特集「AIとデータ活用が拓く新たな社会価値創造」の発行に寄せて

馬 場 定 行

急速に進化するAIは，企業活動や社会生活のさまざまな領域へと広がりつつある．生成AIの登場によって，AIは一部の専門家だけが扱う技術から，多くの企業や組織が日常的に利用できる技術へと変わった．しかし今，求められているのはAIを使うことそのものではない．AIをどのように業務やサービスの中に組み込み，新たな価値へと結びつけていくかが問われている．AIの価値は，技術そのものを導入するだけで生まれるものではない．現場に存在する課題を起点に，業務の中で蓄積されるデータ，専門的な知見，そして人の判断を結びつけることで，初めて価値が生まれる．そして，その価値を高め，顧客に届け，さらには社会課題の解決へと広げていくことが，これからのAI活用に求められている．

私たちは，AI活用を単なる技術導入ではなく，現場の課題を起点として価値を創出し，社会へ実装していく継続的な取り組みとして捉えている．

本特集では，2号に渡り，「AIとデータ活用が拓く新たな社会価値創造」をテーマに掲げ，AIを価値へ変えていくための多様な実践を紹介する．また，その過程を理解するための一つの見取り図として，「AIを理解し活用する」「AIをデータとつなぐ」「AIで業務を変える」「AIで価値を創る」「AIを社会へ広げる」という五つの視点を設定した．

本号に収録した論文は，これらの視点を具体的な現場の取り組みとして示すものである．AIを業務やデータと結び付け，新たな活用の可能性を探る取り組みとして，生成AIと業務ソリューションの連携やデータネットワーキングに関する研究を紹介する．これらは，AIを単独で利用するのではなく，業務データや知識資産と結び付けることで，新たな価値創出へつなげようとする実践である．

また，AIの活用領域が広がるほど，安全に利用し続けるための基盤も重要になる．ローカルLLMに関する論考では，機密情報の取扱い，データ主権，運用ガバナンスといった課題を取り上げている．AIを顧客価値や社会価値へ結び付けていくためには，技術の性能だけでなく，信頼性，セキュリティ，ガバナンス，継続的な運用を含めた実装の視点が欠かせない．

さらに，本号では，製造業における予兆検知，農業経営支援，教育分野における学習支援，エネルギー分野における需要予測など，多様な領域での実践を取り上げている．これらは，AIを活用して現場課題の解決に取り組むだけでなく，新たな顧客価値や社会的価値の創出へとつなげようとする試みである．AIとデータの活用が，個別業務の改善から社会全体の価値創造へと発展し得ることを示している．

次号では，AIをより日常の業務へ組み込み，持続的な価値として定着させるための実践を取り上げる予定である．そこでは，人とAIがどのように協働するかに加え，信頼性と安全性の確保，再現可能な仕組みとしての実装，効果の評価と継続的な改善が重要な論点となる．本号と併せて，AIを価値へ変えていく取り組みの広がりと深まりを感じていただきたい．

本特集で取り上げるテーマや技術は多岐にわたる．しかし，それらに共通しているのは，AI を単なる最新技術として取り扱うのではなく，現場の課題解決や価値創造の仕組みの中に位置づけている点である．AI を価値へ変えるには，技術だけではなく，データ，業務，人材，ガバナンスといった要素を一体でとらえる必要がある．そのため，2 号に渡る多様な実践を通じて，そのための知見を共有することを目指している．

BIPROGY グループは，「先見性と洞察力で技術の可能性を見出し，社会課題を解決する」ことを掲げている．AI もまた，導入すること自体が目的ではない．重要なのは，その可能性を見極め，現場の課題，顧客の期待，社会の要請と結びつけ，持続的な価値として実装していくことである．本特集を通じて紹介する各論文は，そのための実装知を，それぞれの領域から示すものである．

本特集が，読者の皆さまと共に，AI とデータをどのように価値へ変え，顧客と社会へ届けていくかを考える契機となれば幸いである．

（執行役員／ CTO）

MartSolutionを拡張する生成AI連携機能のアーキテクチャ

Architecture of Generative AI Integration to Extending MartSolution

黒田武史

要　約　本稿では，BIPROGYが提供する情報系構築支援BIツール「MartSolution」に実装した生成AI連携機能「レポートコンシェルジュ」について解説する．MartSolutionはレポーティングツールを中核とした製品である．レポーティングツールを長期間運用することで，レポート数が増加し情報探索が困難となる状況を解決するため，MartSolutionに対して，レポートを外部情報としたRAGを構築し，自然言語による情報探索を支援する機能を実装した．レポートを介して情報探索することで，レポートに実装された業務ロジックやアクセス権限を活用した，信頼性の高いデータに基づく回答が得られる環境を実現した．

Abstract　This paper presents Report Concierge, a generative AI integration feature implemented in MartSolution, a business intelligence (BI) tool provided by BIPROGY. MartSolution is a BI product centered on reporting functionality. To address the challenge of increasing report volumes and the resulting difficulty in information retrieval that arises from the long-term use of BI, we propose a Retrieval-Augmented Generation (RAG) approach in MartSolution, which treats reports as external information. We also implemented a feature to support information retrieval via natural language. By enabling information retrieval through reports, we have created an environment that enables response generation based on highly reliable data by leveraging the business logic and access control mechanisms embedded in the reports.

1. はじめに

ビジネス環境のデジタル化に伴い，データの活用は企業活動に不可欠な要素となりつつある．データ活用では可視化が重要な要素の一つであり，その実現手段としてBI（Business Intelligence）ツールが広く利用されている．

BIツールには多様な製品が存在し，主たる利用目的からレポーティングツールとセルフサービスBIツールに大別できる．レポーティングツールは，業務要件に応じたレポート（定型帳票，固定帳票とも呼ばれる）をIT部門などのエンジニアが作成し，事業部門のユーザーに提供するといった利用形態が一般的である．ユーザーは作成されたレポートを利用することで，素早くデータを照会することができる．一方，セルフサービスBIツールは，事業部門のユーザーが自らデータを加工し，レポートを作成することを前提としている．ユーザーはレポートの作成に時間を要するが，理想的な形式でデータを照会することができる．生成AI技術の進展を背景として，セルフサービスBIツールにおいても，生成AIを活用したデータ分析への関心が高まっている．

