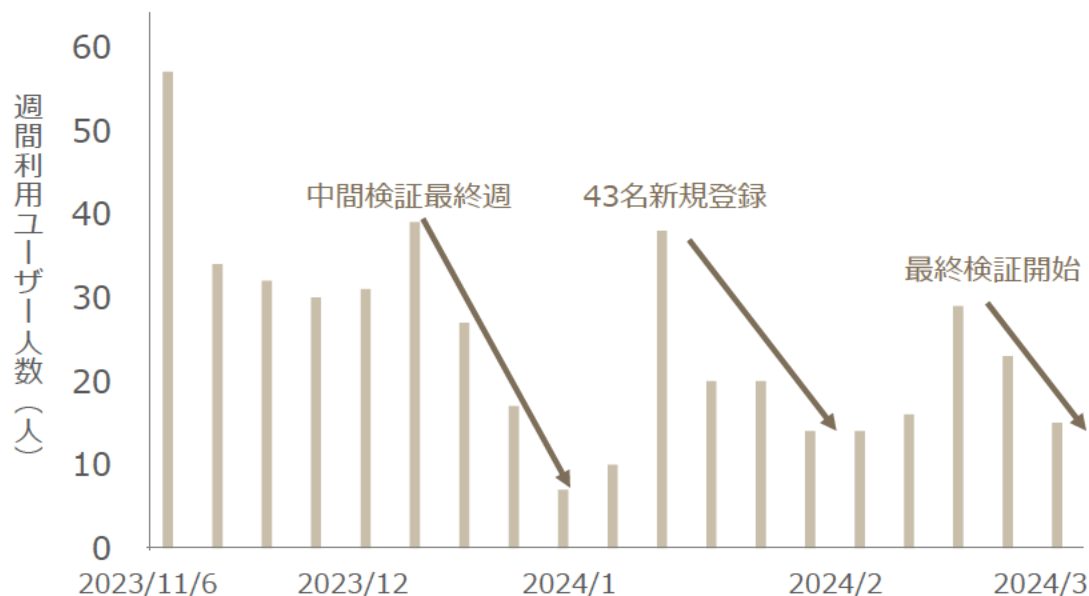
厳密性		類型								
レベル	概要	問い合わせ対応	論点抽出	翻訳	要約	コード生成	データ解析	文案作成	壁打ち	事例収集
高	外部に対し、直接的に影響を与える可能性が高い	✓	✓							
中	間接的に外部に影響をあたえる可能性がある		✓	✓	✓	✓	✓	✓	✓	
低	外部に影響を与える可能性が低い							✓	✓	✓

検証メンバーのうち約半数が徐々に離脱傾向にあり、ユーザーの定着が課題であると読み取れる

ユーザー数の推移

集計期間：2023/11/6～2024/3/22、利用した総ユーザー数¹：114人

ユーザーは生成AI利用に関して前向きな職員層であるにも関わらず、
定着できていない

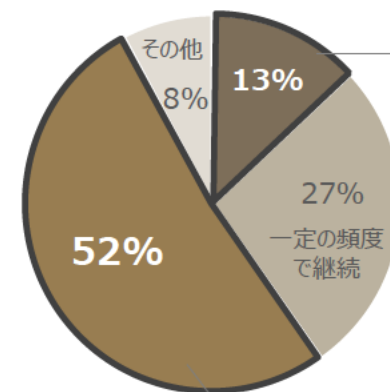


1. 登録されている職員165名の内、ログインしたことがある人数

ユーザーの利用意識

事後アンケート結果より集計 (n=77人)

利用が増えているユーザーが一定数存在するが、約半数が離脱傾向にある



利用が徐々に増加 **13%**

- ・ 業務で必要性を感じる
- ・ ユーザーのスキルが向上
- ・ 便利さを感じる

利用が徐々に減少 **52%**

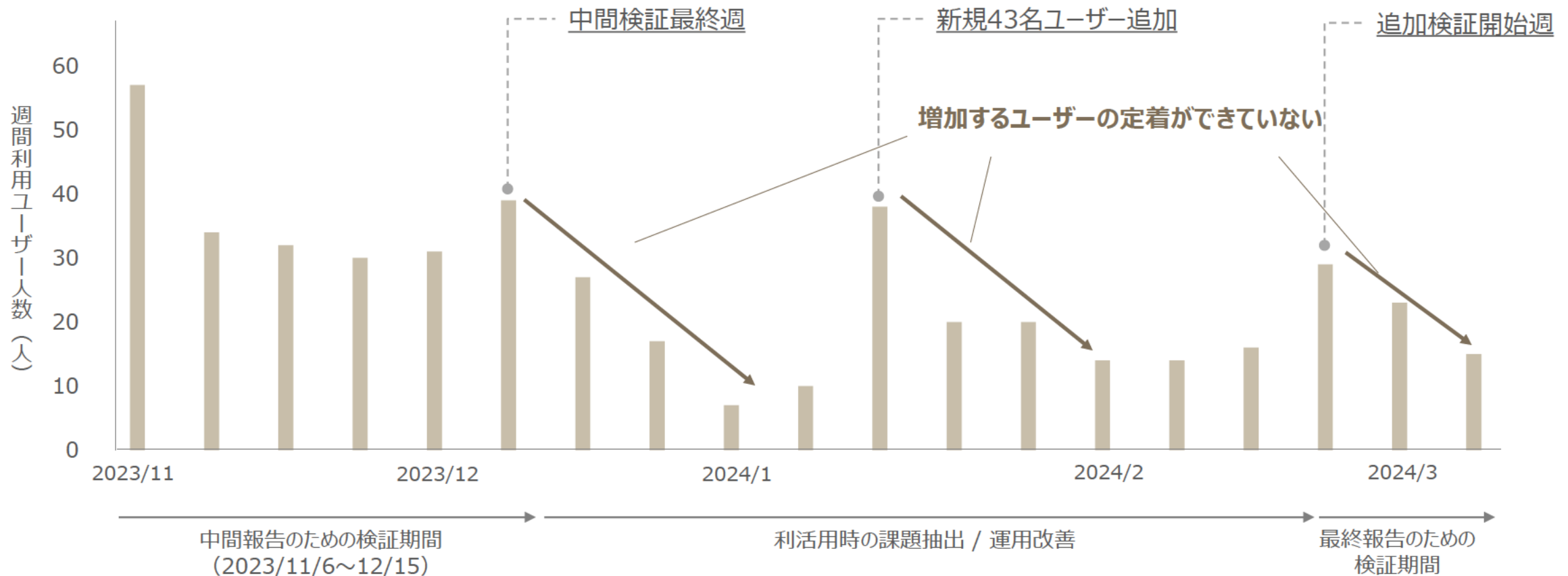
- ・ 性能が業務要求レベルに未達
- ・ 環境が使いづらい
- ・ 業務にフィットしない
- ・ 機密性のある資料が使えない
- ・ ユーザーのスキル起因
- ・ その他（繁忙期等）

4.2. 検証結果の評価 利用動向分析：ユーザー数の推移

実証期間全体を通して、ユーザーの定着が課題である

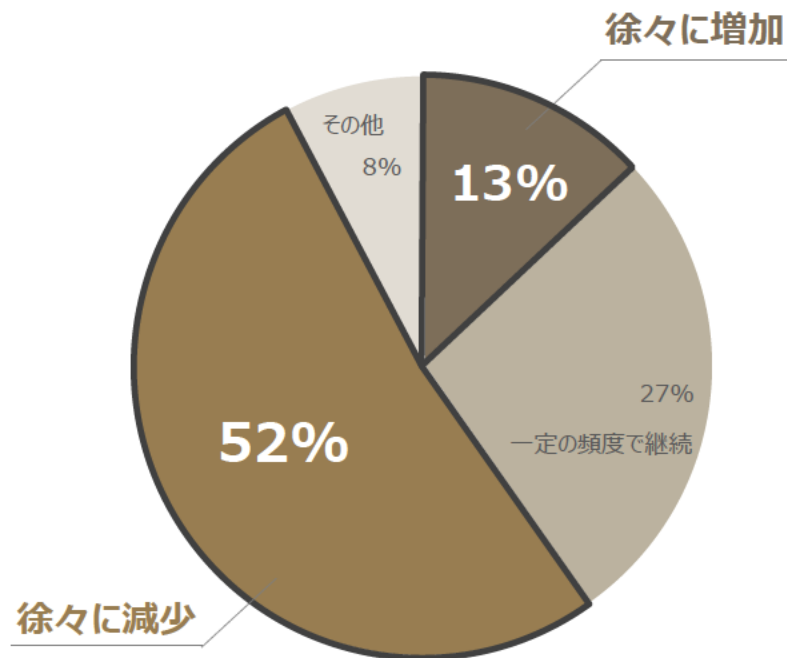
ユーザー数の推移

集計期間：2023/11/6～2024/3/22、利用した総ユーザー数：114人



利用頻度が増えているユーザーが一部存在するものの、約半数が離脱傾向にある

実証期間全体を通した利用動向

利用頻度の変動¹ n=77人、選択回答具体的な理由²（一部抜粋）

利用が徐々に増加（全体の13%）

n=10 記述回答

業務で必要性を感じる

6人/10人

- 一般的な検索ですぐに得ることの出来ない情報をすぐに得ることが出来るため
- 調査、検索、などの作業が増えてきたため

ユーザーのスキルが向上

2人/10人

- 生成AIの活用が有効な場面や、有効なプロンプトの入力方法が分かるようになってきたため

便利さを感じる

2人/10人

- 思っていた以上に使い勝手が良く、便利だった

利用が徐々に減少（全体の52%）

n=40 記述回答

性能が業務要求レベルに未達

10人/40人

- 調べ物をするにしてもまだ人間がやる方が早くて正確であるように感じたため、使う必要を感じなくなったため

環境が使いづらい

7人/40人

- 起動まで、1分程度時間がかかるため。teams等のように、アプリ&自動ログイン機能があると良いと思います

業務にフィットしない

5人/40人

- 最初は活用を意識できていたが、生成AIを使おうと思いつく場面があまりなかった

機密性のある情報が使えない

4人/40人

- 機密性がネックになり、できる作業が少ないと感じたため

ユーザーのスキル起因

2人/40人

- 良いプロンプトが思いつかず、理想の回答が得られなかったため

業務状況の変化/その他

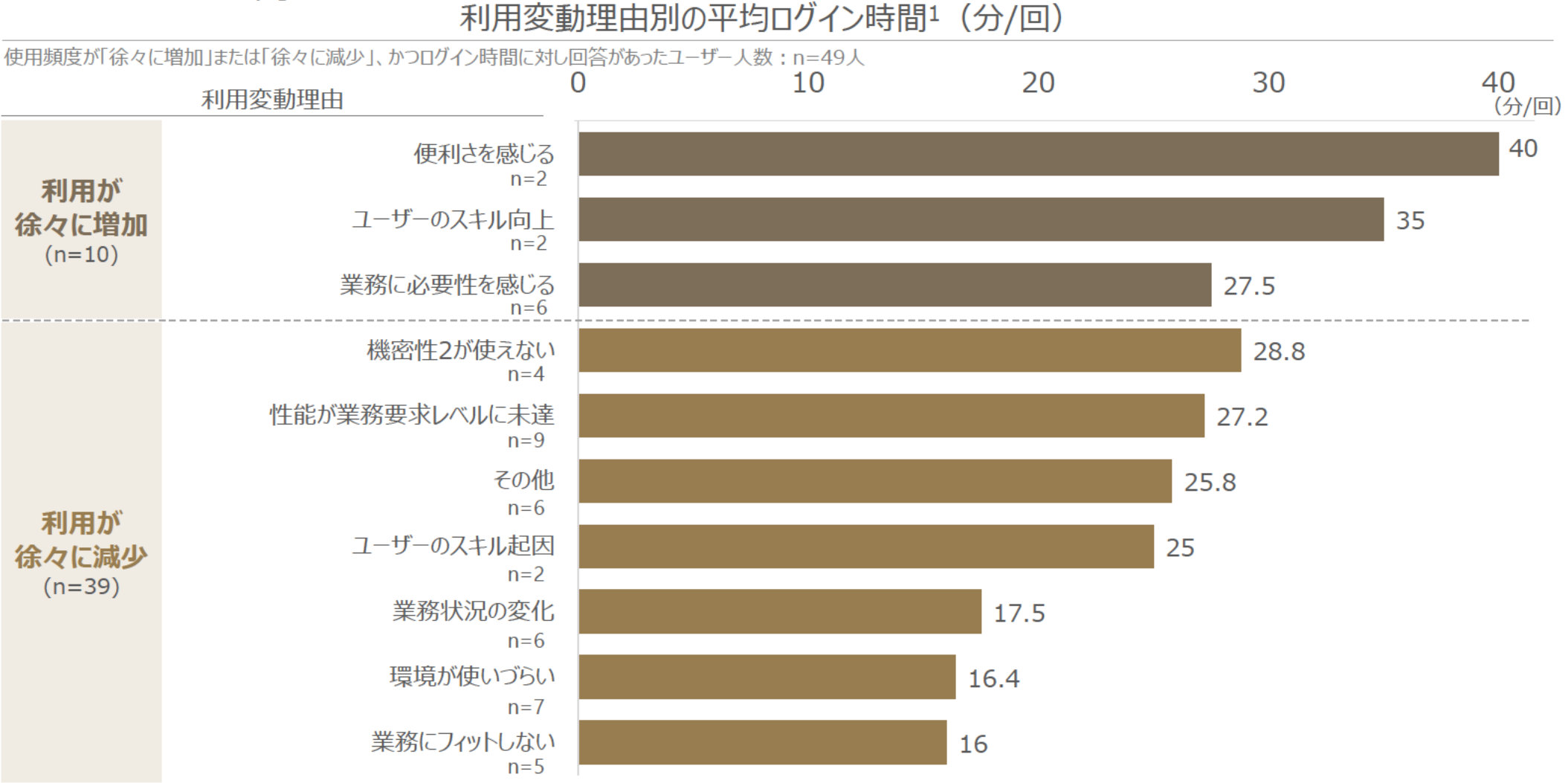
12人/40人

- 業務が多忙になり生成AI実証にかかる時間が取れなくなった

1. 検証後のアンケートより「実証期間全体を通して、生成AIの使用頻度に変動はありましたか？」集計
2. （上記1に対して）その理由は何ですか？

4.2. 検証結果の評価 利用動向分析：利用変動理由とログイン時間

「徐々に増加」の場合、ログイン時間は30分前後。一方、「徐々に減少」した回答したユーザーはログイン時間は「徐々に増加」と比較して、短い傾向にある



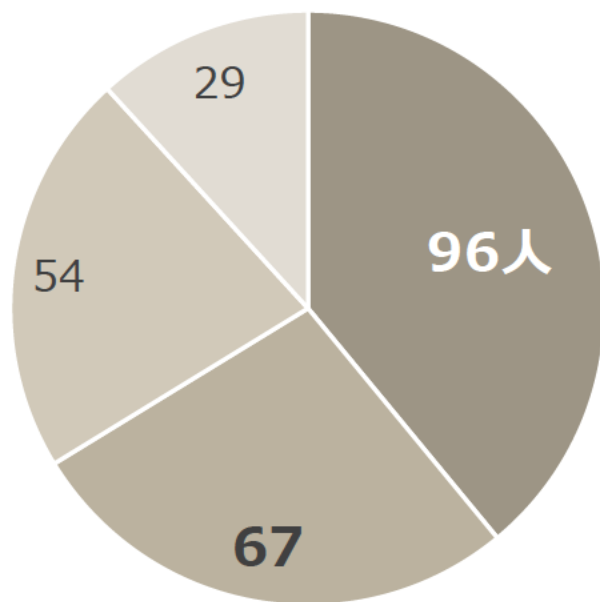
1. 事後アンケートで収集したログイン時間を利用変動理由別に平均値を算出

4.2. 検証結果の評価 利用動向分析：機能別ユーザー数/プロンプト入力回数

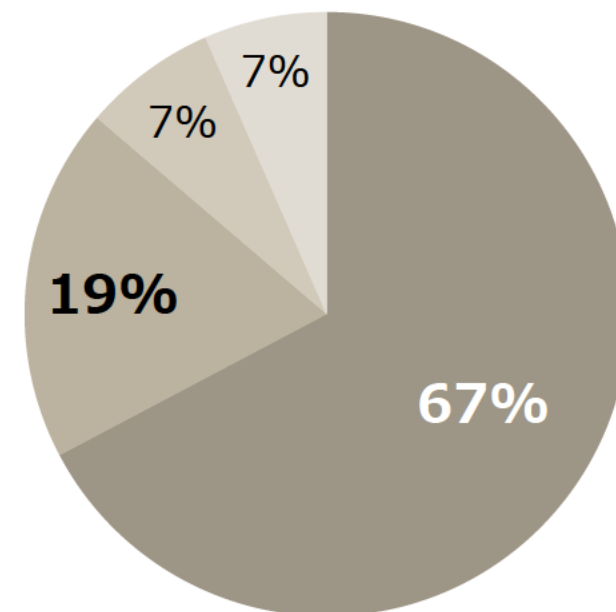
機能別ユーザー数、プロンプトの入力回数を集計したところGeneric Chat機能が最も多く使用されており、次いでPDF Answering機能が使用されている

機能別ユーザー数¹ (n=244人)

■ Generic Chat ■ PDF Answering
■ URL Answering ■ AI Search

プロンプト入力回数¹の割合 (n=5,747回)

■ Generic Chat ■ PDF Answering
■ AI Search ■ URL Answering



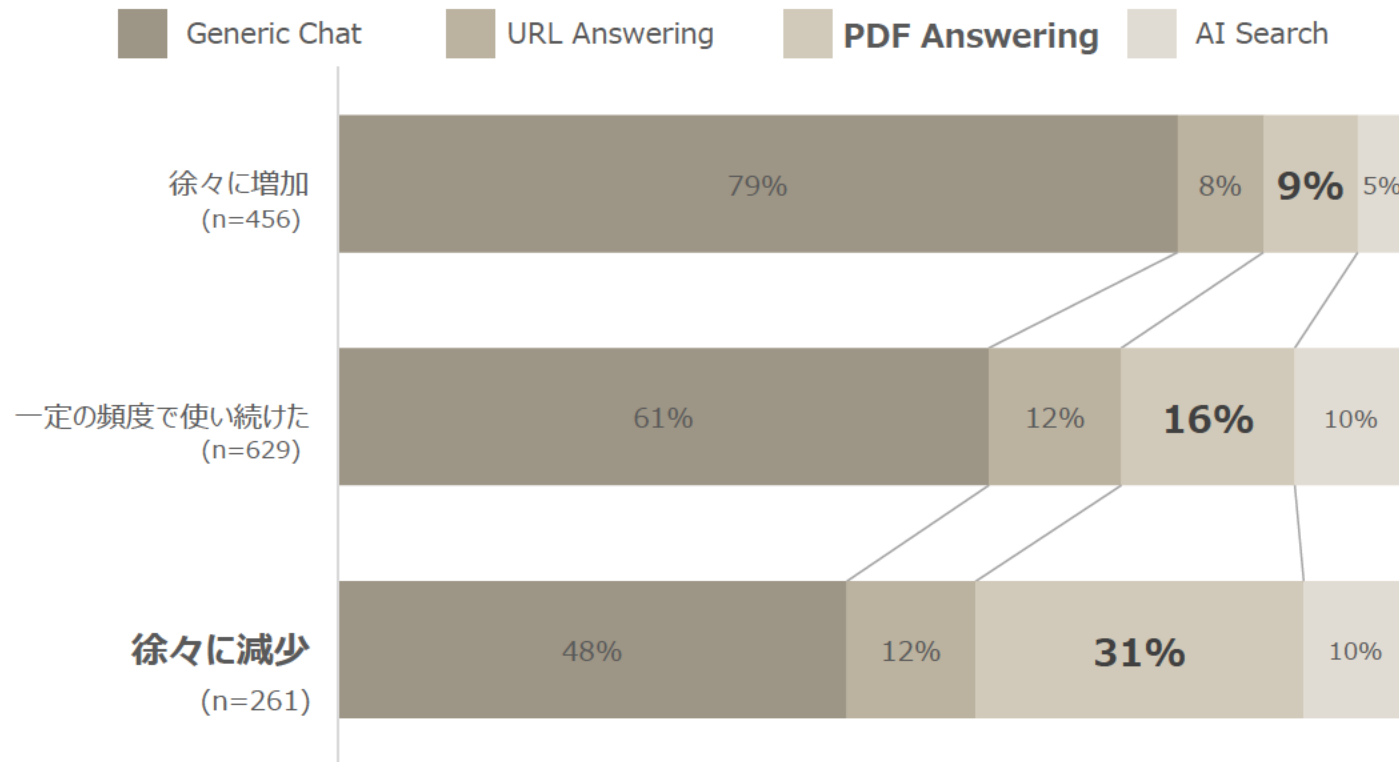
1. 集計期間（2023/11/10～2024/3/22）における集計対象となった独立したユーザー数は107人。機能別の利用ユーザー数は延べ数で算出。プロンプト入力回数も同期間で算出

4.2. 検証結果の評価 利用動向分析：利用変動理由と利用機能の割合

利用頻度が「徐々に減少」した層は他と比較し、Generic Chatの利用が少なく、PDF Answeringの割合が高い。PDF Answering機能に関する意見では、ネガティブな意見も事後アンケートから読み取れた

ユーザー属性別の利用機能の割合

使用頻度が「徐々に増加」「一定の頻度で使い続けた」「徐々に減少」の職員による会話の試行回数^{1,2}：n=1,346回

PDF Answeringに関連するユーザー意見³

“PDF参照は準備に手間がかかる割に、期待した回答が得られず、あまり使わなくなりました”

“URL AnsweringやPDF Answeringを使用する際、資料によっては正しく文字を認識できない場合が存在した”

“PDFやURLは審議会の議事録など、かなり膨大な資料の要約に使用したかったが精度が追いついておらず使えなかった”

“もう少し、長文を長文として参照してほしい（長めのURLやPDFを入れると、全部分を参照してくれない）”

1. 新規に会話を開始した際に、1回とカウント

2. 2023/11/10～2024/3/22の期間にて、会話履歴がある「徐々に増加」：10人、「一定の頻度で使い続けた」：21人、「徐々に減少」：33人分の試行回数から算出

3. 事後アンケート「検証環境の利用時に生じた問題や困難があれば、記載してください。」（自由記述）に対するネガティブな評価のコメントを抜粋

各課室の報告内容より、類型別に定性分析を実施。意見を整理し、課題を抽出した

ユースケース	ペインポイント	ポジティブな意見（職員の声抜粋）	ネガティブな意見（職員の声抜粋）	主な課題種別
問い合わせ対応	- 高い正確性と網羅性が要求される - 問い合わせが抽象的である傾向 - 職員の作業負担が特に大きい	- 具体的かつシンプルな単純な質問への回答は可能 - たたき台作成が可能 - 全て人手作業よりは負荷が低い	- 回答が複数あるとき、抜け漏れがあり、網羅性が低い - 法令の条文が認識できないと活用範囲が大幅に制限 - 現状、回答の手直しで逆に工数が増加する可能性	網羅性 正確性
論点抽出	専門的な内容、委員会資料、議事録から過去の経緯把握や内容を検索	- 回答を絞って質問すると、該当の箇所から回答 - 簡易な質問は、単一PDFから一定水準の回答を取得	- 結論の整理は、意図する回答が得られない - 一部のハルシネーションや参照箇所のスキップがあり、全ての項目を手で確認が必要	網羅性 正確性
翻訳	- 頻度が高い - 一定の工数がかかる定型業務	- 質の高い英語に訳出が可能 - 専門業者よりも質が高いように感じられる場合もある	- 固有名詞や定訳などは別途調査やデータを読み込みが必要 - 一般的な無料翻訳サイトと比較してやや冗長な表現	参照データのオペレーション
コード生成	- コードの調査・実装・検証に時間がかかる - プログラミング知識が必要	- 生成するコードは申し分ない。コードの修正も問題ない - プログラミング知識がなくてもコード作成ができる点が最も大きな効果	- 四則演算の記号の記載が漏れることがある - マクロ作成については正解イメージを持っていないと活用が難しい可能性もある	正確性
要約	報告書の読み込みに時間がかかる	- 英文報告書の要約所要時間の削減効果が大きい（翻訳との組合せ） - 概要を作成させ、内容の理解を素早く実施できる	- 情報の重要度の細かい判別が苦手 - 活用できる範囲が人間のチェックが可能な範囲 - 大量の文書から任意に抽出させると抜け漏れあり	網羅性 正確性
データ解析	グラフの特徴を把握し、コメントを付与する工数が多い	最新の数値が記されたテキスト部分が区別しやすい場合、グラフの内容を回答可能。工数削減も見込める	グラフの傾向は読み取れない	機械判読性
文案作成	- 講演資料の構成案を効率化したい - 知見の無い領域のQ&A作成等	- 人間の介入なく、一般的な講演内容であれば、構成を作成可能（時間配分含む） - アウトプットの質向上（企画骨子作成/アイデア創出）	- 驚くような突飛な回答はない - 業務効率化にはつながっていない - 最終確認ができるほどの精度は出せない	創造性 網羅性 正確性
壁打ち	- 即座に多様なコンセプトの大量アウトプット - 専門的な知見を持つ壁打ち相手の不在	- 人間よりアウトプットは早く、原案作成として有用 - 従来に枠を超えた示唆を受け、実際に具体化している - 非定型業務への活用に期待	- 網羅的にアウトプットは可能だがそれ以上の示唆はない - AIの出力をそのまま採用してもチーム内の納得感醸成に繋がらない	創造性 正確性
事例収集	大量の文書からの素早い事例の収集	- 適切な事例を収集 - 取り出したい情報をすぐに入手でき、大幅な工数削減	事例の一部のみを抜粋しており、情報がやや不足	網羅性

主に組織外部からの問い合わせのユースケースを想定していることから、高い正確性と網羅性が要求される
 加えて、問い合わせが抽象的である傾向があり、生成AIの活用時に職員に作業負担がかかる

	ユースケース	ペインポイント	ポジティブな意見（職員の声）	ネガティブな意見（職員の声）	課題種別
組織外部	問い合わせに対する回答案作成	相談者からの各相談や意見に対して、担当課が一から回答案を作成する必要がある	- 全て人手作業よりは負担が低い - データを読み込ませれば一定数有効 - 検索が楽になり、かなりの負担削減が見込める - 調べものに要する時間が削減 - AI Search導入により、回答精度・検索能力が大幅に向上	- 音声入力を可能にするか、問い合わせフォームの一本化により、電話による問い合わせを可能にすることで削減効果が見込める - 参照箇所を網羅的に使って回答しているため、回答が冗長 - 職員が修正した回答を学習データとして随時追加可能にしたい	機能拡張（音声）
	問い合わせに対する関連箇所の抽出	関連する議題や内容を探す作業に苦労している 工数削減による問い合わせ対応能力の強化したい	- 具体的かつシンプルな質問への回答が可能 - 工夫により一定の工数削減が期待 - 質問に対し具体的な回答をしている - 回答作成時の参考情報を提案している - 問い合わせ対応の実務で役立った	- 回答が複数あるときに抜け漏れがあり、職員による正確性の確認が必要 - 手直しが必要な回答が多い - 質問の仕方によりバラつきがみられる - 正しい法令解釈が困難 - 各ページに散在し、複数回掲載されている情報抽出時に、回答が重複	参照データのオペレーション 網羅性 正確性
組織内部・外部	システムの問い合わせ	問い合わせに対し、シニア職員の力を借りる必要や、マニュアルを読む必要がある	- 一定の工数削減が期待 - 導入として、慣れない職員に有効 - 手直しは必要だがバラつきはない - 質問内容を細かく分けて、分かり易く入力した際に精度向上	- 修正箇所が見え消しで出力されると、業務により有効 - PDFがキャプチャを入れている場合に正しい回答ができない - シニア職員から正しい情報を得たり、マニュアルを読まないと利用が難しい - インプットをAIが解読しやすい形態にする必要がある	プロンプト作成補助 網羅性 正確性 読み取り機能の強化 機能拡張（画像）

工数削減や時間短縮が見込めるものの、生成物の網羅性・正確性については職員による確認が必要である

	ユースケース	ペインポイント	ポジティブな意見（職員の声）	ネガティブな意見（職員の声）	課題種別
内容把握	過去資料から該当部分の抽出や概要の把握	新規着任者が、委員会資料や議事録等から過去経緯や議題、内容を探す作業に苦心している	- 単一の議事録の要約は結論抽出は人間と同程度 - AI Searchは検索ツールとして有効 - 文章の割合が多いほど良い回答	- 結論の集約に課題がある - PDF読み込みの待ち時間が長い - URLの下層まで検索対象としたい	正確性 処理時間の短縮 読み取り機能の強化
	局内有志勉強会における議事とりまとめ	とりまとめ事務の負担を軽減したい とりまとめ時の執筆者の主観排除したい 生成AIならではの斬新な意見を得たい	- 議事から要点を抽出可能 - 一連の質問に対する回答に一貫性がある - 大幅な工数削減、時間短縮に寄与	もう少しオリジナリティのある内容だとより良かった	創造性
	英語レポートにおける論点抽出	レポートの翻訳・論点抽出にかかる工数を削減したい	詳細に場所を絞って質問すると、該当箇所から回答が生成できた	- 数字に関する回答ができなかった - レポート中に記載のある年度が不明、もしくは恣意的に抽出された	正確性
情報収集	ツールの機能制限やエラー回避策の収集（Tableau等）	ツール利用にあたる情報収集作業を効率化したい	—	- 関連記事の情報が出ない - 反応速度が遅い - 適切なURLを入れる必要があり、有用なURLを探すための時間を要する	正確性 処理時間の短縮 参照データのオペレーション
	政策立案等の基礎となる情報収集	過去白書の情報収集や要約する作業を効率化したい	- ファイルのメタ情報が認識可能 - 網羅的な探索が可能	- 事例、コラム、本文が混同して、一部ハルシネーションも存在 - 参照情報をいくつかスキップしている - 正確性・網羅性の担保が必要	網羅性 正確性
	経済産業省HP上の公開情報から事業名・担当部署の抽出	Web情報の抽出を手作業で行っている	- 手作業のコピー＆ペースト作業が減少し、大幅に効率化 - 文章を生成AIが分かり易いよう変更したところ、求める回答が得られた	- 全項目が回答されているかを人の目で確認する必要がある - URL上にあるPDFの読み込みが可能になると良い	網羅性 読み取り機能の強化

4. ユースケース検証及び次期課題の整理

4.2. 検証結果の評価 定性分析：翻訳

STEP6

検証結果の集計・分析

文の長さに関わらず、質の高い英語翻訳ができ、広報関連や情報収集の業務において活用が期待できる

ユースケース		ペインポイント	ポジティブな意見（職員の声）	ネガティブな意見（職員の声）	課題種別
翻訳	ホームページに掲載するニュースリリース等の英語への翻訳・校正	一定の工数がかかり、頻度も高い定型業務を効率化したい	- 質の高い英語に訳出が可能。英訳の精度が高い - 広報室の契約している英語業者よりも質が高い場合があった - データであれば網羅的にまとめたうえでヒントを提供してくれる	- 込み入った質問の場合、別途データを読み込ませることが必要 - 固有名詞や定訳などは別途調査やデータを読み込みが必要 - 一般的な無料翻訳サイトと比較してやや冗長な表現 - 校正した修正箇所が判別できるように出力されると、より業務に使いやすくなる	アプリ・生成結果の視認性向上
	Xアカウントへの投稿テキスト作成	翻訳・校閲・X投稿文作成という一連の業務を効率化したい	- 特に翻訳はX投稿は、作業スピードの向上や効率化に繋がり、ベース作成の業務量減に貢献できる見込み - 広報室の英文校閲者よりもこなれた洗練されたテキストを生成し、英訳も高精度		参照データのオペレーション
	講演資料の英訳	翻訳業務を効率化したい	業務効率化に繋がる	—	—
	精度内容の抽出にあたる海外文書の翻訳	質問に回答する際にかかる英訳・要約業務を効率化したい	- 英訳に関しては完璧 - 簡単なものに関しては過不足ない	整理が必要な制度については対応できない	網羅性

その他の考えられるユースケース： イベント等の案内メールのひな形作成等

開発やデバック等、プログラミング知識を要する作業の効率化が期待できる

	ユースケース	ペインポイント	ポジティブな意見（職員の声）	ネガティブな意見（職員の声）	課題種別
開発補助	コード生成補助 (Power BIメジャー/Excel VBA等)	コードの実装に時間がかかる 一定のプログラミング知識が必要である Excelマクロ作成の効率化したい	- 正常に動くコードが生成され、エラー解決策に至るまで詳細な情報が提供された - Excel VBAやプログラミングに不慣れな職員でも、コード作成が可能であり、大幅な工数削減が期待 - 過不足なく分かり易い説明である - コード修正も問題ない	- 四則演算の記号の記載が漏れることがある - エラーが出ないようなフォローがあると良い - マクロ作成については正解イメージを持っていないと活用が難しい	正確性
	ローコードシステムの開発補助 (Power Automate等)	PCに詳しくない職員によるシステム開発を可能にしたい	- 詳しい情報を提示することでかなり正確なプログラムを提示してくれる - エラーが出た際に原因追究させると一定数明確な回答が得られる - システム開発初心者でも簡単なものは作成可能 - AIの得意不得意の分野を理解すればかなり業務効率が向上	作成されたコードや、DBを極力増やさない制御方法等、局員が作成したものと大差ないので、問題解決に至るかは怪しい	機能拡張 (データ分析)
デバック補助	コードデバッグ作業	プログラミングにおけるエラーを効率的に解消したい	- エラー解消には完璧であった - Web検索でエラー解消することと比較して時間短縮に繋がった	プロンプト（質問）の作成に時間がかかる。ここを効率化できると最高	創造性
					プロンプト作成補助

要約に加えて翻訳を行うケースも見られた。課題として、膨大なデータの要約と、詳細な情報の重要度判別が挙げられ、ミスが許されない行政目的では活用が限定される可能性がある

	ユースケース	ペインポイント	ポジティブな意見（職員の声）	ネガティブな意見（職員の声）	課題種別
要約のみ	局長年頭所感・挨拶文の要約	指定字数の中で作成する必要がある	- たたき台としては、若干の工数削減になる - 完璧ではないが概ね内容は網羅されていた	- 情報の重要度の細かい判別が苦手 - 職員による修正を前提とした利用 - 残すべき要素の判断はLLMには難しく、多少の手直しが必要	正確性
	レポートの要約	人間（特に初任者）がレポートを要約する作業に負荷がかかっている	- 初任者の担当業務のキャッチアップ等における活用が期待 - Word由来のテキスト中心のPDFは、一定の実用的な効果を示唆	- 現状タスクを任せられるだけのレベルにはない - PowerPoint由来の図表中心のPDFは、実用的に不向きと思われる	機械判読性
要約＋翻訳	英文報告書の要約	様々な報告書の読み込みに時間がかかっている	- 英文報告書については、所要時間の削減効果大きい - 概要を作成させ、その後読み始めることによって、理解が容易になった	- 行政目的だとミスがあってはいけないので、活用範囲が人の確認が可能な範囲（英語以外は確認不可） - 再現性がない	正確性
	特定分野の要約及び論点抽出	要約にかかる時間も、ファイルからコピー＆ペーストして翻訳ツールに逐一入力していて効率が悪い	- 翻訳と要約を同時に行うことが可能 - 予め部分を指定して要約させると良い結果となった	- 膨大なデータを一度に読ませると、詳細な内容が返ってこない - 大量の文書から任意に抽出させると抜け漏れが発生	網羅性 正確性
要約＋外部機能	スキャンしたPDFからOCRした文書の校正	会計手続きにおける訳文作成業務を効率化したい	- プロンプト内で不正な校正は行わないよう指示も理解し、適切に処理 - 生産性の向上に寄与	OCRのプロセスにGPT-4VやPower Apps AI Builder等が利用できるとプロセスがさらに短縮する	機能拡張（画像）
	利用検証中の議事録アプリ（AmiVoice ScribeAssist）	議事録作成業務に負荷がかかっている	- クバ取り、文意に合わない語の判断 - 分量が増えるほど時間短縮の効果があり、生産性向上に寄与	Word形式での出力が可能になると、テンプレートへ埋め込んだドキュメント作成・検証が可能になる	外部ツール連携 外部ツール連携