このような状況下において，BIPROGY株式会社（以下，BIPROGY）は，レポーティングツールに生成AIの技術を適用して，新たな価値向上の可能性を検討した．本稿では，BIPROGYが提供するレポーティングツールを中核とした情報系構築支援BIツール「MartSolution[1]」

を対象に，業務課題の解決に向けて実装した生成 AI 連携機能「レポートコンシェルジュ」について，その背景とアーキテクチャを解説する．まず 2 章で MartSolution が提供するデータ可視化のための機能を紹介し，3 章で活用する生成 AI 技術の選定理由について述べた後，4 章でレポートコンシェルジュ機能を説明する．5 章では生成 AI と連携するアーキテクチャ，6 章では効果と精度調整について述べる．

2. MartSolution

MartSolution は BI システムを効率的に構築するためのツール群であり，.NET Framework を基盤としてサーバー上で動作する Web アプリケーションである．MartSolution には，データを可視化するための機能として「定型レポート機能」と「自由検索機能」が用意されている．いずれもレポートとして Web ブラウザにデータを出力する機能である（図 1）．定型レポート機能は，ユーザーが必要な情報をワンクリックで照会できるようにすることを目的とした機能で，あらかじめ定義された情報を基にデータを出力する仕組みである．自由検索機能は，定型レポート機能と比較するとやや検索の自由度が高く，用意されたデータ項目の中からユーザーが任意の項目を選択して出力できる仕組みである．

以下の各節で，定型レポート機能および自由検索機能のアプリケーション構成を説明する．なお，MartSolution にはデータの可視化に限らず様々な機能が備わっているが，本稿ではデータを可視化することを目的とした機能に限定して述べる．

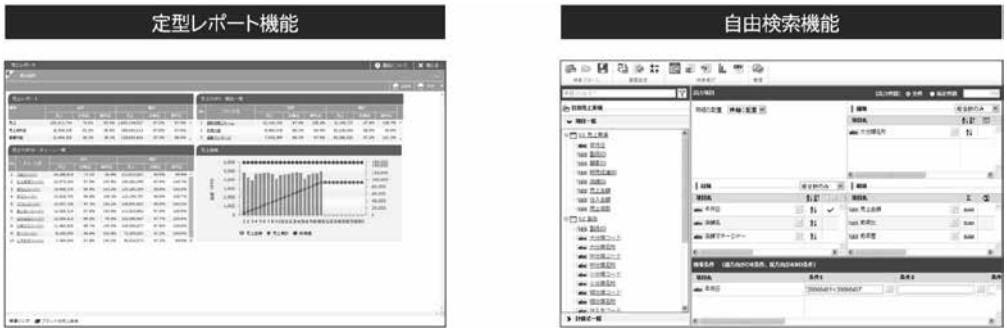

図 1　MartSolution によるデータの可視化

2. 1　定型レポート機能

定型レポート機能は，事前にレポートに出力される情報を「レポートファイル」に定義して利用する．レポートファイルに，データベースからデータを抽出するためのロジックや，出力する項目をプログラムとして実装し，サーバーに配置する．ユーザーが Web ブラウザを介してレポートの画面にアクセスすると，レポートファイルの定義情報に従いデータベースからデータを取得し，レポートとして出力する（図 2）．

なお，レポートファイルにはアクセス権限を設定して，ユーザーの権限に応じた参照可否を制御することができる．

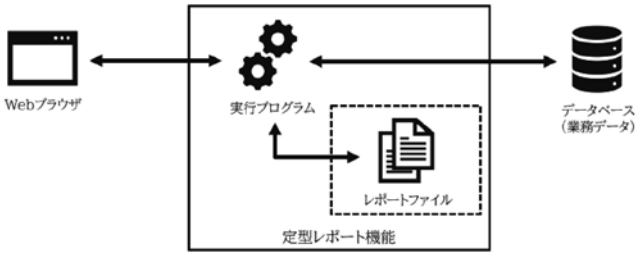

図 2　定型レポート機能のアプリケーション構成

2. 2　自由検索機能

自由検索機能は，レポートに出力される項目をユーザーが指定して利用する．ユーザーは Web ブラウザを介して自由検索機能の画面にアクセスし，レポートに出力する項目を選択する．その定義情報は「リポジトリデータベース」に登録される．その後，ユーザーがレポートの画面にアクセスすると，リポジトリデータベースに登録された定義情報に従いデータベースからデータの取得が行われ，レポートとして出力される（図 3）．

定型レポート機能のレポートファイルと同様に，ユーザーの権限に応じた参照可否を制御することができる．

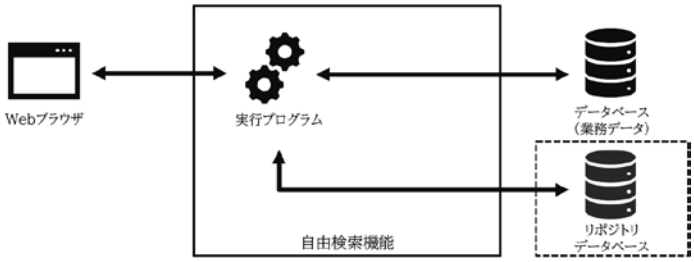

図 3　自由検索機能のアプリケーション構成

3.　レポーティングツールにおける生成 AI の活用

2 章で述べたように，MartSolution はレポーティングツールを中核とした製品であり，業務要件に基づいて定義されたレポートを通じて，正確なデータを素早くユーザーに提供することを目的としている．一方で，レポーティングツールを長期間運用すると，ビジネスの変化に応じてレポート数が数十件から数百件と増加し，情報探索が困難となる状況が散見される．

この問題に対して BIPROGY は，MartSolution に生成 AI の技術を適用して解決することを検討した．その結果，MartSolution を大規模言語モデル（Large Language Models：以下，LLM）と連携し，ユーザーが MartSolution に自然言語で質問すると，検索対象のレポートやデータが回答として得られる機能の提供を目指した．これを実現するには，LLM がレポートやデータに関する情報を参照できる構成にしなければならない．その手法としては，ファインチューニングや RAG（Retrieval-Augmented Generation）が候補となる．