4.2. 検証結果の評価 定性分析：データ解析

将来的な外部ツールとの連携を想定した検証において、グラフの傾向は読み取れないものの部分的な数値読取りが可能なケースも確認でき、更なる検証にて業務適用の可能性はある

ユースケース		ペインポイント	ポジティブな意見（職員の声）	ネガティブな意見（職員の声）	課題種別
グラフの解析	統計データ分析におけるコメント生成の自動化	グラフの特徴を把握し、コメントを付与する工数を削減したい	最新の数値が記されたテキスト部分が区別しやすい場合、グラフの内容を回答可能。工数削減も見込める	- グラフの傾向は読み取れない - 正確に演算させようとする（分析・集計）とPythonコードのみが回答されるため、Python実行環境があると良い - 入力時にテキスト入力欄を拡大できると良い	機械判読性
					機能拡張（データ分析）
					アプリ・生成結果の視認性向上

共通して一定の効率化は見られる一方で、新規創出のケースにおいては主に創造性、既存情報から作成するケースでは主に網羅性に課題が見られた

	ユースケース	ペインポイント		ポジティブな意見（職員の声）		ネガティブな意見（職員の声）		課題種別
新規創出	プレゼン資料の構成案作成	頻繁に依頼のあるテーマについて、講演資料の構成案作成を効率化したい	>	人間の介入なく、一般的な講演内容であれば構成の作成が可能	>	- 驚くような突飛な回答はない - 生成された資料構成と、既存資料を組み合わせた資料作成ができない	>	創造性
	外部への発信文作成	発信文の内容によっては職員が文案を考える負荷が大きい	>	- 状況に応じた表現が可能 - 文案検討時の工数及び負荷を下げる効果が期待	>	微修正が必要	>	正確性
	外部への発信文作成（祝辞・挨拶文）	指定字数の中で作成する必要がある	>	完璧ではないが、概ね必要な内容は網羅されていた	>	- 時間・工数の削減は望めない - データを学習させ、ゼロベースでの挨拶文の作成が可能になることを期待	>	参照データのオペレーション 創造性
	企画の仕様検討	企画のアウトプットの質を向上させたい	>	アウトプットの質向上に役立つ（企画の骨子作成やアイデア創出等）	>	- 業務効率化にはつながっていない - 入力する枠が狭くて入力しづらい	>	アプリの視認性向上
既存情報から作成	外部への発信文作成（プレスリリースからX・Facebook）	より興味を持ってもらえるSNS投稿文を作成したい	>	- 抜け漏れなく回答された - 作業時間の短縮が見込まれる他、文字数に収める際の負担感が軽減	>	- 工数削減にはつながっていない - 思うようなXの文案が生成されない	>	創造性
	外部への発信文作成（プレスリリースからメルマガ）	記事作成にかかる委託コストや作業時間を削減したい	>	- 記事の下書きとして活用可能 - コスト低減が可能	>	- 情報が不足している - 最終確認の作業が発生する	>	網羅性 正確性
	Q&A作成（組織内向け）	知見の無い職員が質問を検討する工程が必要である	>	- 質問作成は問題ない - 手作業と比較してかなり効率化	>	- 作成された回答に抜け漏れがある - 機密性を含む情報の取扱いが必要	>	網羅性 機密性2以上
	YouTubeの公開動画をTeamsの音声認識機能で文字起こししたものを修正し、議事録作成	議事録作成業務に負担がかかっている	>	—	>	- PDFが長いと網羅的に参照されない - 同音異義語を文脈に合うように正確に修正できていない	>	網羅性 正確性

作業工数の削減は期待できるものの、正確性や創造性に課題があり、実務での利用は現状期待できない

	ユースケース	ペインポイント		ポジティブな意見（職員の声）	ネガティブな意見（職員の声）		課題種別
情報収集・壁打ち役の設定	白書の調査設計に関する指摘	専門的な知見を持つ壁打ち相手の不在	➤	- 原案作成としては有用 - ファイル数が増加したことで網羅的な探索が可能に	- 実務に使うハードルは高く、正確性・網羅性の担保が重要 - 事例、コラム、本文が混同 - 参照箇所が誤っている - 論点漏れがある	➤	網羅性
	政策等のアイデア出しに向けた壁打ち役の設定	過年度白書との差別化を図りたい 壁打ちに向けた参照資料内容を把握・抽出する際の工数を削減したい	➤			➤	正確性
	政策議論にあたる論点整理、ヒアリング候補者の特定	ヒアリング候補者を探す作業を効率化したい	➤	- 議論の論点をピックアップするために有効 - 情報収集の時間が削減可能	具体的なヒアリング候補者の選定については誤りが多く、有効に活用することは難しい	➤	正確性
	事例マップや組織マップの作成	各事例との関係性を素早く把握したい	➤	情報収集の時間が削減可能	- 正確性の確認が必要である - テキストから図を作成する機能があると良い	➤	正確性 機能拡張（データ分析）
創造	新事業におけるキャッチーなコンセプトネームのアイデア出し	即座に多様なコンセプトを大量に考えることの限界	➤	- 非定型業務への活用に期待 - プレストの手間が短縮可能	- AIの出力をそのまま採用しても納得感の醸成に繋がらない - ビジュアルコンセプトを検討したい	➤	機能拡張（画像）
	地域企業の新陳代謝をテーマとした新政策の検討	問題意識を共有したうえでより良い施策を提案したい	➤	- 施策整理や不足の言及が可能。短期間に整理された回答が得られた - 回答にバラつきがなかった	- 人と同程度のアイデアにとどまる - 反対主張や意見を学習させ、仮想の壁打ち相手が作れると良い	➤	創造性
広報	メディアの想定問い立ての迅速な作成	想定される問い立てを効率的に考案したい	➤	- 網羅性が高い - 担当者のバイアスなく速やかな問い立てが可能	- 当たり前の問い立てのみであった - 機密性がないケースは少ない	➤	創造性 機密性2対応
	政策広報の手段の検討	従来の枠を超えた広報手段を検討したい	➤	- 従来の枠を超えた示唆を受け、実際に具体化している - 人間よりアウトプットが速い	- 網羅的だが、それ以上の示唆がない - 機密性のある情報が扱えない - Web検索機能がない	➤	創造性 機密性2対応 参照データのオペレーション

4.2. 検証結果の評価 定性分析：事例収集

適切な事例を迅速に収集できるため大幅な工数削減が期待できるものの、事例の一部のみの抜粋であるなど網羅性に課題がある

業務効率化	ユースケース	ペインポイント	ポジティブな意見（職員の声）	ネガティブな意見（職員の声）	課題種別
	ロボット導入事例の調査	システム内容から金額や規模感を把握したい	- 適切な事例を収集している - 事例集の中から取り出したい情報をすぐに入手可能なため、大幅な工数削減となる - より多くのデータを読み込ませることで活用の幅が広がるのでは	事例の一部を抜粋しており、情報がやや不足している印象	網羅性

4.2. 検証結果の評価 定性分析を踏まえた課題抽出

定性分析より抽出された課題と、対応するユースケース類型を一覧化する

抽出された課題種別		対象ユースケース類型								
		問い合わせ対応	論点抽出	翻訳	コード生成	要約	データ解析	文案作成	壁打ち	事例収集
精度向上	・ 網羅性	✓	✓	✓	✓	✓		✓	✓	✓
	・ 正確性	✓	✓		✓	✓		✓	✓	
	・ 創造性		✓		✓			✓	✓	
データマネジメント	・ 機械判読性					✓	✓			
	・ 参照データのオペレーション	✓	✓	✓				✓	✓	
	・ 機密性のある情報の取扱い							✓	✓	
	・ 読み取り機能の強化	✓	✓							
拡張	・ 機能拡張（画像/音声/データ分析）	✓			✓	✓	✓		✓	
	・ 外部ツール連携					✓				
ユーザビリティ向上	・ アプリ・生成結果の視認性向上			✓			✓	✓		
	・ 処理待機時間の短縮		✓							
	・ プロンプト作成補助	✓			✓					

4. ユースケース検証及び次期課題の整理

4.2. 検証結果の評価 定量分析：サマリ

現時点で翻訳、コード生成、要約の類型が業務高度化において有意義な可能性が高い

類型	厳密性 (レベル：低～高)	工数削減効果 (週間の工数削減人時)	生成された回答の質 (利用者の評価)	活用満足度 (5段階の平均)
問い合わせ対応	高	70 人時/週	悪い	2.9
論点抽出	高	5.3	悪い	3.6
翻訳	中	16	良い	4.8
コード生成	中	5.9	良い	4.4
要約	中	2.0	良い	4.3
データ解析	中	0.9	—	3.5
文案作成	低～中	1.3	良い	4.0
壁打ち	低～中	0.2	良い	4.4
事例収集	低	0.5	良い	4.0

業務への適用可能性

業務への利活用が限定的となる可能性が高い

- 工数削減効果は大きい
- しかし、回答の質の評価や活用満足度が低く、利活用にあたっては、課題の把握と対応方針の検討が必要

業務高度化に資する可能性が高い

- 1週間あたり、「～約2日人時」の工数削減が見込める
※「要約」は工数削減効果がやや低い
- 生成された回答に対する利用者評価も「良い」
- 活用満足度が「人手作業よりも良い」という評価

工数削減効果は低いが業務高度化が見込める

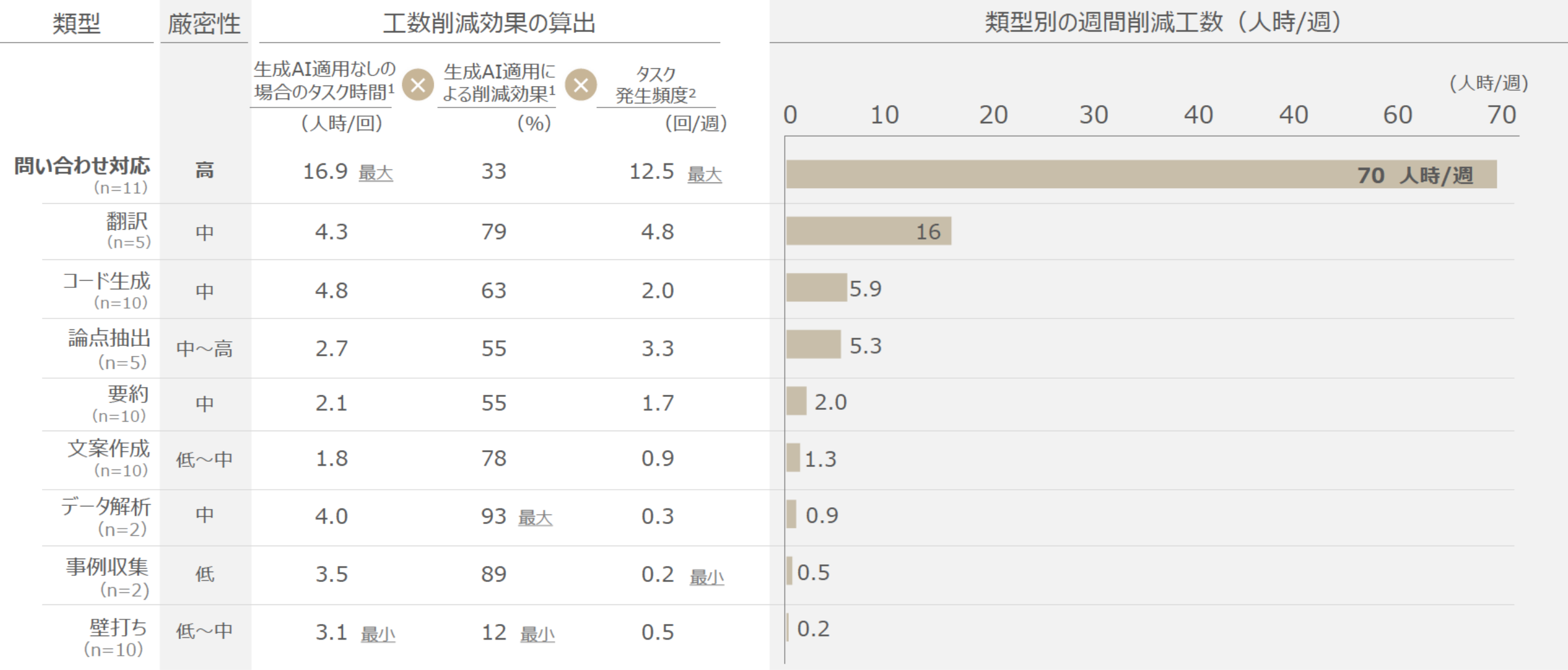
- 工数低減効果は他類型と比較し、小さい
- 生成された回答への利用者の評価は「良い」
- 活用満足度もおおむね「人手作業よりも良い」と評価

※「データ解析」は、定性分析で一定の成果が確認できているが、生成AIが不得手とする数値を取り扱うことを想定しているため、業務利活用においては引き続き検証が必要な類型

4. ユースケース検証及び次期課題の整理

4.2. 検証結果の評価 定量分析：業務改革性（工数削減効果）

厳密性が高く求められる問い合わせ対応等ほど、生成AIによる週間の工数削減効果が高い傾向にある



1. 各原課の最終報告及び中間報告における、“生成AIを活用した際の工数”、“職員のみで対応した際の工数”より平均値（人時/回）および生成AIによる削減効果（％）を算出

2. 各原課の最終報告及び中間報告における、“当該タスクの発生頻度”の平均値（回/週）

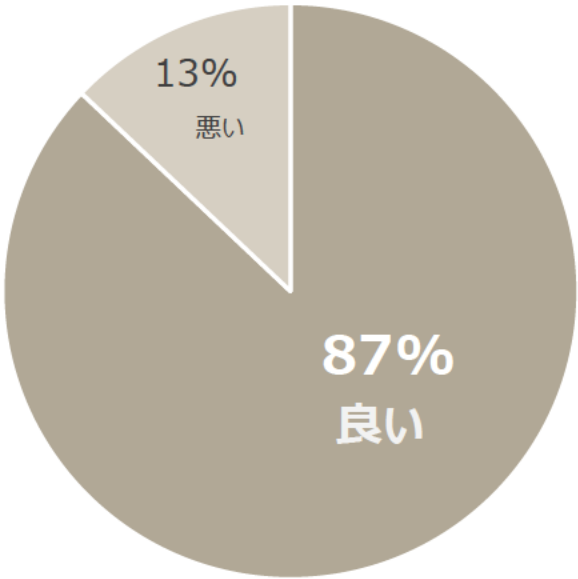
4. ユースケース検証及び次期課題の整理

4.2. 検証結果の評価 定量分析：回答生成評価

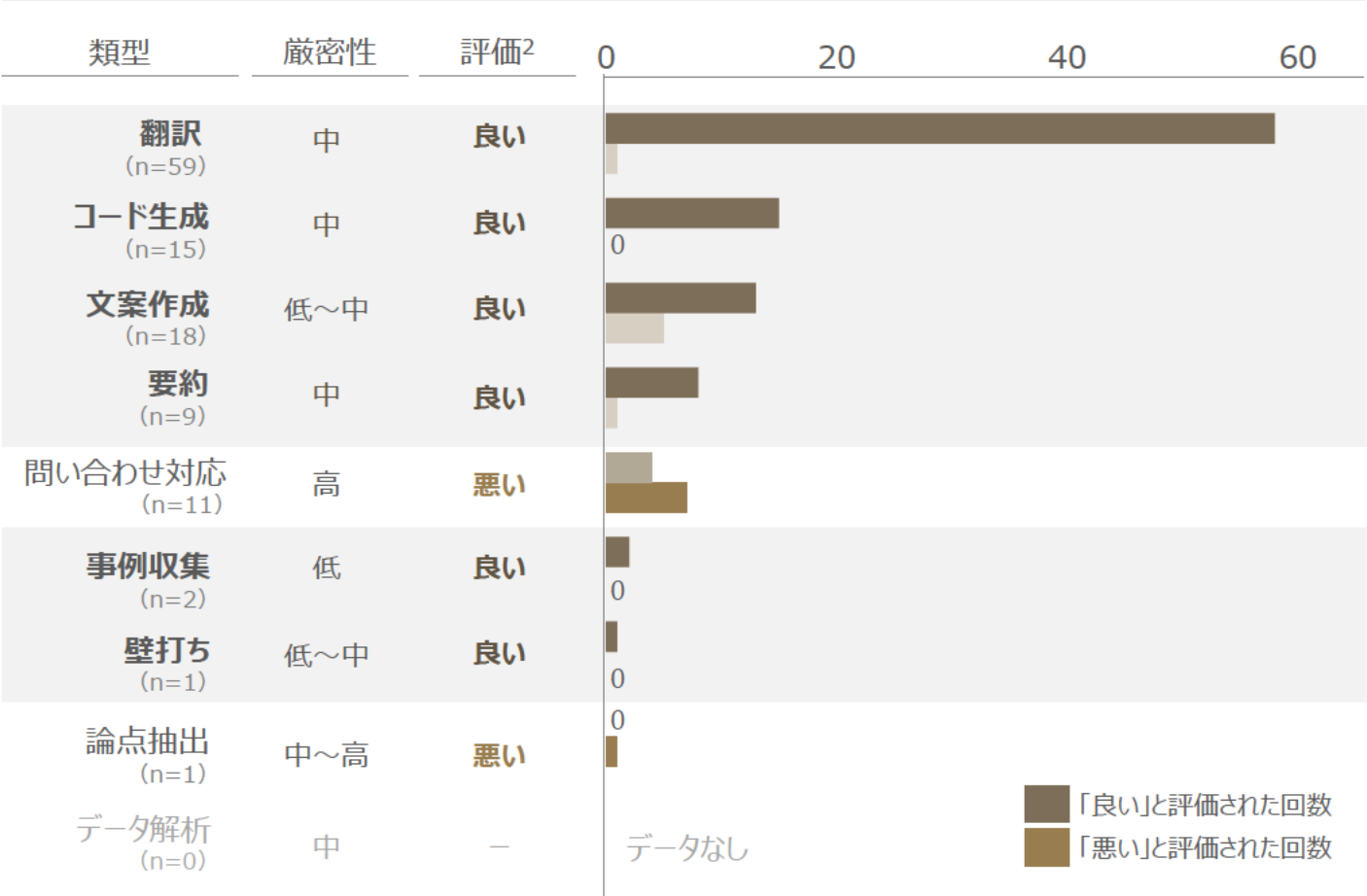
生成された回答の約90%が「良い」と評価。問い合わせ対応、論点抽出を除く類型に「良い」評価が多い傾向にある。※個々の生成結果について良い/悪いをGood/Badのボタンで職員が評価する機能を設けた