ファインチューニングは，LLM に対して特定の業務やドメインに特化したデータを用いて追加学習し，専門知識を踏まえた回答に対応できるようにする手法である．RAG は，LLM と

外部データの検索結果を組み合わせて回答を生成する手法である．LLM 単体では参照できない企業内のデータや最新情報を反映した回答の生成を実現できる．ファインチューニングと RAG の違いを表 1 に示す．

表 1　ファインチューニングと RAG

観点	ファインチューニング	RAG
知識を与える手段	モデルに対して必要な知識を追加学習させる	必要な知識を外部情報として利用させる
知識を更新する方法	モデルに再学習させる	外部情報を更新する
回答根拠の提示可否	根拠となる知識はモデルに埋め込まれるため困難	根拠となる知識は外部情報のため提示できる

ファインチューニングは専門知識を前提とした回答に優れる一方で，回答に必要な知識を与えるためにはモデルの再学習を要する．MartSolution と LLM の連携において，回答に必要な知識はレポートとデータである．特にデータに関しては日々の業務において頻繁に追加や更新が発生することから，モデルの再学習を伴うファインチューニングは運用面で適合しない．これに対して，質問ごとに外部情報を利用して回答を作成する RAG の手法は，データおよびレポートの追加や更新による影響を受けないことから，本機能の要件に適合していると判断した．また，RAG は外部情報を利用して回答を生成することから，回答の生成に利用したレポートやデータを特定できる．この特性を利用すると，回答と合わせて回答根拠を提示できる．

以上のことから，MartSolution と LLM の連携では RAG の手法を採用して，レポートとデータの探索を支援する機能「レポートコンシェルジュ」を実装した．

4.　レポートコンシェルジュ機能

レポートコンシェルジュのチャット画面（図 4）からユーザーが自然言語で質問すると，レポートコンシェルジュは質問の内容に応じて，MartSolution に格納されたレポートの定義情報を検索し，適切な回答を生成する．質問の内容がレポートに関するものであれば該当レポートへのリンクを提示し，データに関するものであれば該当レポートの検索結果を基にした回答を提示する．

レポートコンシェルジュは回答の生成時に，レポートとデータを外部情報として利用する（図 5）．その際に，データについては常にレポートの検索結果のみを利用し，レポートの参照先であるデータベースから直接データを取得しない仕様とした．これは，信頼性の高いデータを回答に利用することと，データに対する適切なアクセス制御を実現するためである．

図 4　レポートコンシェルジュの画面

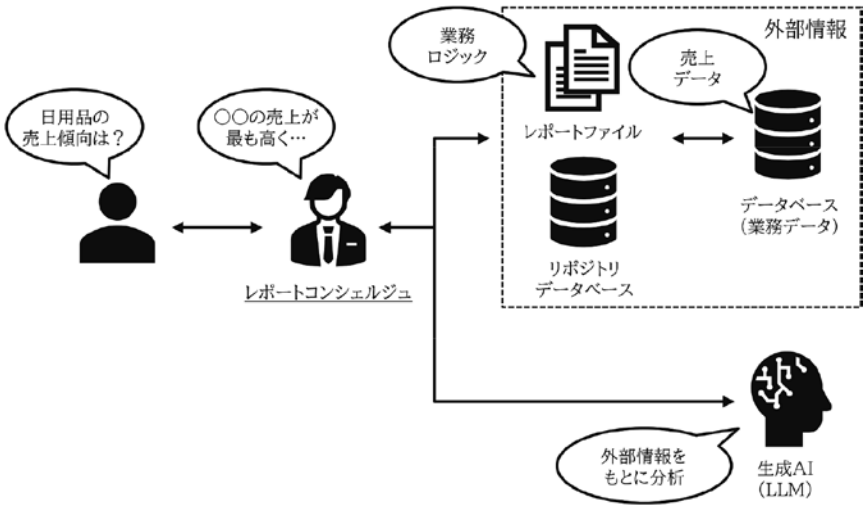

図 5　レポートコンシェルジュの処理の流れ

IT 部門のエンジニアや事業部門のユーザーが事前に作成する，定型レポート機能のレポートファイルや自由検索機能のリポジトリデータベースといったレポートの定義情報には，正しくデータを出力するための業務ロジックが含まれている．データを用いた回答の生成においては，それらの業務ロジックを利用することで信頼性の高いデータが取得できる．

また，データガバナンスの観点からも，データについては常にレポートの検索結果のみを利用している．レポートの定義情報にはユーザーのアクセス権限に関する情報が含まれているため，常にレポートを介してデータを参照することで，該当のユーザーが参照できないデータにアクセスすることを防止している．

5. レポートコンシェルジュのアーキテクチャ

レポートコンシェルジュの実装にあたり，生成 AI との連携の効率化を目的として「Azure OpenAI Service スターターセット Plus[2]開発キット」を採用した．Azure OpenAI Service スターターセット Plus 開発キットは，生成 AI の業務利用を支援するために BIPROGY が提供している環境構築サービスである（図 6）．本サービスには，RAG を実現するための Azure の各種サービスが用意されている．

レポートコンシェルジュを実現するため，MartSolution では主に，回答生成機能，ブリッジ機能，レポート情報クローラーを実装した．以下の各節で，これらの機能の役割を述べる．