生成された回答のユーザーによる評価¹

集計期間:2024/2/19~2024/3/22, 総回答数:n=116回(Good/Bad選択), 総回答人数:19人

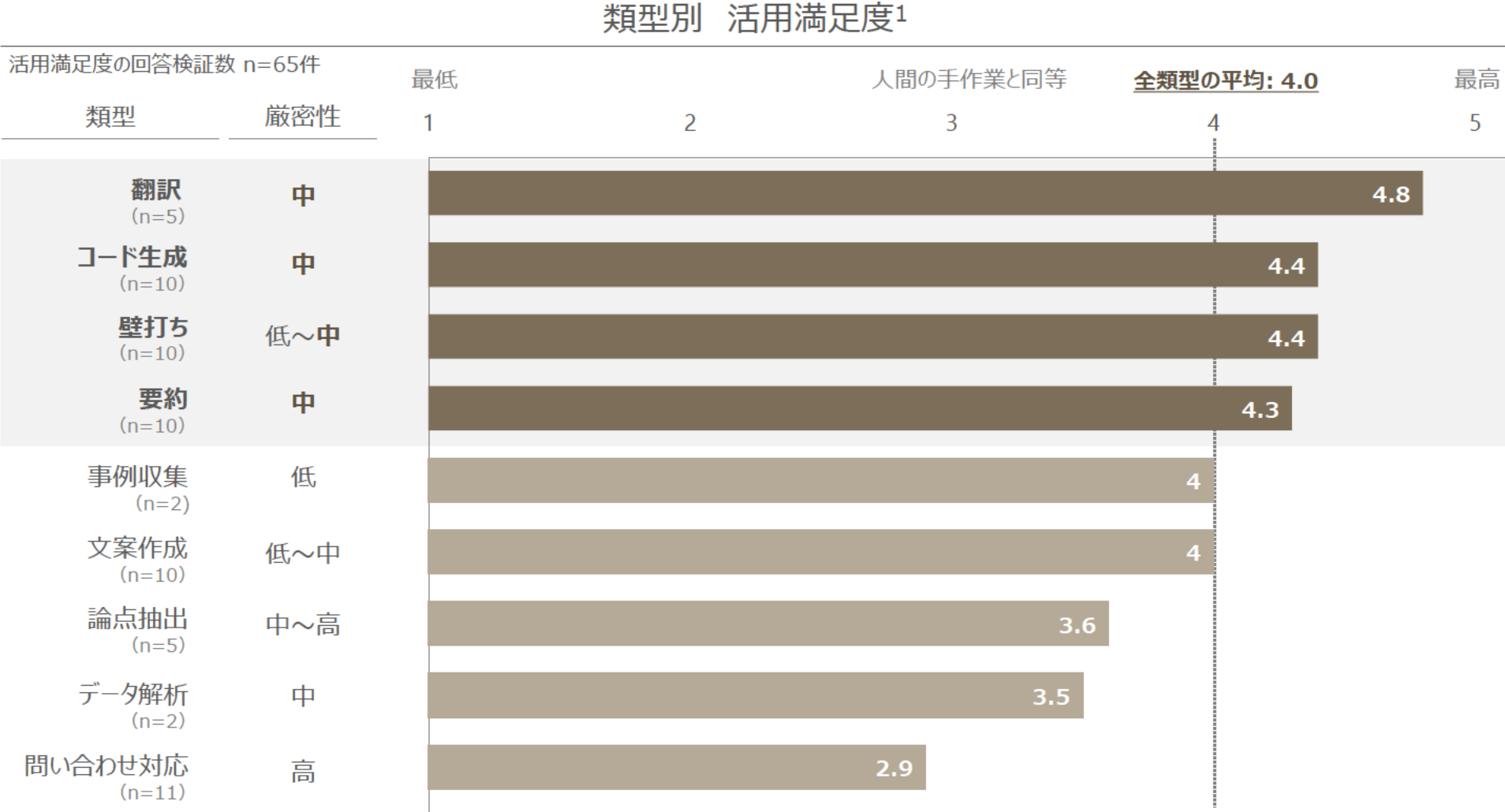


類型別 生成された回答の評価¹内訳



1. 最終検証期間においてユーザーから評価された個々の回答について集計
2. 「良い」が「悪い」を上回った場合は、「良い」と判定

翻訳やコード生成等の類型は、活用満足度が全類型の平均より高い傾向にある



1. 中間及び最終報告時に設定した評価項目。検証したタスクについて、各原課で生成AI活用に対する満足度を5段階で評価したものを類型別に平均し、集計

4.2. 検証結果の評価 定量分析：問い合わせ対応の課題

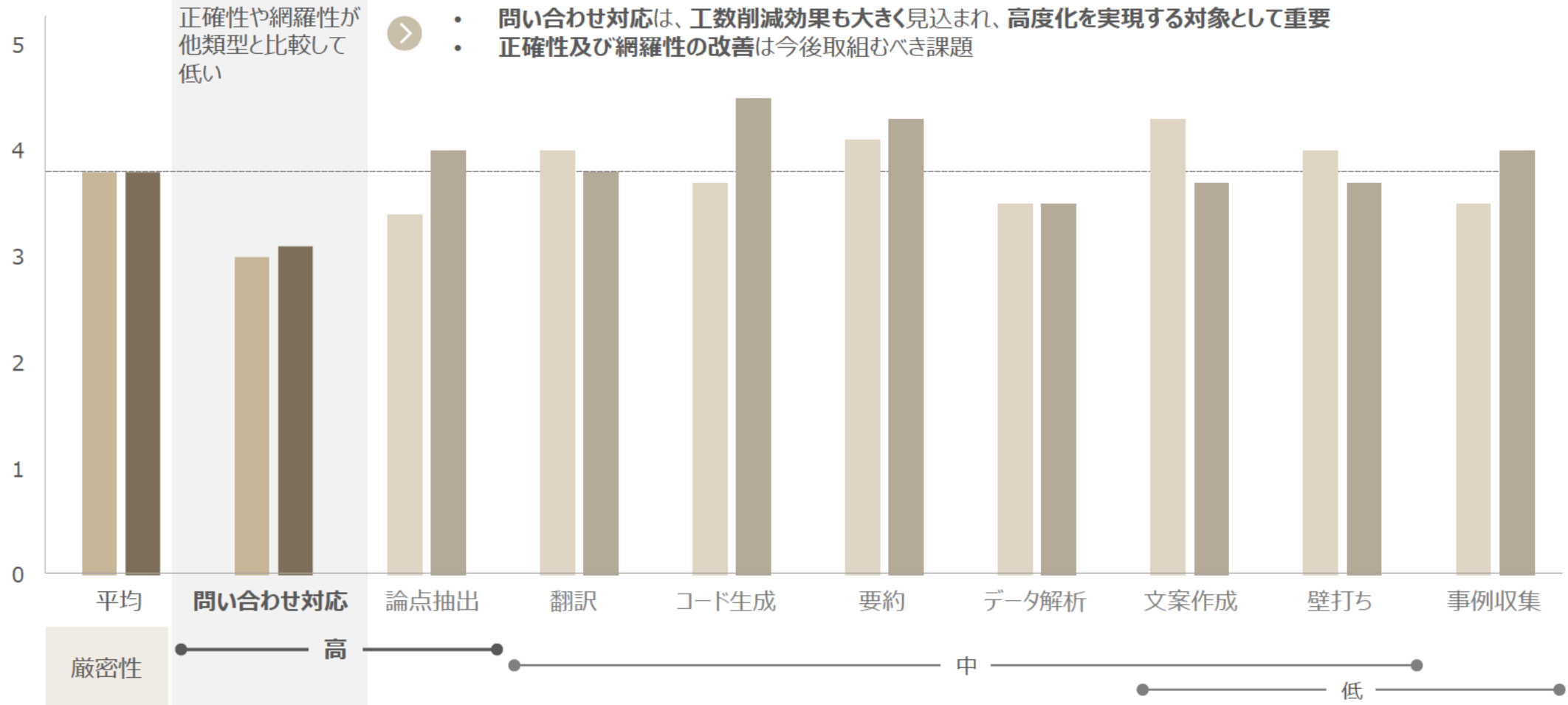
問い合わせ対応は厳密性を高く求められ、工数低減効果が見込まれるが、正確性や網羅性は全類型の平均値を下回る

正確性及び網羅性の評価 類型別

正確性、網羅性の回答検証数 n=65件

回答の正確性

回答の網羅性



4. ユースケース検証及び次期課題の整理

4.2. 検証結果の評価 検証結果の総括

STEP6

検証結果の集計・分析

定性分析及び定量分析の結果を踏まえ、今年度の検証により推定される類型別の業務適用の可能性をまとめた

業務への適用可能性

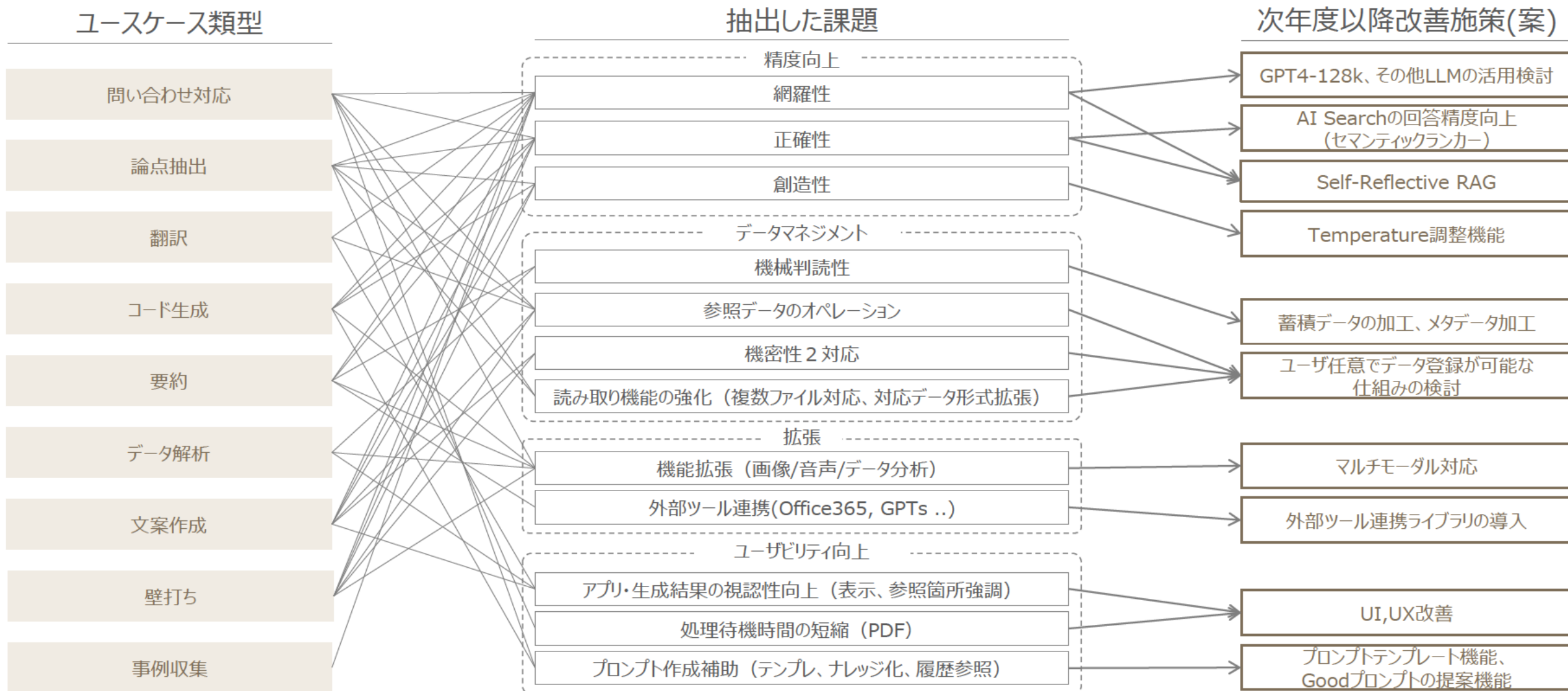
業務への利活用が限定的となる可能性が高い	- 回答の精度及び工数低減のともに一定の効果を確認したが、現時点では直接的な業務利活用は困難 - 特に問い合わせ対応は、工数削減効果が大きく、課題解決時の改善効果に期待
業務高度化に資する可能性が高い	- 回答の精度及び工数低減の効果ともに業務の高度化を実現する可能性が高いユースケース - 業務への活用を希望する声も大きい
工数削減効果は低いが業務高度化が見込める	工数の削減効果は小さいものの、特にアイデアの創出や案の作成などのユースケースにおいて補助する役割で有効

類型別の総括

問い合わせ対応	複雑な問いに対する回答は抜け漏れがあり、たたき台を作成する上では現時点でも有効だが、職員の手直し工数も大きい。検索の網羅性/正確性の精度向上は引き続き検討が必要
論点抽出	限定的な範囲においては一定水準の論点抽出ができたが、広い範囲を対象とする場合、一部のハルシネーションや参照箇所の漏れがあり、人手の確認は必須
翻訳	質の高い英語に訳出、英文校閲等が可能（専門業者よりも質が高いケースもあり） 一般的な無料翻訳サイトと比較してやや冗長な表現となる場合もある
コード生成	生成するコードはコードの修正も含めて問題ない業務利活用上、問題ないプログラミング知識がなくてもコード作成ができる点も大きな利点
要約	行政目的では確認が可能な範囲とはなるが、短時間での概要作成、内容の理解促進が可能。 膨大な英文を読むケースでは、翻訳と組合せて、要約を日本語で回答できることも利点
データ解析	グラフは傾向は読み取れないものの部分的な数値の読み取りは可能であることを確認したが、生成AIが不得手とする数値を取り扱う業務を想定しているため、引き続き検証が必要
文案作成	業務効率化にはつながらないが、企画の骨子作りやアイデア創出等、アウトプットの質向上に役立つ。 一般的な講演内容であれば、人間の介入なく、構成を作成することが可能
壁打ち	原案作成としては有用。従来に枠を超えた示唆を受け、実際に具体化しているケースもある。 人間よりもアウトプットは早い。非定型業務への活用に期待
事例収集	適切な事例を迅速に収集できるため大幅な工数削減が期待できる。他方、事例の一部のみの抜粋であるなど網羅性には課題がある

4.3. 重要検討課題の整理 課題・改善施策の全体像

定性評価から抽出された課題に対する改善施策の紐づけを実施した



1. 検証後アンケート調査:「業務効率化のため、検証環境に追加して欲しい機能があれば記載してください。」(自由記述)

課題の詳細化・改善施策検討にあたり、定性分析に加えアンケート¹も参照

4. ユースケース検証及び次期課題の整理

4.3. 重要検討課題の整理 注力すべき課題の特定

整理した課題に対し、利用者の維持・増加及び業務への適用可能性向上の観点から、注力すべき課題の絞込みを実施した

行政事務の高度化において重要な課題	注力すべき課題の特定		施策の検討
ユーザーからの声を元に、行政事務の高度化において重要な課題種別を特定	利用者の維持・増加や、業務への適用可能性を高めるために注力すべき課題を特定		注力すべき課題を解決し、より行政事務の高度化に資するための改善施策を検討
定性分析	利用動向分析（離脱要因）	定量分析	
精度向上	性能が業務要求レベルに未達	特に問い合わせ対応で課題の解決が求められる	問い合わせ対応における課題を解決でき、かつ幅広いユースケースに効果のある施策の検討が必要
データマネジメント	機密性のある情報が使えない	—	システムとしてどのように機密性のある情報を取り扱うべきかセキュリティの観点で検討が必要
機能拡張	—	—	現時点で直接的に離脱要因や業務への適用可能性の低下に繋がる課題とは考えにくい
ユーザービリティ向上	環境が使いづらい	—	生成AI技術が業務適用できるか否かによらずユーザーの離脱へ繋がっている可能性があるため、ユーザーが使いやすい仕組みの検討が必要
	業務にフィットしないユーザーのスキル起因	—	利活用ガイドライン作成や研修を実施することで対応（Appendix）

4.3. 重要検討課題の整理 コスト対効果・重要度評価

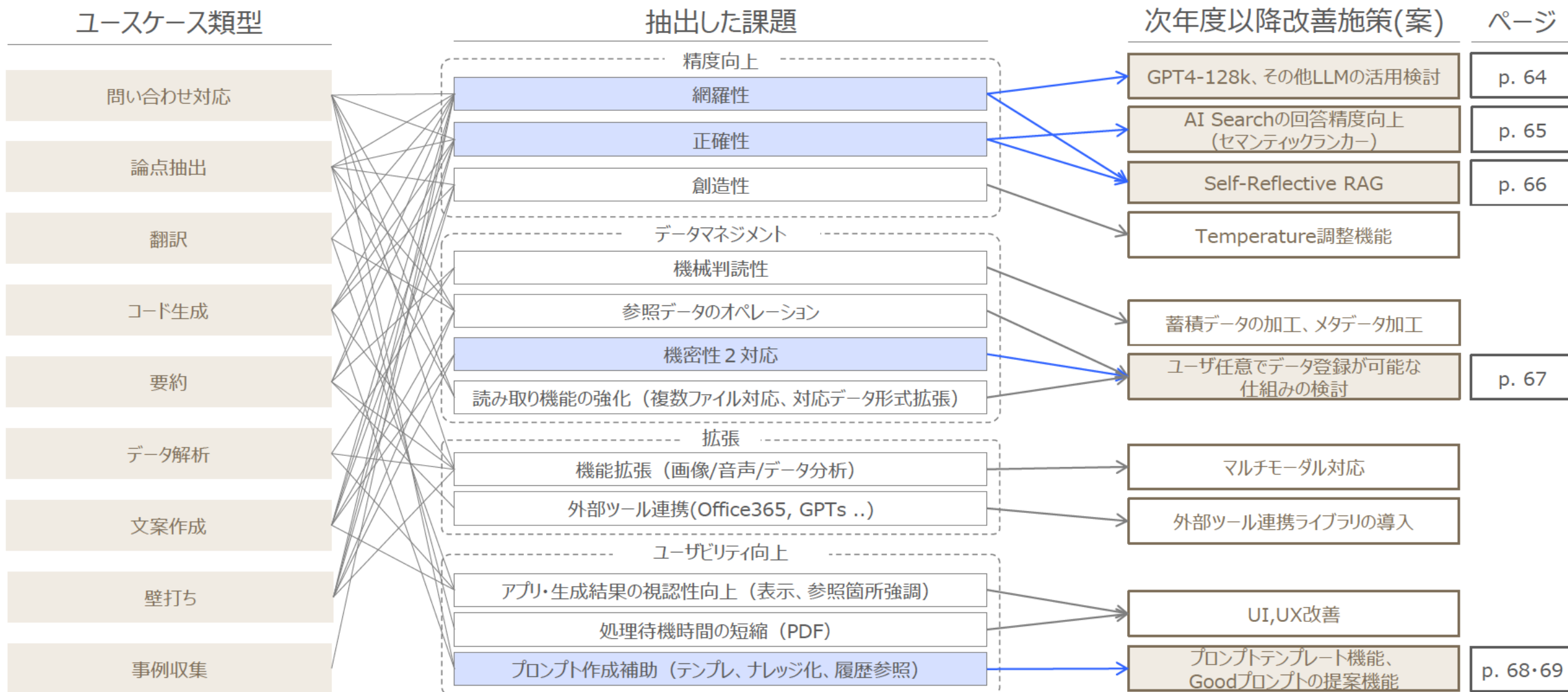
改善施策の検討

各改善施策に対し、コスト対効果、重要度評価を行い、改善施策に対し評価を行った

X 優先順位			コスト/難易度		
			低	中	高
システム改修			既存システムへの軽微な機能追加	既存システムへの大規模な機能追加	新システムの構築
効果\重要度	高	効果大かつ影響範囲も広い	1 • プロンプトテンプレート機能	3 • GPT4-128k、その他LLMの活用検討 • Goodプロンプトの提案機能 • ユーザ任意でデータ登録が可能な仕組みの検討 • Self-Reflective RAG	6
	中	効果大だが影響範囲は狭い 影響範囲は広いが効果は限定的	2 • AI Searchの回答精度向上（セマンティックランカー） 優先度高	5 • UI,UX改善	7 • マルチモーダル対応 • 外部ツール連携ライブラリの導入
	低	効果も限定的で影響範囲も狭い	4 • Temperature調整機能	8 • 蓄積データの加工、メタデータ加工	9 優先度低

4.3. 重要検討課題の整理 重要検討課題の絞り込み

コスト対効果・重要度評価から重要検討課題を絞り込み、重点的に取り組むべき改善施策を明らかにした



回答の網羅性向上の改善施策として、GPT4-128k、その他LLMの活用を検討する

網羅性の課題認識と施策

本検証のなかでも、RAGプロセスを利用した機能(URL,PDF Answering, AI Search)を用いたユースケース検証において、**網羅性**が課題となり、業務適応レベルに至っていないという意見が多かった。

網羅性を高める施策として、**コンテキストサイズが大きいモデルの採用**が挙げられる

昨今、LLMのモデルサイズが増々拡大しており、それに伴い、利用可能なコンテキストサイズも大きくなっている

これにより、一度の問合せで参照させられる文書量が多くなり、状況においては、トークン分割せずに参考文書をそのまま参照させられる可能性も高まっている

今後の展望

- ✓ Azure OpenAIモデルの場合、「**GPT-4-Turbo**」を利用することで、32Kから128Kと4倍の参照量を確保可能となる
 - ✓ その他パブリッククラウドの場合、昨今、Anthropic社が提供しているLLMモデル「**Claude**」もコンテキストサイズが大きいモデルを公開している(Amazon BedrockでAPI提供されている)
- これらAPIプロバイダが提供しているモデルを今後、注視しつつ、最適なモデル採用を目指す必要がある

Azure OpenAIモデルの料金体系¹

モデル	コンテキストサイズ	入力	出力
GPT-3.5-Turbo	4K	\$0.0015	\$0.002
GPT-4	32K 本検証環境利用	\$0.06	\$0.12
GPT-4-Turbo	128K	\$0.01	\$0.03

1. [Azure OpenAI Service - 価格 | Microsoft Azure](#)
2. [LLM API Provider Leaderboard | Artificial Analysis](#)

LLM APIプロバイダリーダーボード²

N o	モデル	コンテキスト サイズ	モデル精度 (Normalized avg)	費用 (USD/1M tokens)	スループット (token/s)	レイテンシー First Chunk(s)
1	Claude 3 opus	200K	100	\$30.00	22.9	1.56
2	Claude 2.1	200K	59	\$12.00	32.1	0.61
3	GPT-4-Turbo	128K	100	\$15.00	8.6	0.65

✓ モデル回答精度：OpenAI GPT4 Turboを基準に精度比較
✓ コンテキストウィンドウ：大きいほど、一度に参照できる文書量が多くなる
✓ スループット：単位時間当たりの処理量。高いほど良い
✓ レイテンシー：回答速度。低いほど良い

4.4. 今後のシステムのあるべき姿 セマンティックランカー

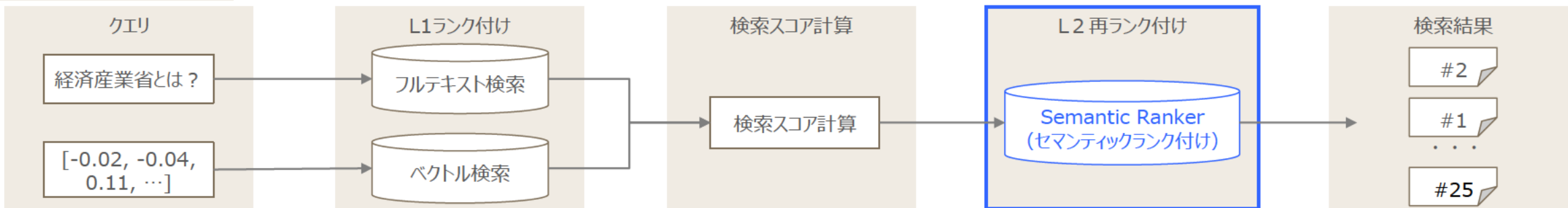
回答の正確性向上の改善施策として、AI Searchの機能であるセマンティックランカーの導入を検討する

AI Search機能の課題と施策

本検証のなかでも、AI Search機能への期待が高かった分、AI Searchを用いたユースケース検証では、**正確性**が十分ではなく改善してほしい意見が集まった
今後、省内データの利活用をしていくにあたり、AI Search機能の利用を促すうえでも、以下Semantic Rankerによる正確性向上は十分、検討の余地があるとする

検索オプション	詳細	フィルター/メタデータによる絞り込みの併用
Hybrid search（ハイブリッド検索） ¹	キーワード検索とベクトル検索を組み合わせる検索オプション。 ①全文検索：BM25を用いた単語ベースでのランク付け ②ベクトル検索：cos類似度によるベクトル情報に基づいたランク付け ③ ①②の相互スコアを評価し（RRF）、候補をランク付けする	○
Semantic Ranker （セマンティックランク付け） ²	Semantic RankingというMS提供のアルゴリズムを利用し、 検索精度を向上させることが可能なオプション。 但し、リクエストごとに課金される。（¥147per1,000request）	○

ハイブリッド検索のフロー



1. [Hybrid search scoring \(RRF\) - Azure AI Search | Microsoft Learn](#)
2. [セマンティック ランク付け - Azure AI Search | Microsoft Learn](#)

回答の正確性、網羅性向上の改善施策として、Self-Reflective RAGの導入を検討する

RAG(情報参照による回答生成)の課題認識と施策

本検証のなかでも、RAGプロセスを利用した機能(URL,PDF Answering, AI Search)を用いたユースケース検証において、**正確性**が課題となり、業務適応レベルに至っていないという意見が多かった

これらの課題として、まず質問クエリの回答となる検索結果が一回では得られていないケースが見受けられる

適切な検索結果が出るように、ユーザが質問クエリを調整することは現実的に難しく、ユーザの質問に対し、適切な検索結果・根拠を取得する手段として、Self-Reflective RAGに着目している

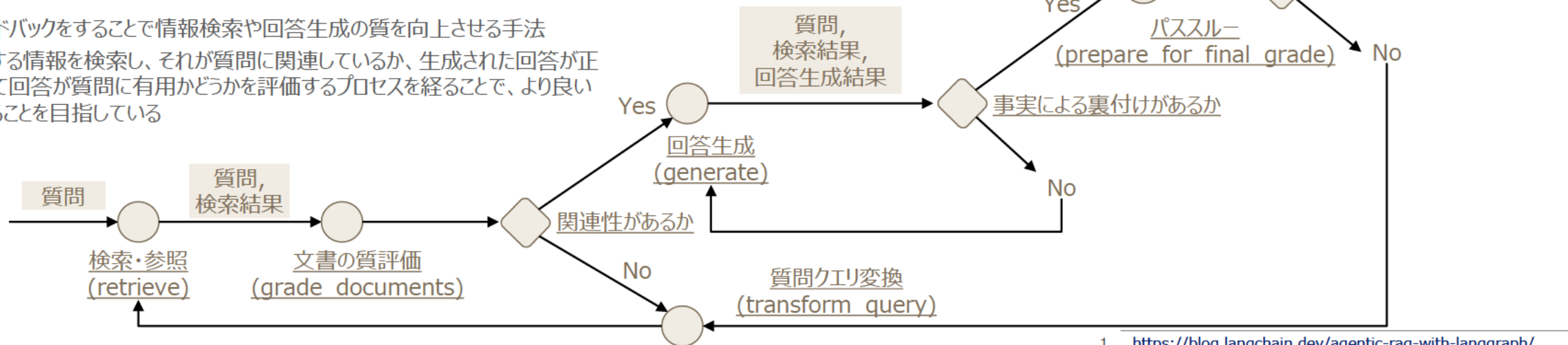
今後の展望

- ✓ 省内データの暗黙知となっている資料を問合せすることで、データの利活用を図ることは、今後の行政事務高度化において重要なテーマであると考え
- ✓ Self-Reflective RAGの方法論を活用し、適切な検索結果を得られる処理フローへ改善していくことが、今後の回答の正確性を高め、業務利用可能なレベルの実現に向けて有効な一手になるのではないかと考える

Self-Reflective RAG ^{1,2}

自己フィードバックをすることで情報検索や回答生成の質を向上させる手法

質問に関する情報を検索し、それが質問に関連しているか、生成された回答が正確か、そして回答が質問に有用かどうかを評価するプロセスを経ることで、より良い結果を得ることを目指している

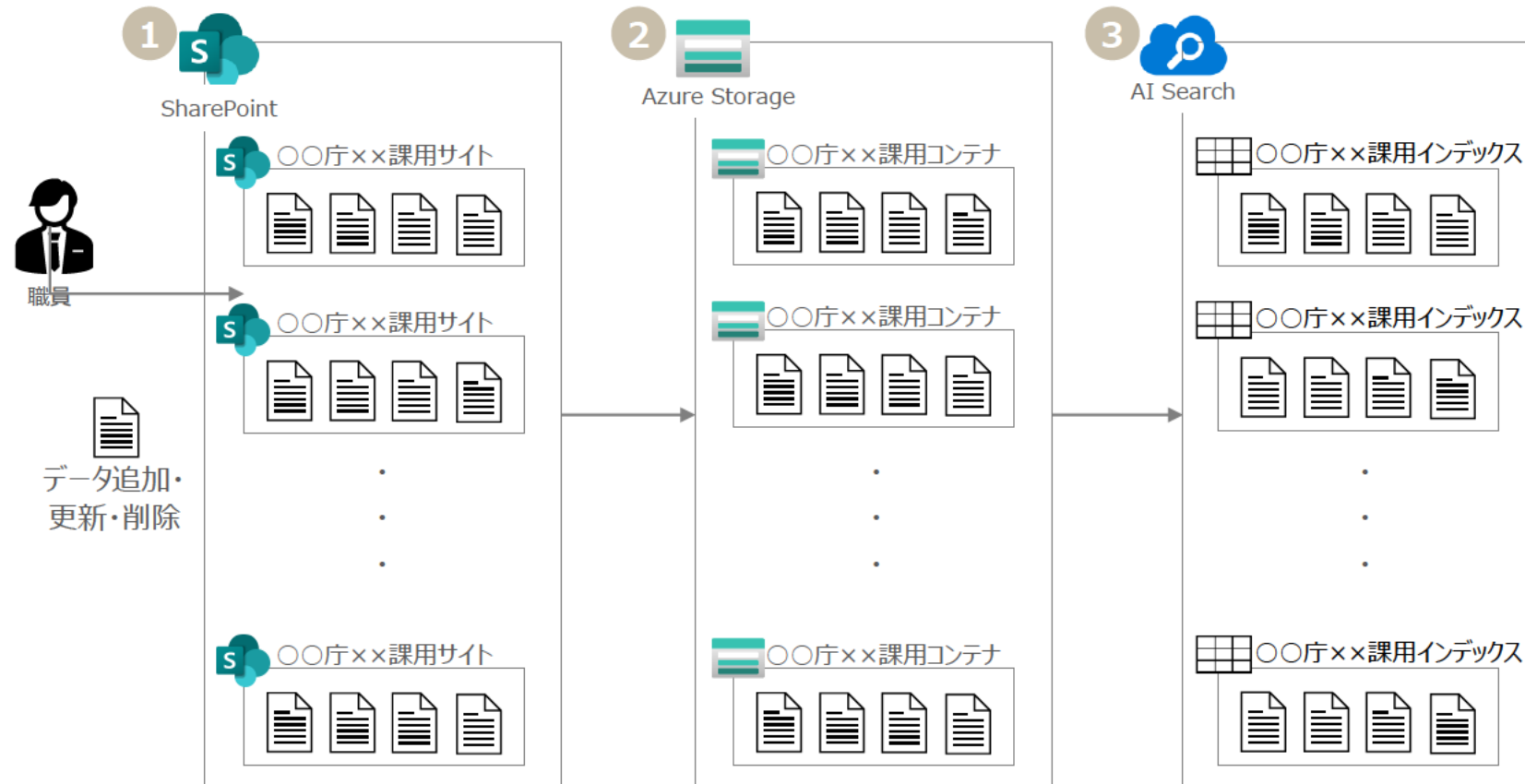


1. <https://blog.langchain.dev/agent-rag-with-langgraph/>
2. <https://arxiv.org/abs/2310.11511>

4.4. 今後のシステムのあるべき姿 ユーザ任意でデータ登録が可能な仕組みの検討

機密性を含む情報へ対応するための改善施策として、高いセキュア環境上でユーザが任意にデータ登録可能な仕組みを検討する

データ登録の仕組み



1 SharePointでのファイル管理

SharePointに課ごとにサイトを作成し、課ごとに機密性のある情報を管理する。職員には所属する課のサイトへのアクセス権のみを付与し、他組織のデータへのアクセスを制限する

2 Azure Storageへファイルを格納

SharePointに格納されたファイルをAzure Storageのコンテナにコピーする。課ごとにコンテナを作成してデータを管理する

3 検索用インデックス作成

Azure Storageに存在するコンテナごとに検索用のインデックスを作成する。課ごとにインデックスを作成することでアプリケーション上でデータを利用する際も各ユーザーの権限に合わせてデータへのアクセスを制限できる

ユーザビリティ向上の改善施策として、プロンプトテンプレート機能の実装を検討し、その要件を整理する

プロンプト作成の課題認識

本検証のなかで収集した意見の中で、「どのような質問をしたらいいかわからない」といった意見を確認している

アプリの利用し始めの職員などは自由質問型であると、どのような質問が良いか検討する必要があるため、重要検討課題として挙げている

プロンプトをゼロから検討するのは業務負荷にもなるため、あらかじめテンプレートとなり得るプロンプトをユーザが選択可能とするなどして、プロンプトの修正工数の低減、さらにはプロンプトの均一化を図ることが求められると考える

想定している施策

上記課題に対する施策としてプロンプトテンプレート機能を想定している

プロンプトの内容が記載されたコンポーネントが複数配置されており、当該コンポーネントを押下すると、質問入力エリアにプロンプト情報が配置され当該エリアにて編集可能になる、あるいは、プロンプトがコピーされ、質問入力エリアにペースト可能になる

今後の展望

- ✓ 業務の中で職員各位が検討・考案したプロンプトは、今後、アプリを利用する職員にとって再利用価値のあるものである
- ✓ 今後、利用ユーザの拡大、持続的な利用を促していくためには、このようなUI,UX面での改善を継続的に進めていく必要があると考える

プロンプトテンプレート機能の画面イメージ

ユースケース

文案作成

問い合わせ対応

壁打ち

要約

事例収集

論点抽出

翻訳

データ解析

コード生成

選択ユースケース：文案作成

xxシーンの
文案作成プロンプト

Copy All

yyシーンの
文案作成プロンプト

Copy All

zzシーンの
文案作成プロンプト

Copy All

⋮

4.4. 今後のシステムのあるべき姿 Goodプロンプトの提案機能

ユーザビリティ向上の改善施策として、Goodプロンプト蓄積・提案によるプロンプト工数削減施策を検討する

プロンプトノウハウの課題認識

前項「今後の展望」にて記載した、業務の中で職員各位が検討・考案したプロンプトは、職員それぞれが持つノウハウになっており、今後、アプリを利用する職員にとって再利用価値のあるものであるが、個々に閉じてしまう課題も想定される

想定している施策

職員がアプリを利用して生成結果が優れている場合、Goodボタンを押下してもらうことで、その質問(プロンプト)と回答のセットを好事例として蓄積し、ノウハウの蓄積を目指す