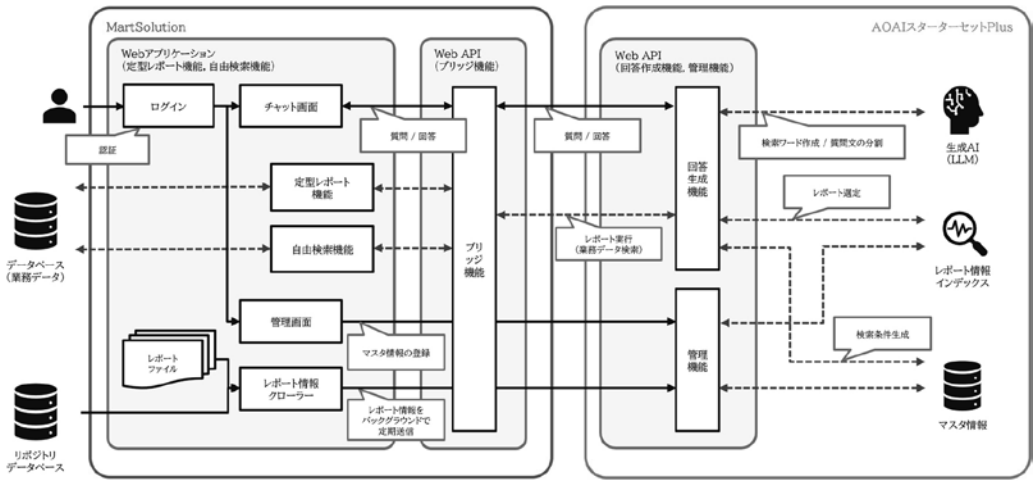

図 6　レポートコンシェルジュのアーキテクチャ

5. 1　回答生成機能

回答生成機能は，ユーザーがチャット画面から入力した質問の内容に応じて適切なリソースを利用し，回答を生成する機能である．回答生成機能はユーザーからの質問を受信すると，質問の内容がレポートに関するものか，データに関するものかを判定する．それぞれの場合に分けて説明する．

5. 1. 1　回答生成機能（レポートに関する質問の場合）

回答生成機能は，チャット画面から質問を受信すると，質問を単語単位に分解し，レポートの検索に利用する検索ワードとして取得する．取得した検索ワードを基に，レポート情報インデックス（レポートに関する情報を格納した検索用のインデックス）からレポートを選定する．さらに，マスタ情報（項目の表示名と変数名のマッピング）を基にレポートに設定すべき検索条件式を作成する．これらの処理により，検索条件がプリセットされたレポートへのリンクを回答として出力する（図 7）．

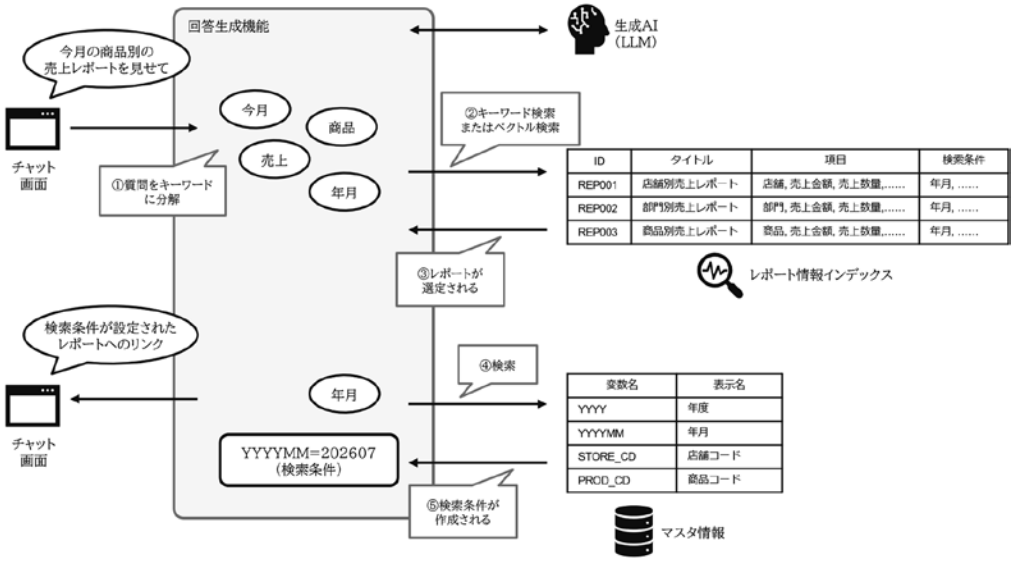

図 7　回答生成機能のイメージ（レポートに関する質問の場合）

レポート情報インデックスとは，定型レポート機能および自由検索機能における，レポートの出力項目や検索条件項目といった基礎情報，並びにレポートに任意で付与できる説明（コンテキスト）を格納したものである．レポート情報インデックスは Azure AI Search によって構築されており，検索ワードを基にキーワード検索またはベクトル検索を行うことで，目的のレポートを選定できる．

マスタ情報とは，レポートの検索条件項目の表示名と，レポートファイル（図 8）やリポジトリデータベースの内部で利用される項目の変数名とのマッピング情報である．レポートファイルやリポジトリデータベースの内部では，項目は変数名で扱われるため，質問の内容からレポートの検索条件式を作成するにはマスタ情報が不可欠である．

```
// タイトルの定義
SetCaption("組織別売上レポート");

// 検索条件項目の定義
ID("ORG_CD", "組織コード", "S");

// 結果表のヘッダー項目の定義
HD("組織名");
HD("売上金額");
・・・

// 検索クエリの定義
rs = OpenRecordSet("SampleDatasource", @"
  SELECT
    ORG_CD AS 組織コード,
    ORG_NM AS 組織名,
    SUM( NT_AMT ) AS 売上金額,
・・・
```

図 8　レポートファイルのイメージ

5. 1. 2　回答生成機能（データに関する質問の場合）

回答生成機能は，レポートに関する質問の場合と同様の手順で，対象となるレポートを選定する．加えて，データに関する質問の場合は，回答の生成においてレポートの検索結果を用いるため，対象のレポートを実行し，その検索結果を基に回答を生成する（図 9）．

質問の内容によっては，一回のレポートの検索結果のみでは回答として十分な情報が揃わないことがある．そのような場合，回答生成機能は継続して他のレポートも実行し，複数のレポートの検索結果を基に回答を生成する．

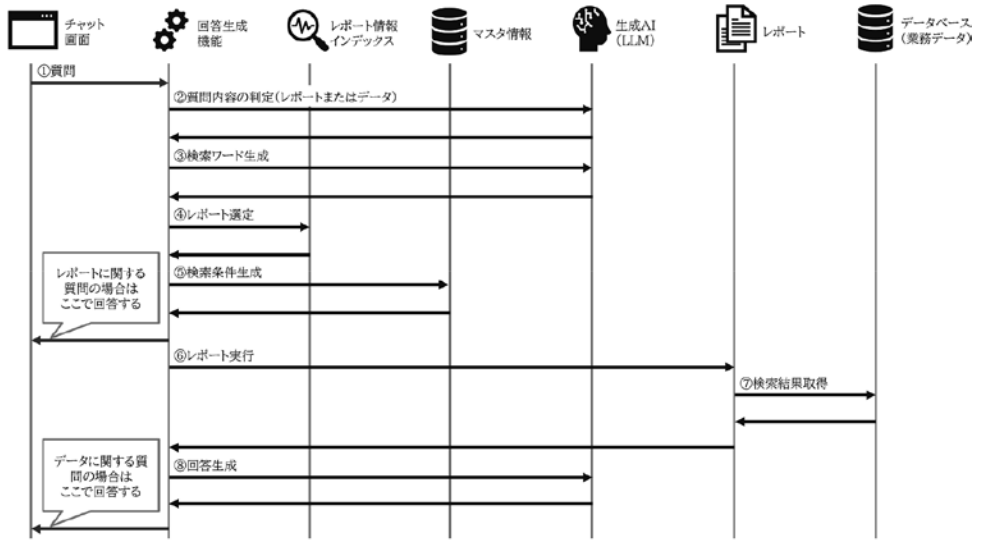

図 9　回答生成機能の処理の流れ

5. 2　ブリッジ機能

ブリッジ機能は，MartSolution の Web アプリケーションと，回答生成機能の Web API を連結する機能である．チャット画面から入力された質問は，MartSolution の Web アプリケーションからブリッジ機能の API に対するリクエストとして送信される．ブリッジ機能はリクエストを受信すると，回答生成機能の Web API を呼び出し，回答を取得する．取得した回答内容は，ブリッジ機能を介して MartSolution の Web アプリケーションに返却される．

質問の内容がデータに関するものである場合，回答生成機能はレポートの実行依頼をブリッジ機能の API に対するリクエストとして送信する．ブリッジ機能はリクエストを受信すると実行依頼をキューに登録する．MartSolution の Web アプリケーションはキューの状態を監視しており，実行依頼が登録されると，レポートの検索処理を実行し，検索結果をブリッジ機能に送信する．送信された検索結果は，ブリッジ機能を介して回答生成機能に返却される．

レポートの定義情報はレポートファイルやリポジトリデータベースに格納されており，それらには業務ロジックが実装されているため，極力外部からのアクセスを避けるべきである．このため，回答生成機能が MartSolution の Web アプリケーションにアクセスすることなく，レポートの検索結果を取得する構成としている．

5.3　レポート情報クローラー

レポート情報インデックスに格納された情報は，RAG における外部情報であるため，質問に対する回答の精度を維持する観点から，適切な更新が不可欠である．レポート情報クローラーは，MartSolution に格納されたレポートの定義情報を監視し，更新を検知した場合はレポート情報インデックスを更新する（図 10）．

定型レポート機能のレポートファイルはプログラムとして実装されており，そのままの状態では定義情報を抽出できない．このため，レポート情報クローラーはプログラムを動的にビルドし，インスタンス化して，定義情報を取得している．

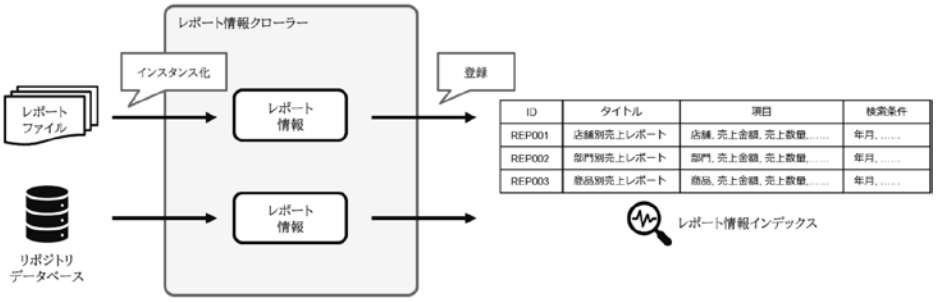

図 10　レポート情報クローラー

6.　レポートコンシェルジュの実装による効果と精度調整

最後に，実装されたレポートコンシェルジュの評価結果について述べる．約 100 件のレポートが格納された，スーパーマーケットの販売管理システムを想定した MartSolution のデモンストレーション環境において，レポートコンシェルジュを利用してレポートやデータに関する質問をした．「今週の X 店の商品売上ランキングを見せて」「今月の Y 店の売上金額の傾向を教えて」といった比較的単純な質問であれば，期待どおりの回答を 30 秒以内には取得できることが確認できた．この結果から，当初想定していたレポート数が増加して情報探索が困難となった状況への対策として，一定の効果が得られたことを確認できた．

しかしながら，MartSolution を利用して作成されるレポートや参照されるデータは，対象のシステムによって様々であり，業種や業界によっても使用される用語やレポートの形式は異なる．したがって，適切な回答を効率的に得るためには，個別環境に応じた精度調整が不可欠である．レポートコンシェルジュには，次のようなチューニングポイントを用意している．

1）回答生成機能

5.1 節で述べたとおり，回答生成機能では質問の内容に応じて適切なリソースを利用している．回答の生成において，LLM に送信するプロンプトがあらかじめ実装されているだけでなく，対象システムの特性に合わせてプロンプトの内容を調整することができる．

2）レポートの説明（コンテキスト）

定型レポート機能および自由検索機能の各レポートには，そのレポートが示す内容を記述するための説明（コンテキスト）機能が用意されている．このコンテキストは，レポート情報クローラーによってレポート情報インデックスに格納され，回答生成機能によるレポートの選定に利用される．期待通りの回答結果が得られない場合は，対象レポートの説明に適切

な内容を記述することで，選定精度の調整に役立てることができる．

7. お わ り に

本稿では，レポーティングツールを中核とした製品である MartSolution を対象に，生成 AI を連携させてレポートおよびデータの探索を支援する機能「レポートコンシェルジュ」の実装における背景とアーキテクチャについて述べた．レポーティングツールの長期的な運用における課題に対し，RAG の手法を適用することで，自然言語で効率的にレポートやデータを探索できる可能性を示した．

RAG を既存のアプリケーションに適用する際には，回答機能の実装に着目しがちであるが，外部情報の収集や更新の仕組みについても同時に検討することが肝要である．

本機能では，レポートを外部情報として利用することで，業務ロジックおよびアクセス権限を維持したまま，信頼性の高いデータに基づく回答生成を実現した．これにより，生成 AI の適用に伴うデータガバナンスの課題にも対応した構成となっている．