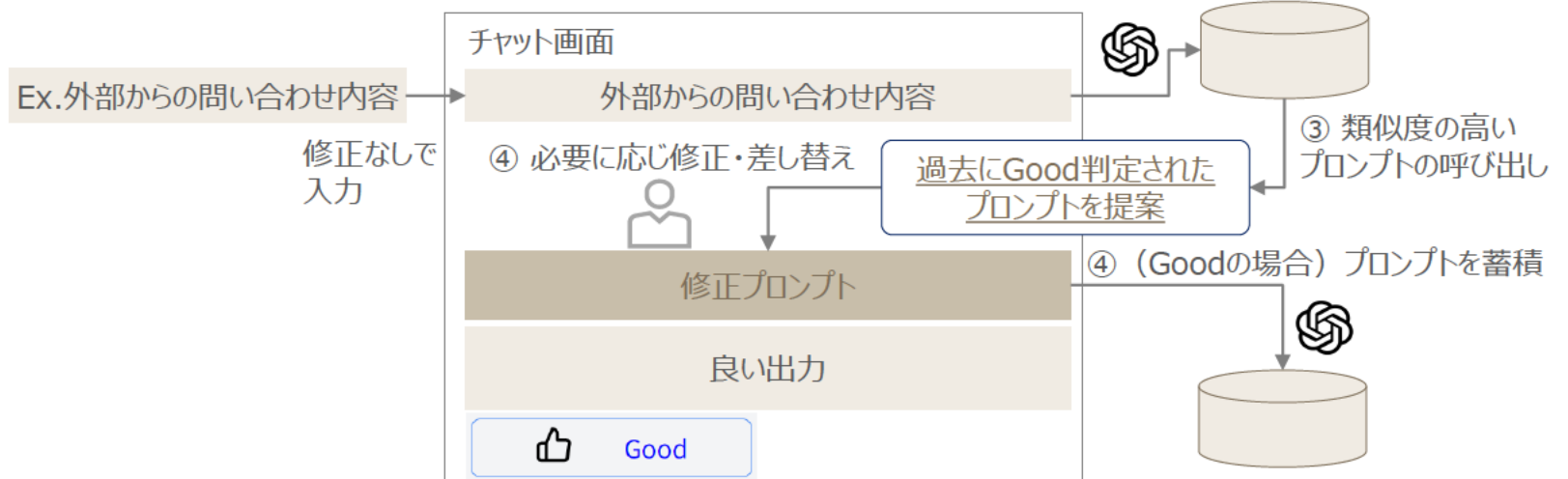
この蓄積データを活用し、外部からの問合せを解釈して、過去の好事例だったプロンプトを提案することで、良いプロンプトを再利用可能な仕組みを検討する

良質なプロンプトを提案するための仕組み

(1) “Good”プロンプトのDB蓄積

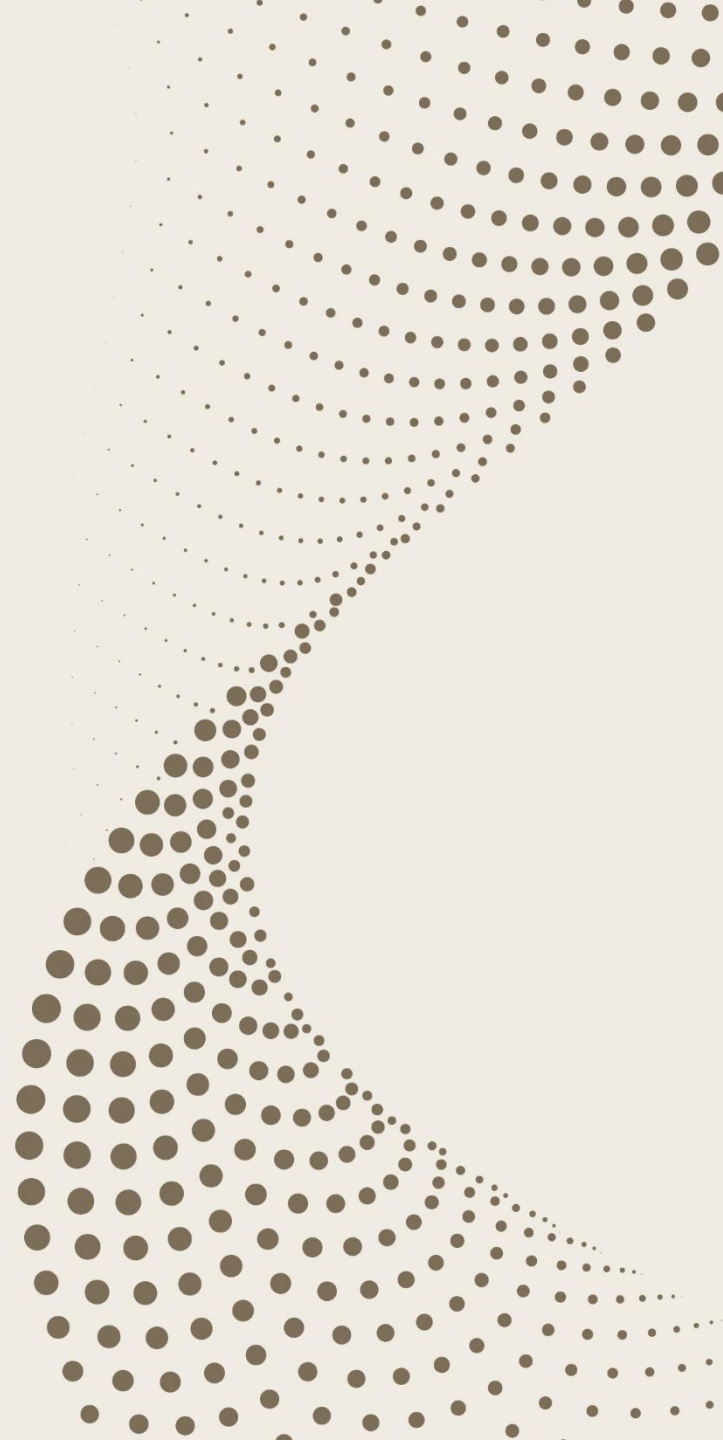


(2) 蓄積されたDBから“Good”プロンプトを提案



5. 生成AI利活用を前提としたデータに関する調査結果

- 5.1. 非構造データの利活用とデータマネジメントの目的
- 5.2. 機械判読性の調査
- 5.3. 権限管理とセキュリティ
- 5.4. 持続可能な運用の仕組み



5. 生成AI利活用を前提としたデータに関する調査結果

5.1. 非構造データの利活用とデータマネジメントの目的

省内に埋もれている資料を政策立案の高度化に利活用すべく、効率化・持続可能性を追求したデータマネジメントを検討する

非構造データの利活用における経済産業省の目標

「暗黙知として埋もれた過去の資料」をデータベースとして一元化して政策立案の高度化を図るとともに、職員が様々な領域から手動で検索する手間を効率化する



想定手段

- 職員の工数や機能の拡大撤退の柔軟性等の持続可能性も含めたメリット・デメリットも考慮した上での検証を実施する
- R5年度の検証事業の中で、生成AIで読み取るデータの形式を比較するほか、省内のデータを広く生成AIに格納する

持続的なデータマネジメントにおける検討要素

機械判読性

生成AIにより判読可能なデータ形式がどのようなものを明らかにすることで、必要に応じて文書作成ルールの策定と、職員自身による機械判読可能なデータの判断を可能にする

権限/セキュリティ管理

行政機関におけるデータのアクセス権限は、各所掌範囲等に基づいて厳密に管理されていると推察されるため、権限に対して柔軟に対応できるシステムの構築し、不適切なアクセスや情報漏洩の防止を図る

持続可能な運用の仕組み

職員自身が大量の文書情報を取り扱うにあたり、データ登録の運用上の負担を軽減することは持続性に大いに影響を与えるため、登録されたデータ自体の情報陳腐化の防止策を検討しつつ、持続的な運用方法の実現を図る

5. 生成AI利活用を前提としたデータに関する調査結果

5.1. 非構造データの利活用とデータマネジメントの目的 As-Is/To-Beの整理

機械判読性・権限/セキュリティ・持続可能な運用に対し、As-IsとTo-Beを整理する

機械判読性

As-Is



文書が機械判読性が良い/悪いの基準が不明確



ルール化等の標準化をすべきかどうか判断できない



To-Be



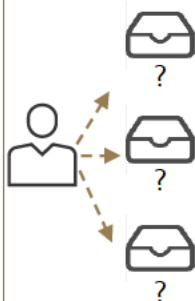
機械判読性の良し悪しが明らかである



必要に応じて、ルール化等の施策を打つことができ、データの管理が標準化される

権限管理/セキュリティ

As-Is

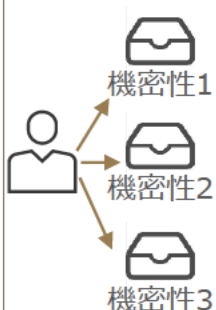


データの取扱いにおける権限管理の全容が把握できていない

セキュリティの観点で、データの権限および機密性のレベルに応じた取扱いを検討する必要がある



To-Be



各データの権限管理が明確になり、かつ、権限の追加・変更・削除といった権限の更新にも柔軟に対応できる

データの機密性に応じたセキュリティレベルが担保される

持続可能な運用

As-Is



データの登録は手作業での更新が必要なため、運用上の負荷が高い

情報が追加がされなくなり、登録されている情報の鮮度が低下していく可能性がある



To-Be



データの登録は職員がいつでも容易にできる



情報は常に最新のデータが登録され、継続できに利用する価値のあるツールになる

5. 生成AI利活用を前提としたデータに関する調査結果

5.2. 機械判読性の調査 機械判読性調査の全容

Azure AI Searchを用いた機械判読性の良し悪しを明らかにするために、下記調査を実施した

調査項目		対象ファイル	調査観点	調査結果（各検証項目に対して1～2データで検証）
ファイル形式 による精度 比較	PowerPoint ⇄ PDF	- PowerPoint - PDF（変換後）	PowerPoint/PDFで同じ質問を投げ、以下の観点を確認 ① 正確性 : 虚偽の情報なく想定する回答が得られるか ② 網羅性 : 想定する回答を漏れなく得られるか ③ 図表情報の理解度 : 図表情報を取得し、文脈を理解できるか	① PDFの方が高い ② PowerPointの方が高い ③ オブジェクトやグラフの文脈をある程度理解して取得可能
	Word ⇄ PDF	- Word - PDF（変換後）	Word/PDFで同じ質問を投げ、以下の観点を確認 ① 正確性 : 虚偽の情報なく想定する回答が得られるか ② 網羅性 : 想定する回答を漏れなく得られるか ③ 図表情報の理解度 : 図表情報を取得し、文脈を理解できるか ④ 構造化の有効性 : スタイルの有無が精度向上に寄与するか	① Word/PDF共に高い ② Word/PDF共に高い ③ 表の文脈をある程度理解して取得可能 ④ スタイルの有無による大きな精度差はない
テーブル構造の理解		- Excel - PDF（変換後）	① 単純なテーブル構造の理解度 : 1セル1単語の表を用いて、テーブル構造を把握可能か ② 複雑なテーブル構造の理解度 : 1セル複数文章で複数列に及ぶ表を用いて、テーブル構造を把握可能か	① Excel/PDF共に基本的に把握不可。質問によっては稀に正しく返答されることがある ② Excel/PDF共に把握可能
縦書きファイルの理解		- PDF（縦書き版） - PDF（横書き版）	① 正確性 : 虚偽の情報なく想定する回答が得られるか ② 網羅性 : 想定する回答を漏れなく得られるか	① 縦書き/横書きの間に差はない ② 縦書き/横書きの間に差はない ※法令データを参考にしたため、正確性・網羅性はやや低い
英語ファイルの理解		- PowerPoint（英版） - PowerPoint（日版）	日本語/英語で同じ質問を投げ、以下の観点を確認 ① 正確性 : 虚偽の情報なく想定する回答が得られるか ② 網羅性 : 想定する回答を漏れなく得られるか	① 英語の方が低く、一部誤った情報を取得する ② 英語の方が低く、情報が取得できないこともある ※日本語で質問し、日本語で回答された
文章と図表の割合に対する 文書理解		PDF（委託報告書）	① 各割合に対する文書理解度 : 文章と図表の割合によって、どの程度の情報取得が可能であるか（検索/内容把握/思考補助）	※P74参照

5. 生成AI利活用を前提としたデータに関する調査結果

5.2. 機械判読性の調査 委託調査報告書の機械判読性調査

文章の割合が多いほど、委託調査報告書の検索・内容把握・思考補助としての活用が容易であった

委託調査報告書（PDF）の特徴 ※比率は目安

1 大部分が文章で構成されている報告書

(割合) 文章：9割 図表：1割

2 図表と比べて文章の割合が多い報告書

(割合) 文章：7割 図表：3割

3 文章と比べて図表の割合が多い報告書

(割合) 文章：3割 図表：7割

4 大部分が図表で構成されている報告書
(図表内の文字を取得可)

(割合) 文章：1割 図表：9割

5 大部分が図表で構成されている報告書
(図表内の文字を取得不可)

(割合) 文章：1割 図表：9割

想定プロセス

所望の報告書の検索

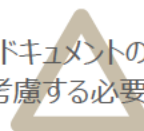
検索した報告書の内容把握

報告書内容を元に思考補助

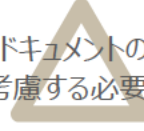
○：対応可能 △：要検証 ×：対応不可



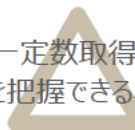
プロンプトやドキュメントの質を上げることも考慮する必要がある



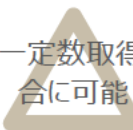
プロンプトやドキュメントの質を上げることも考慮する必要がある



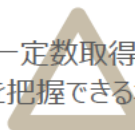
図表情報を一定数取得できるうえ、図表の文脈を把握できる場合に可能



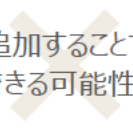
図表情報を一定数取得できている場合に可能



図表情報を一定数取得できるうえ、図表の文脈を把握できる場合に可能



OCR機能を追加することで図表情報を取得できる可能性がある



5. 生成AI利活用を前提としたデータに関する調査結果

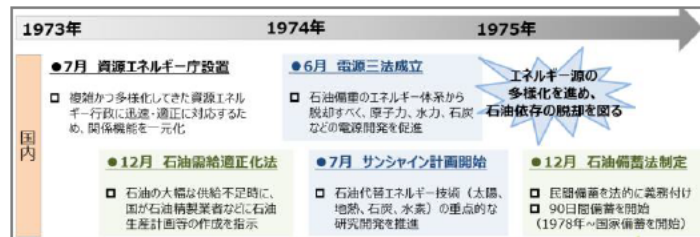
5.2. 機械判読性の調査 図表（グラフ・オブジェクト・表）の機械判読性調査

図表の文脈や構造を完全に把握することは難しいが、LLMの性能により概ね理解したうえでの回答が可能であった

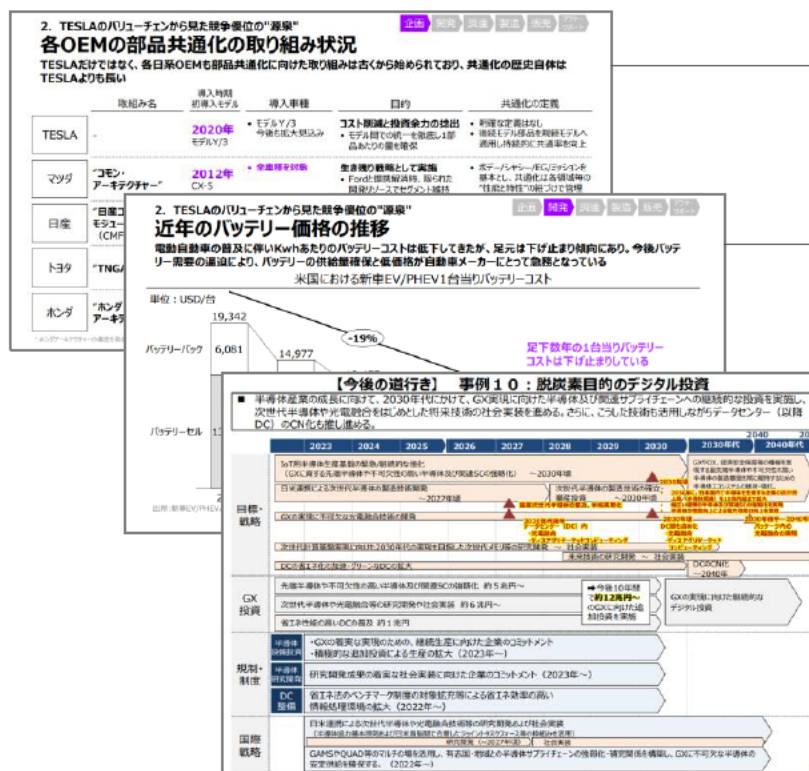
AI Searchで取得されるテキスト

LLMで生成させた検証結果

例 PowerPointオブジェクトから取得されるテキスト



国内国外1973年1974年1975年●7月資源エネルギー庁設置複雑かつ多様化してきた資源エネルギー行政に迅速・適正に対応するため、関係機能を一元化●10月 第1次オイルショック●第4次中東戦争を契機に～