今後は，定型レポート機能および自由検索機能のレポートだけではなく，ポータル機能に登録されたレポートやドキュメントなど，MartSolution に格納されたあらゆる情報を対象とし，生成 AI と連携して活用することにより，さらなる利便性の向上を目指す予定である．

最後に，本稿の執筆にあたり，多くの助言をいただいた関係者の皆様に対し，ここに深く感謝の意を表する．

参考文献　［1］ 情報系 構築支援 BI ツール MartSolution，BIPROGY，
https://www.biprogy.com/solution/service/martsolution.html
［2］ Azure OpenAI Service スターターセット Plus，BIPROGY，
https://www.biprogy.com/solution/service/rinzatalkplus.html

※　上記参考文献に示した URL のリンク先は，2026 年 7 月 8 日時点での存在を確認．

執筆者紹介　黒　田　武　史（Takeshi Kuroda）
2001 年日本ユニシス（株）入社．流通系の BI/DWH システム構築に従事した後，現在は BIPROGY が提供する情報系構築支援 BI ツール「MartSolution」の開発責任者として，企画・開発を担当．

グラフ構造を用いた探索プロトタイプの開発

Development of a Graph-based Search Prototype

植　木　達　彦，牧　原　幸　伸

要　約　現代の情報化社会において，蓄積される情報資源の量は増加している一方で，データは組織や個人のもとに分散して保管されており，データのサイロ化が問題となっている．この課題を解決するため断片化した情報を，グラフ構造を用いてネットワーク化することで，網羅的な探索を行うことを目指す．本取り組みではこのアプローチの第一歩として，データをハイパーグラフへと変換し，重み付けランダムウォークを用いて網羅的な探索を行うプロトタイプを構築した．実験として，知的財産分野における特許調査業務をユースケースに選定し，特許のデータをグラフに変換し，探索の精度を評価した．その結果，提案するプロトタイプは参考値としたキーワード検索を上回る精度を達成した．今後は，探索パラメータ調整の自動化や異種データの結合の研究を進める予定である．

Abstract　In today's information society, while the volume of accumulated information resources is burgeoning, data remains dispersed across organizations and individuals, leading to the issue of data siloing. To address this challenge, we aim to facilitate exhaustive information discovery by networking fragmented data through graph-based structures. As an initial step in this approach, we developed a prototype that transforms data into hypergraphs and performs comprehensive searches using weighted random walks. In our experiment, we selected patent search within the field of intellectual property as a primary use case, mapping patent data onto graph structures to evaluate search precision. The results demonstrate that the proposed prototype achieves accuracy superior to that of the keyword search. Future research will focus on the automation of search parameter tuning and the integration of heterogeneous data sources.

1. はじめに

情報化が進展する社会環境において，蓄積される情報資源の量は急速に増加の一途を辿っている．しかしながら，これらの膨大なデータの多くは組織や個人のもとに分断された状態で保管されており，「データのサイロ化」という構造的な問題となっている．結果として，ユーザが検索によって発見し活用できるデータは，特定のプラットフォームやインターネット上に一般公開された極めて限定的な資源に留まっているのが実情である．このように分散し断片化したデータベース群を横断的かつ網羅的に探索し，求める情報を高精度に抽出することは，キーワードマッチングに代表されるような既存の検索手法では困難である．

上記のような課題を根本的に解決するためには，検索の文字列などのユーザのバイアスに過度に依存しない客観的な網羅性を担保しつつ，日々増大するデータを低コストで処理できるスケーラブルな探索技術が求められる．本取り組みでは，この探索を実現するアプローチとして，データを「ノード（頂点）」と「エッジ（枝）」からなるネットワーク構造として表現し，要素間の結びつきの強さを元に解析を行うグラフ構造に着目した．断片化した情報を，グラフ構造

を用いてネットワーク化することで，網羅的な探索が可能となるだけでなく，グラフのトポロジー（接続形態）から，従来の手法では見出すことのできなかった潜在的な関係性や新たな洞察を抽出することが可能となる．

本取り組みでは，上記のアプローチを実現するための第一歩として，単一のデータソースからグラフを構築し探索を行うプロトタイプを開発した．具体的には，企業に広く普及しているリレーショナルデータベース（RDB）形式のデータをグラフ構造に変換し，ノードおよびエッジに独自の重み付けを施したランダムウォーク（Random Walk：RW）ベースのグラフ探索を行うプロトタイプの構築，およびその実証実験の成果について報告する．実証実験においては，本来であれば多様なユースケースに対して適用可能であるが，本取り組みでは，プロトタイプの有効性を定量的に確認するため，知的財産分野における「特許調査業務」をユースケースとして取り上げる．特許のデータをグラフに変換，探索を実施し精度を評価する．

本稿の構成は以下の通りである．2 章では関連研究と RDB のグラフ変換における既存手法の課題，および本研究の前提技術となる重み付けフレームワークについて整理する．3 章では前提技術を含むプロトタイプシステムの構築について詳述する．4 章では実際の特許データを用いた実証実験の設計と結果を示す．5 章では得られた結果に基づく考察と技術的課題を述べ，6 章で本稿のまとめと今後の展望を総括する．

2. 解決のアプローチと前提となる技術

情報資源に対し，グラフを用いて構造化し，そこから適切な関係性を引き出すアプローチには多数のバリエーションが存在する．本取り組みでは，ノードやエッジの属性情報に依存せず，グラフの接続構造そのものをベースに重要度を測る PageRank 中心性を採用する．

また，ベースとなる情報資源は一般的な企業で広く利用されているリレーショナルデータベース（RDB）形式を前提とした．

本章では，前提となる PageRank 中心性について説明し，既存手法の課題，および本取り組みの前提技術となる藤井ら[1]の「重み付けフレームワーク」のモデルについて説明する．

なお本取り組みは，大日本印刷株式会社と慶應義塾大学金子研究室との共同研究に基づくものである．

2.1 PageRank 中心性とランダムウォークによるグラフ解析

PageRank（PR）は，ネットワーク上においてランダムウォーク（確率的な遷移）によって到達しやすい，ハブとして重要なノードやエッジを評価する中心性指標である．この指標はもともとウェブページの相対的な重要度を評価するために考案されたアルゴリズムであるが，現在ではソーシャル・ネットワーキング・サービス（SNS）の解析など，多様なアプリケーションにおいて利用されている．