一部行列関係を誤って取得したが、各項目の内容を大筋を外すことなくLLMが回答できていた

ある程度並びに規則性があればLLMが回答しやすい

一部誤っている数値もあったが、グラフ付近の参考情報を元に、大筋を外すことなく回答できていた

付近に参考情報があればLLMが回答しやすい

オブジェクトの対応関係を維持した状態で、正確・網羅的に情報を取得できていた

ある程度並びに規則性があればLLMが回答しやすい

文脈や構造は把握されず文字の羅列になり、LLMが回答しにくい

文脈や構造を完全に把握することは難しいが、一定の文脈や構造を理解したうえで回答することが可能

LLM単体の性能で、一定の文脈や構造を理解したうえで回答することが可能である

LLMが理解しやすいよう前処理を施すことで更なる精度向上が見込める

検証における参考資料

<https://www.enecho.meti.go.jp/category/others/gx/data/03.pdf>

<https://www.meti.go.jp/medi lib/report/2021FY/050113.pdf>

5. 生成AI利活用を前提としたデータに関する調査結果

5.2. 機械判読性の調査 構造化された文書の機械判読性調査

スタイルの有無によって機械判読性（構造把握能力）に大きな差はなかった

Word形式の資料を用いた検証結果（目次の章配下にある情報を網羅して回答できるか質問）

例1 「経済産業省デジタルガバメント中期経営計画」の第Ⅱ章への問い合わせ



データ駆動型行政への転換に向けた方針を教えてください。

Ⅱデータ駆動型行政への転換に向けて

1.サービス改革と業務改革の徹底（BPR）

2.データ利活用環境整備

(1)データ基盤の整備

(2)データ利活用を見すえた行政サービスのデジタル化の推進

(3)内部業務におけるデータ利活用の推進

3.データ駆動型行政に向けたガバナンスの強化

(1)価値を生み出すガバナンス

(2)情報セキュリティ、個人情報保護、業務継続性の確保

(3)人材確保・育成

(4)デジタル庁が整備する共通機能の活用の徹底、実現プロセスの整備、推進体制の強化



スタイルの有無に関わらず、章・項配下の文章が長い項目が生成されなかった

太字：スタイル有・無の両方で生成されず

例2 サンプルとなる調達仕様書への問い合わせ



情報システムに求める要件を全て教えてください。

7.情報システムに求める条件

(1)機能要件

(ア)ログイン機能

(イ)コンテンツ生成機能

(ウ)参照又は学習に供する独自データ登録および学習機能

(エ)利用状況分析機能

(2)非機能要件

(ア)キャパシティ要件

(イ)利用環境要件

(ウ)ユーザビリティ要件

(エ)アクセシビリティ要件

(オ)可用性・継続性要件

(カ)性能要件

(キ)セキュリティ要件

(ク)教育要件

(ケ)移行要件

(コ)運用要件

(サ)保守要件



スタイルの有無に関わらず、章・項配下の項目が多い場合の生成が不安定だった

太字：スタイル有・無のいずれかで生成されず

スタイルの有無による構造把握能力に大きな差はない

5. 生成AI利活用を前提としたデータに関する調査結果

5.3. 権限管理とセキュリティ 機密性2情報の取り扱い

機密性2情報を省外で取り扱う場合、厳重なデータ管理とセキュリティ対策が求められる

※「政府機関等の対策基準策定のためのガイドライン（令和5年度版）」、「政府機関等の対策基準策定のためのガイドライン（令和5年度版）」に従い要件を整理

機密性2情報の管理要件

1 暗号化及び電子署名

暗号化及び電子署名の必要性を確認し、必要に応じて対応する。暗号化方式はガイドラインで指定されている方式に従う。また、クラウドサービス上で要機密情報を扱う場合は通信経路全般が暗号化されていることを確認する

2 省外へのデータ送信時のセキュリティ対策

省外へデータを送信する場合は安全確保に留意し、VPNや信頼できる通信経路を使用してデータを送信する

3 権限管理とアクセスの監視

保存する情報にアクセス制限を設定するなど、情報の格付及び取扱制限に従って情報を適切に管理するとともに、情報へのアクセスを監視する。また、データが不要になった場合は速やかに削除する

4 NDA締結

委託先において要機密情報を取り扱う場合は、秘密保持契約（NDA）を締結する

次年度における対応方針

- クラウド上のデータベース、ストレージのデータをすべて暗号化し、キー管理用のサービスを使用して暗号化キーを管理する
- クラウドサービス内の通信がすべて暗号化されていることを確認し、必要に応じて通信経路を暗号化する措置を講じる
- データ送信時に使用するサービス、通信経路を検討し、対象のサービスおよび通信経路を使用したデータ連携の安全性について確認する
- Entra IDを使用してユーザーごとに適切な権限を付与し、情報へのアクセスを制限する。また各ユーザーのアクセスログを取得し、機密情報へのアクセスを監視する
- 要機密情報の提供を受ける前に、秘密情報の定義や秘密情報の保持・管理などについて委託元と委託先で合意し、秘密保持契約（NDA）を締結する

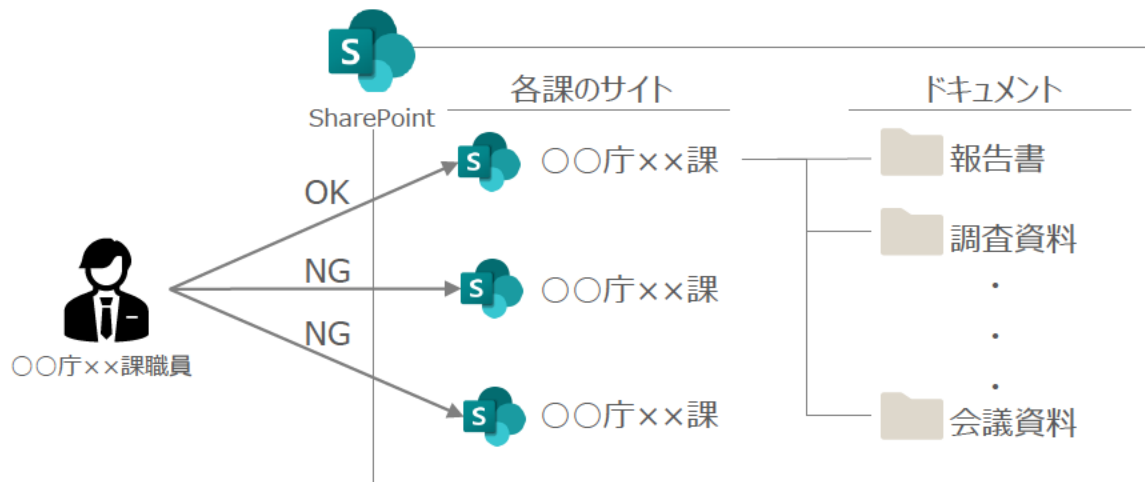
5. 生成AI利活用を前提としたデータに関する調査結果

5.3. 権限管理とセキュリティ 権限管理

機密性2情報は課ごとの非公開データとなるため、課ごとに参照可能なドキュメントの範囲を定義づけ、セキュアにデータにアクセスする仕組みを構築する

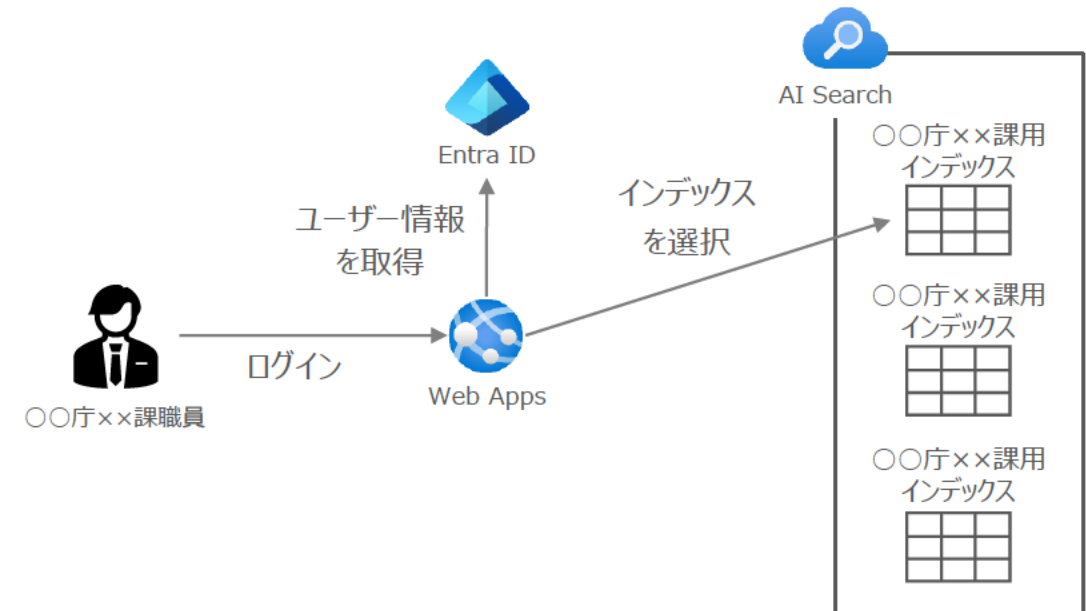
SharePoint上での権限管理

- SharePointに各課のサイトを作成して機密性2情報を管理
- 関係者のみが対象データにアクセスできるようにユーザー・グループに必要最小限のアクセス権限を付与



アプリ上での権限管理

- 機密性2情報に関しては検索用のインデックスも課ごとに作成
- アプリにログインしたユーザーの情報をEntra IDから取得し、対象ユーザーが保有する権限に対応するインデックスを選択して検索することで、権限外のデータが検索結果に出力されない

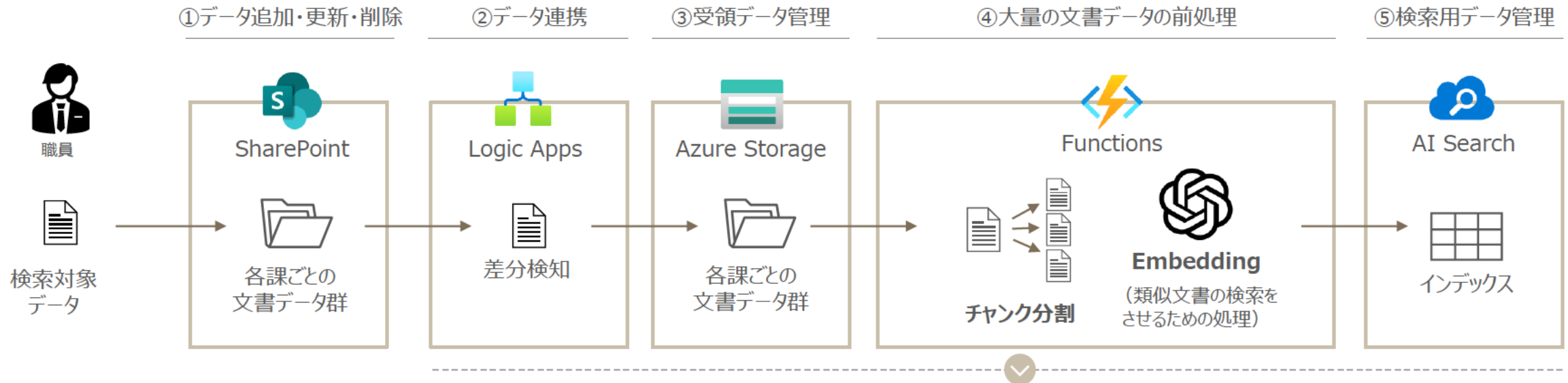


5. 生成AI利活用を前提としたデータに関する調査結果

5.4. 持続可能な運用の仕組み データ登録までの全体像

データの追加/更新/削除を容易に行えるように、データ登録のプロセスを自動化する

データ連携及びデータ処理のプロセス



データ連携からインデックス作成までを自動化

各プロセス詳細

- ① 職員が対象のデータを任意のタイミングでSharePointにアップロードする
- ② SharePointにデータがアップロードされるとLogic Appsがそれを検知し、データをAzure Storageに連携する
- ③ Logic Appsから連携されたデータがAzure Storage内に作成される
- ④ FunctionsとAzure OpenAIを使用してデータの前処理を行う。この処理はスケジュールされた時間に自動で実施される
- ⑤ 検索対象のデータがインデックス化され、AI Searchに格納される

5. 生成AI利活用を前提としたデータに関する調査結果

5.4. 持続可能な運用の仕組み Azure AI Searchで発生した課題一覧

本事業で発生したAzure AI Searchにおける課題と、その対応方法を一覧化する

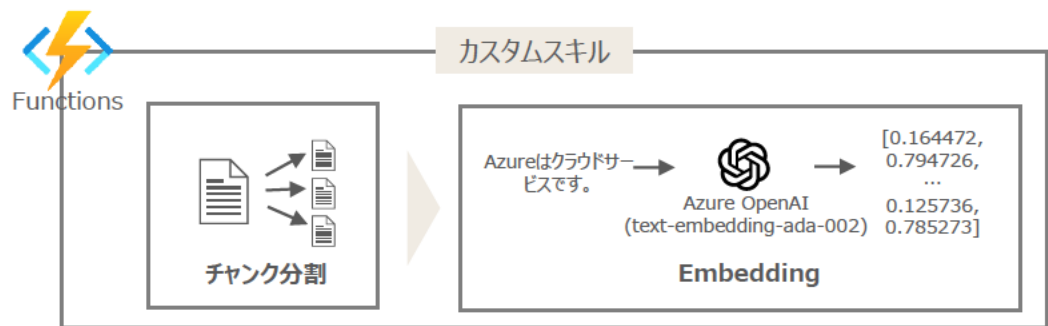
No	課題詳細	制限事項	対象 ファイル数	本事業での対応方法	来年度事業での対応想定
1	1ファイルあたりのサイズ制限が原因で 取り込み失敗	Basicプラン :16MB/1ファイル Standard1プラン:128MB/1ファイル Standard2プラン:256MB/1ファイル	0 ※検証時に発生	Standard1プランの利用で制限緩和	運用ルールを策定 ※P83参照
2	1ファイルあたりの文字数制限が原因 で取り込み失敗	Basicプラン :6.4万字/1ファイル Standard1プラン:400万字/1ファイル Standard2プラン:800万字/1ファイル	2	Standard1プランの利用で制限緩和 制限を超過するファイルは取り込み対象外	運用ルールを策定
3	1ファイルあたりの画像数制限が原因 で取り込み失敗	1000画像/1ドキュメント ※PDF1ページを1画像として扱うのでページ数制限と同義	2	制限を超過するファイルは取り込み対象外	運用ルールを策定
4	1ファイルあたりのファイル名の長さ制 限が原因で取り込み失敗	約70字 ※Blobパスを変換したキーの1024字制限が起因している	8	制限を超過するファイルは手動でファイル名を短 縮	運用ルールを策定
5	AIPや暗号化等の保護が原因で取り 込み失敗	AIP・暗号化ファイルは取り込み不可	0 ※検証時に発生	取り込み対象外	運用ルールを策定
6	インデクサーのタイムアウトが原因で取 り込み中断（長時間の実行時間）	2時間/1実行	相当数	当初は小分け投入により対応したが、カスタ ムスキルのスリープタイム設定の変更に伴い改善	本事業と同様に対応予定
7	一部のPDFのコンテンツを抽出できず、 テキスト取得失敗	-	10	カスタムスキルのスクリプト内で、ストレージから 直接テキストを取得することでエラー回避	本事業と同様に対応予定
8	一部のPDFが特殊な文字列で Encodeされており、文字化けが発生	-	5	課題No.7と同様の方法でテキストを取得する 際、PyMuPDFを利用することで解消	本事業と同様に対応予定
9	カスタムスキルのWebAPIコールのタイ ムアウトが原因で取り込み失敗	3分50秒/1コール	相当数	当初はファイル分割の前処理で対応したが、ス リープタイム設定を変更することで改善	Azure OpenAIの並列処 理も検討 ※P81参照

5. 生成AI利活用を前提としたデータに関する調査結果

5.4. 持続可能な運用の仕組み データ登録時のエラー解消（カスタムスキルのタイムアウト）

設定変更により処理速度が大幅に向上し、運用課題に直結していたタイムアウトエラーを解消した

タイムアウトエラーで想定された原因



原因①

1チャンクをEmbeddingする毎に、1秒間のスリープタイムを設定している

- Azure OpenAIのクォータ¹制限を回避するための待機時間を設けている
- スリープタイムが1秒の場合、230チャンク以上存在すると、その他処理の実行時間に関わらずタイムアウトが発生してしまう状況（3分50秒 = 230秒）

原因②

1つのAzure OpenAIからEmbeddingモデルを呼び出している

- 各チャンクを順に処理するため実行時間が長くなり、タイムアウトが発生しやすい状況

タイムアウトエラーに対する施策

施策①（実施済み）

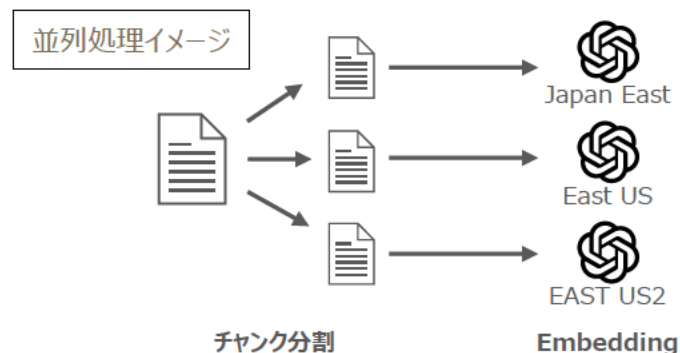
スリープタイム設定を処理に影響がない範囲で短縮（0.1秒など）する

結果

- 処理速度が大幅に改善され、分割処理なしでエラー無く処理が終了（規格外データを除く）
- 処理できるデータ量が大幅に増え、データを一括投入することが可能に

施策②（必要に応じて実施予定）

複数リージョンでAzure OpenAIを作成し、並列処理を実装することで負荷分散（クォータ制限の回避）と処理の高速化を狙う



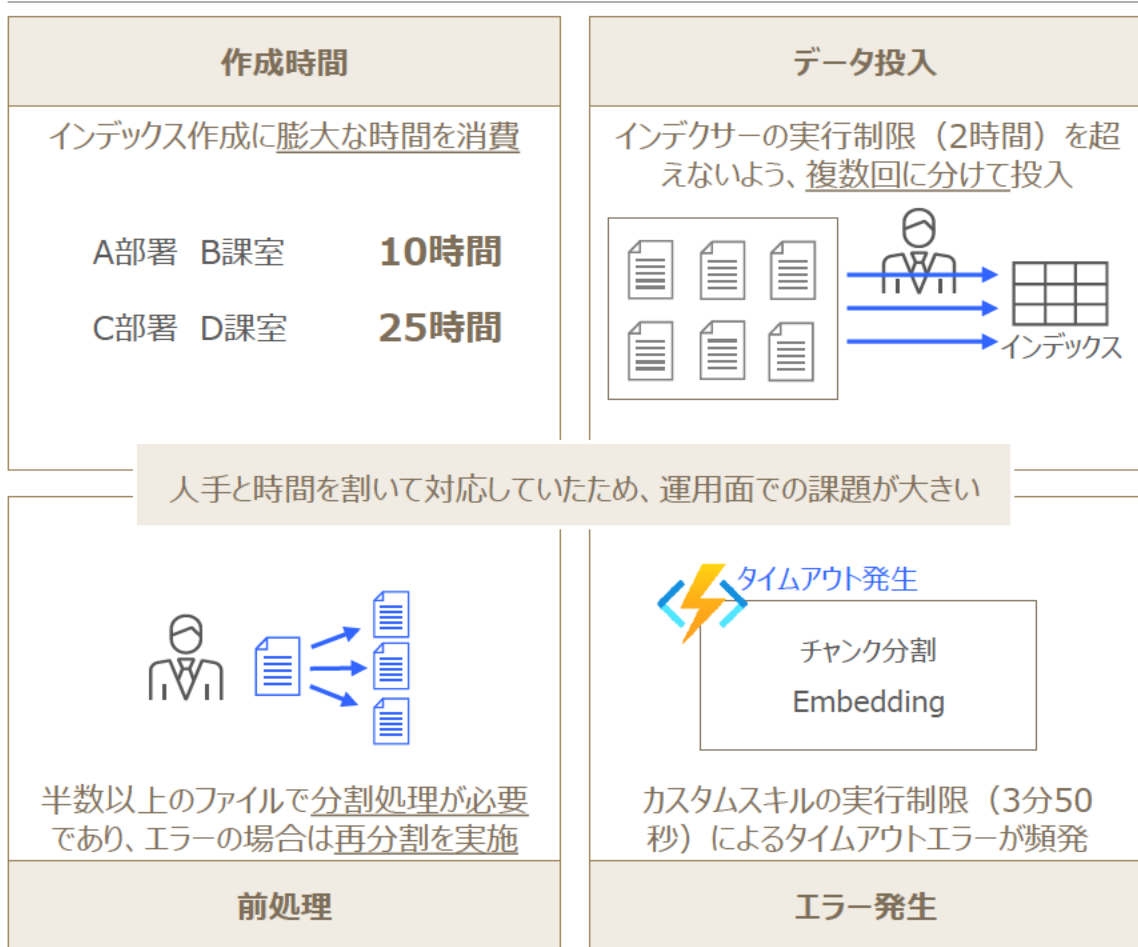
1. Azure OpenAIのリージョンごと、モデルごとに、TPM（Token Per Minute）単位で割り当てられるレート制限。デプロイした各モデルに対してTPMを割り当てることができ、制限を超えるとモデルをコールする際にエラーが発生する。