PR の特徴は，単なるノード同士の接続数（次数）の多寡を評価するだけでなく，ネットワーク全体のグローバルなトポロジーを反映して各要素の重要度を算出できる点にある．これにより，直接的な繋がりを持たないノード間であっても，ネットワークの構造を介した間接的な影響力や関連性を定量的に測ることが可能となる．

さらに，PR を特定の探索に特化させた指標として，Personalized PageRank（PPR）が存在する．通常の PR が「ネットワーク全体における各ノードのグローバルな重要度」を測るの

に対し，PPR は特定の起点ノード（またはノード群）からランダムウォークを開始したと想定し，その「起点ノードに対する相対的な関連性の強さ」を評価する指標である．本研究で対象とする特許探索タスクにおいては，検索クエリとなる特定の特許（起点ノード）に焦点を当て，そこから関連の深い特許をランキング形式で見つけ出す必要があるため，通常の PR ではなく PPR を用いて探索処理を実行する．なお，本稿では特記しない限り，PPR を用いた評価も含めて広義の PR として表記する．

2.2　RDB のグラフ変換手法

RDB のデータをグラフ解析のアルゴリズムに適用するためには，テーブル内に格納されたレコード群とカラムの関係性を，ノードとエッジからなるネットワーク構造に適切に変換する必要がある．代表的な既存の変換手法として，クリーク展開とハイパーグラフ変換の二つが存在するが，それぞれが課題を抱えている．

2.2.1　クリーク展開

クリーク展開は，RDB の各レコードを一つのノードに変換し，同一のカラム値を持つレコード間を全て通常のエッジ（二つのノードのみを結ぶ一対一のエッジ）で相互に結合してグラフを形成するアプローチである．この変換手法は直感的であるが，例えば人事情報のようなデータベースには「性別」や「国籍」，あるいは特許データにおいては「最上位の国際特許分類（IPC）」のように，極端に多くのレコードが該当する巨大なカラムが存在するという特性がある．

この手法でグラフを構築した場合，巨大なカラムに属するレコード群の間に膨大な数のエッジが生成される．その結果，ランダムウォーカーがそれらの巨大なクラスタによる引力に捕らわれやすくなり，PR の確率分布が「サイズの大きいカラム」や「次数の高い有名ノード」に極端に引きずられる現象が発生する．

これにより，ノードの PR 値と次数の順位相関（以降，NPR 相関度）およびエッジの PR 値とエッジサイズの順位相関（以降，EPR 相関度）が過剰に高くなる．結果として，PR が巨大な属性による単純な次数効果以外の微細な特徴をほとんど反映できなくなり，探索のための評価指標としての価値が著しく一面的なものになってしまう．

2.2.2　ハイパーグラフ変換

ハイパーグラフとは，通常のエッジが二つのノード間のみを結ぶのに対し，図 1 のように一つのエッジが二つ以上の複数ノードを同時に包含できる拡張されたグラフ構造である．ハイパーグラフ変換では，同一のカラム値に属するレコード群を一つのハイパーエッジとして定義し，該当する全レコードをその内部のノードとして接続する．

ハイパーグラフ上のランダムウォーク処理では，現在地のノードを含む全エッジ群の中から等確率で一つのエッジを選択したのち，その選択されたエッジに含まれる全ノードの中からさらに等確率で次の遷移先ノードを選択するという二段階のステップを踏む．図 2 の例であれば，ノード a から開始したランダムウォーカーがエッジ b を選択し，さらにエッジ b の中からノード e を選択し遷移している．この手法では，最初のエッジ選択ステップにおいてエッジサイズ（内包するノードの絶対数）の大小による遷移確率への影響が排除されるため，2.2.1 項のク

リーク展開のように巨大なエッジにランダムウォーカーが過剰に吸い込まれる問題を回避できる．

しかし，ハイパーグラフ変換では逆に，NPR 相関度と EPR 相関度が極端に低くなりすぎるという弊害が生じる．ネットワークの解析において「次数が高い（多くの共通属性を共有している，あるいは特許であれば多数の文献から引用されている基本技術である）」という事実は，情報伝播の要となる重要な指標の一つである．しかし，HT 手法では PR アルゴリズムがこの次数の重要性を構造的にほとんど加味できなくなり，やはり探索のための評価指標としての有用性が損なわれてしまう．