5. 生成AI利活用を前提としたデータに関する調査結果

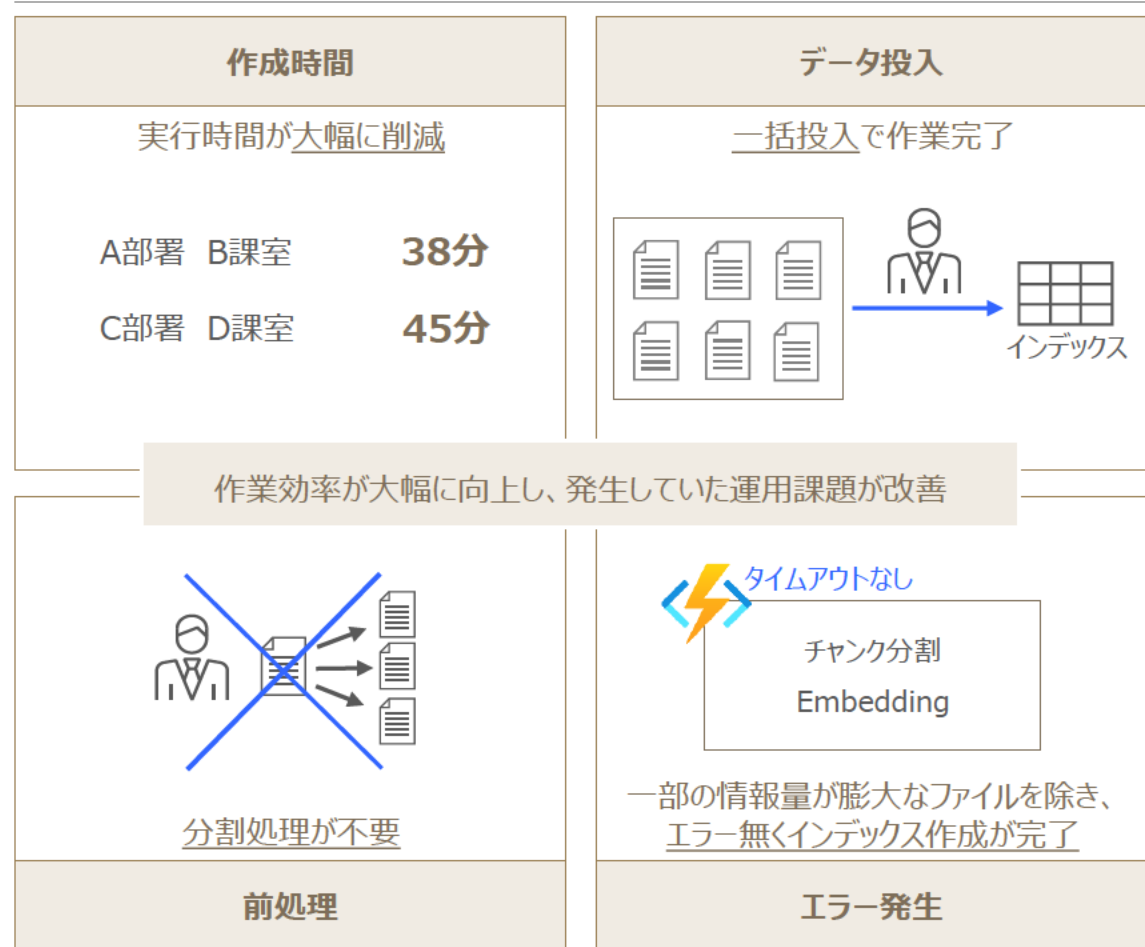
5.4. 持続可能な運用の仕組み インデックス作成効率の改善

タイムアウトエラーが解消されたことでインデックス作成の作業効率が大幅に向上し、運用課題が改善した

当初のインデックス作成状況



現在のインデックス作成状況



5. 生成AI利活用を前提としたデータに関する調査結果

5.4. 持続可能な運用の仕組み 制約事項を踏まえた運用ルール策定の

技術的な制約事項と経済産業省の業務要求を踏まえ、運用面でのルールを策定する

データの制約事項

1 ファイルの情報量

以下制限の範疇を超えるファイルは読み取り不可

- 容量 : 128MB以上 (256MB以上※)
- 文字数 : 400万字以上 (800万字以上※)
- ページ数 : 1000ページ以上 (PDFのみ)
- ファイル名文字数 : 約70字以上

※プランを上げた場合

2 保護ファイルの取り扱い

以下保護が施されているファイルは読み取り不可

- AIP保護 (秘密度がPublicの場合は可)
- 暗号化 (パスワード含む)

3 画像・オブジェクト・グラフ・表からの情報抽出

以下項目が困難なため、文書理解度が低下

- 埋め込み画像内のテキスト抽出
- テーブル情報の行・列関係の理解
- 情報量が少ないオブジェクトやグラフの文脈理解

4 外国語の理解

以下場面での回答精度が低下

- 外国語ファイルの内容に対し、日本語で質問

運用ルール候補

ルール候補① 制限の範疇を超えるファイルは取り込み対象外とする

ルール候補② 制限内に収まるよう、ファイルに加工処理を施したうえでSharePointに格納する

判断基準例

- 取り込まないことによる業務への影響度 (制限超過ファイルの数、ニーズ等)
- 加工処理されたファイルが取り込まれることの可否 (元ファイルとは異なるファイルが参照される)
- 機密性の高いファイルをはじめ、ファイルに加工処理を行うことの可否
- 業務の中でファイル分割やファイル名変更等の手作業を行うことの可否 (作業負荷)

ルール候補① 保護ファイルは取り込み対象外とする

ルール候補② 予めAIP保護や暗号化を解除したうえでSharePointに格納する

判断基準例

- 取り込まないことによる業務への影響度 (保護ファイルの数、ニーズ等)
- 職員が保護や暗号化を解除する権限の有無 (保護解除の容易性)

ルール候補 情報抽出が困難な場合を周知したうえでの利用とする

別途改善策

- OCR処理の追加 (有料)
- 情報抽出されやすいプロンプト例の提供

ルール候補① 外国語ファイルに対しては、対象となる言語で生成AIに問い合わせる

ルール候補② 予め外国語ファイルを日本語訳してSharePointに格納する

判断基準例

- 対象となる言語で問い合わせることが可能か (能力・作業負荷)
- 日本語訳されたファイルやテキストを用意することが可能か (作業の容易性・翻訳精度等)

5. 生成AI利活用を前提としたデータに関する調査結果

5.4. 持続可能な運用の仕組み Azure AI SearchとAmazon Kendraの運用観点での比較

Amazon Kendraの利用も想定し、Azure AI Searchの運用における制約事項を中心に比較する

		Azure AI Search	Amazon Kendra
情報量の制限	容量	128MB (Standard2は256MB)	50MB (※制限緩和可能)
	文字数	400万字 (Standard2は800万字)	約170万字 (最大5MBのテキスト量制限、約0.028MB/1万字)
	ページ数	1000ページ (PDFのみ)	なし (3000ページ以上の場合にエラーなし)
	ファイル名長さ	約70字	なし (受領ファイルの内、最長96字の場合にエラーなし)
保護ファイルの 取り扱い	AIP	取り込み不可	取り込み不可
	暗号化	取り込み不可	取り込み不可
画像・図表の 抽出	オリジナル	抽出可	抽出可
	埋め込み	抽出不可	抽出不可
運用性能	インデックス 作成時間	例 A部署 B課室 : 約40分	例 A部署 B課室 : 約45分
	同期処理	【差分検知】追加/更新/削除 【スケジュール実行】分/時間	【差分検知】追加/更新/削除 【スケジュール実行】時間/日/週/月
メタデータの取 得	ページ番号	デフォルトでは取得不可	デフォルトではPDFのみ取得可
	カスタム	Azure Functionsを用いて付与可能	AWS Lambdaを用いて付与可能

5. 生成AI利活用を前提としたデータに関する調査結果

5.4. 持続可能な運用の仕組み 総括

持続的なデータマネジメントにおける各検討要素のTo-Beに対し、現在の対応状況を整理する

機械判読性

To-Be



機械判読性の良し悪しが明らかである



必要に応じて、ルール化等の施策を打つことができ、データの管理が標準化される

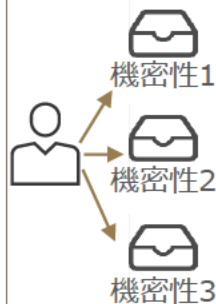


対応状況

- ファイル形式による精度差・構造化されたファイルの機械判読性・グラフや表の機械判読性等について調査し、生成AIに利活用する際の良し悪しを明らかにした
- 判明したデータの制約事項に対しては、予め運用ルールを検討。次年度策定・周知予定

権限管理/セキュリティ

To-Be



各データの権限管理が明確になり、かつ、権限の追加・変更・削除といった権限の更新にも柔軟に対応できる

データの機密性に応じたセキュリティレベルが担保される



対応状況

- SharePoint上（格納先）とアプリ上（利用先）で権限管理する仕組みを検討。次年度構築予定
- 機密性2情報を扱う上での管理要件を整理し、次年度に向けた対応方針を策定

持続可能な運用

To-Be



データの登録は職員がいつでも容易にできる



情報は常に最新のデータが登録され、継続できに利用する価値のあるツールになる

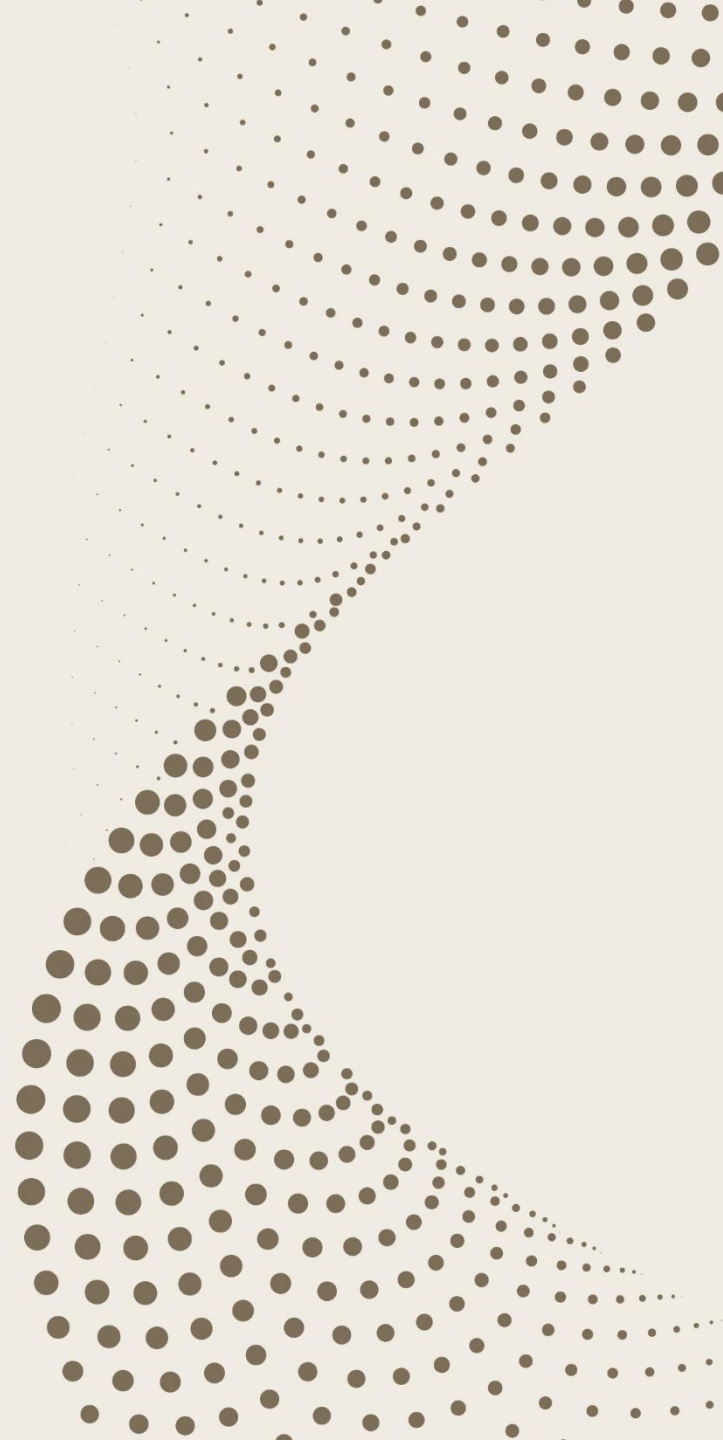


対応状況

- インデックス作成における運用課題を解決し、速く・確実なインデックス作成を可能にした
- SharePointからインデックス作成までのプロセスを自動化する仕組みを検討。次年度構築予定
- 判明したデータの制約事項に対しては、予め運用ルールを検討。次年度策定・周知予定

6. 本調査事業の総括

6.1. 総括



6. 本調査事業の総括

6.1. 総括

本事業の調査内容及び、当該技術を導入する際に具備すべき要件・制約事項のまとめとして、第4、5章を総括する

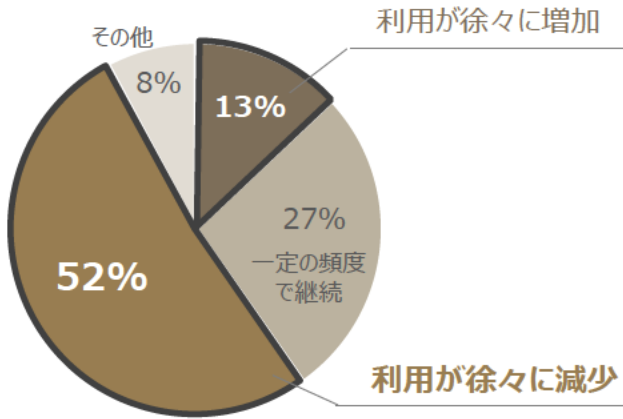
利用実態と離脱理由

第4章

ユーザーの離脱防止は省内の行政事務高度化において重要な課題であり、その具体的な理由も分析

利用実態

利用頻度が増えているユーザーが一部存在するものの、約半数が離脱傾向



主な離脱理由

事後アンケートから離脱理由を抽出・整理

- 性能が業務要求レベルに未達
- 環境が使いづらい
- 業務にフィットしない
- 機密性のある情報が扱えない
- ユーザーのスキル起因

離脱防止のための改善取り組み対象として重要な項目

業務への適用可能性の評価、課題抽出、及び施策案の検討

第4章

分析結果から業務適用可能性を整理。洗い出した課題に対し、生成AI技術を導入する際、具備すべき要件・施策案を検討

行政業務への適用可能性

適用可能性の評価	類型
業務高度化に資する可能性が高い	翻訳/コード生成/要約
工数削減効果は低いが見込める	データ解析/文案作成/壁打ち/事例収集
業務への利活用が限定的となる可能性が高い	問い合わせ対応/論点抽出 ※「機械判読性」の向上が重要

課題と施策案

洗い出した12つの課題を重要度・費用の観点で4つに絞込み、施策案を検討

課題	施策案（要件）
網羅性	・ GPT4-128k/その他LLMの活用検討
正確性	・ Azure AI Searchの回答精度向上 ・ Self-Reflective RAG
機密性 2 未対応	・ データ登録が可能な仕組みの検討 ※「権限管理/セキュリティ」の十分な考慮
プロンプト作成補助	・ プロンプトテンプレート機能、 ・ Goodプロンプトの提案機能

生成AI利活用を前提としたデータに関する調査結果

第5章

生成AIの業務導入時の制約事項として、データマネジメントと持続可能性の観点から考慮すべき重要事項をまとめた

機械判読性



- ファイル形式や構造化されたファイル違い、図表有無に依存する機械判読性の調査
- 生成AIに利活用する際の良し悪しを明らかにし、運用ルールを検討

権限管理/セキュリティ



- 機密性2情報を扱う上での管理要件を整理
- 権限管理する仕組みと具体的な対応方針案を策定した

持続可能な運用



(Azure AI Searchを前提)

- インデックス作成における運用課題を解決し、速く・確実なインデックス作成を検討し、実現
- SharePointからインデックス作成までのプロセスを自動化する仕組みを検討

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.



Build Beyond As One.