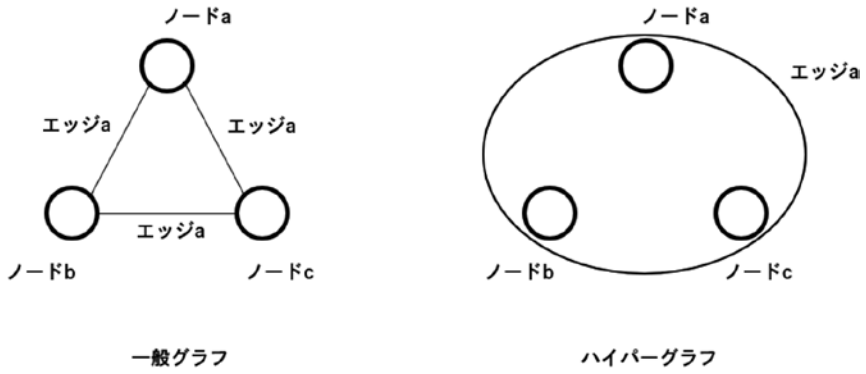

図 1　一般グラフとハイパーグラフ

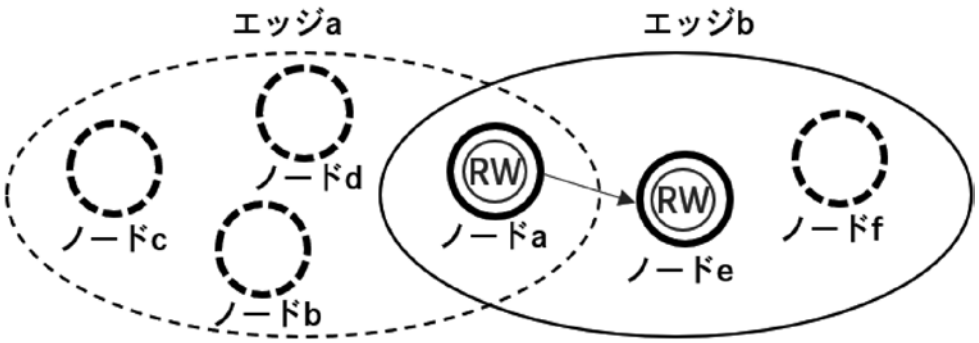

図 2　ハイパーグラフ上のランダムウォーク処理

2. 3　解決すべき技術的課題

既存のグラフ変換手法では，PR 相関度の値が両極端（高すぎる，または低すぎる）となり，ユーザの探索要件に合致した適切な依存度を持つ PR を提供できないことが最大の課題である．特許調査のような複雑な探索タスクにおいては，特許分類などの巨大なクラスタを形成する属性の影響を適度に抑えつつ，特有の専門用語などのニッチな属性が持つ関連性を適切に拾い上げるバランスが求められる．また，基本特許と呼ばれるような次数（被引用数など）の大きな重要なノードの影響力も完全に無視することはできない．

したがって，NPR 相関度と EPR 相関度を自由に調節し，探索の遷移確率をデータセットや探索目的に応じて最適化することが必要となる．

2.4　重み付けフレームワーク

前節で述べた技術的課題を解決するため，本取り組みでは前提技術として，データをハイパーグラフ構造に変換した上でエッジサイズやノードの次数に応じた「重み付け」を動的に行うフレームワークを採用する．

本項では，このフレームワークを説明する

2.4.1　ハイパーグラフ上のランダムウォーク遷移確率

重み付きハイパーグラフ上のランダムウォーカーは，以下の二段階の独立した確率的ステップを実行してネットワークを巡回する．

1）エッジの選択

ランダムウォーカーが現在滞在しているノードに接続されている全エッジの中から，エッジの重みに応じた確率で遷移先のエッジを選択する．エッジ e_i が選択される確率 P_{e_i} は，以下の式(1)で定義される．

$$P_{e_i} = \frac{w_{e_i}}{\sum w_e} \tag{1}$$

ここで，w_{e_i} は該当する遷移先エッジに付与された個別の重みであり，分母の $\sum w_e$ は現在地ノードを含む全エッジの重みの総和である．重みの大きいエッジへの遷移確率が相対的に高くなり，重みの小さいエッジへの遷移確率は低く制御される．

2）ノードの選択

次に，第一段階で選択されたエッジの内部に含まれる全ノードの中から，各ノードの重みに応じた確率で最終的な遷移先のノードを選択する．ノード v_i が選択される確率 P_{v_i} は，以下の式(2)で定義される．

$$P_{v_i} = \frac{w_{v_i}}{\sum w_v} \tag{2}$$

ここで，w_{v_i} は該当ノードに付与された個別の重みであり，分母の $\sum w_v$ は現在地のエッジが包含する全ノードの重みの総和である．

2.4.2　エッジサイズに応じた重み付けモデル

RDB において極端に多くのレコードを保持する巨大なカラム（例：多くの特許を含む技術分類など）への過剰な遷移を抑制し，EPR 相関度を低下させるため，エッジサイズの大小に応じた重み付けを導入する．

エッジ e_k に対して，式(3)のように重み w_{e_k} を付与する．

$$w_{e_k} = \left(S_{e_k} - 1\right)^{1-\alpha} \tag{3}$$

ここで，S_{e_k} はエッジ e_k のエッジサイズ（エッジが包含するノード数）であり，α は $0 \leq \alpha \leq 1$ の範囲を取るエッジ重み付けパラメータである．

図 3 に示すように，$\alpha = 0$ に設定した場合，エッジサイズに比例してランダムウォーカーが吸い込まれるクリーク展開と同様の振る舞いを示す．一方，α の値を徐々に大きくするにつれ

て，包含ノード数が多いサイズの大きいエッジの重みが減衰する．これにより，巨大なエッジによるランダムウォーカーの流量の独占が抑えられ，EPR 相関度を単調に減少させることが可能となる．$\alpha = 1$ の場合は，全てのエッジの重みは 1 となりハイパーグラフ変換と同様の振る舞いとなる．

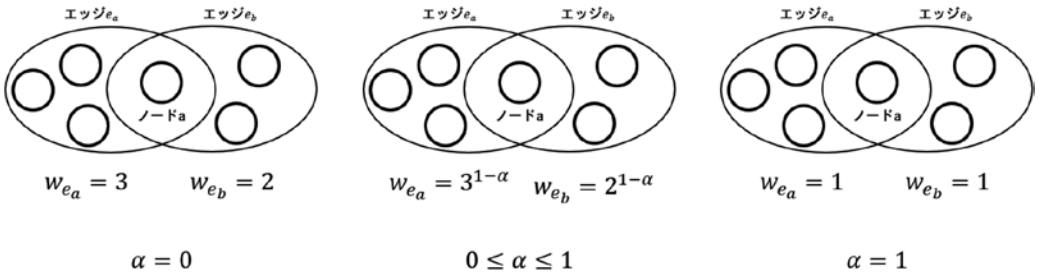

図 3　エッジサイズに応じた重み付け

2. 4. 3　ノードの次数に応じた重み付けモデル

巨大なエッジに属することで結果的に高次数となってしまったノード（例：様々な分類に重複して登録されている汎用的な特許など）への過剰な到達を抑制し，NPR 相関度を低下させるため，ノードの次数の大小に応じた重み付けを導入する．

ノード v_k に対して，式(4)のように重み w_{v_k} を付与する．

$$w_{v_k} = d_{v_k}{}^{-\beta} \tag{4}$$

ここで，d_{v_k} はノード v_k の次数であり，β は $0 \leq \beta \leq 1$ の範囲を取るノード重み付けパラメータである．

図 4 に示すように，$\beta = 0$ に設定した場合，全てのノードの重みは 1 となり，選択されたエッジ内において等確率でノードが選択される．β の値を大きくするにつれて，高次数ノードの重みが相対的に小さくなり，高次数ノードへのランダムウォーカー流量が抑制される．これにより，ノード間の次数による遷移確率の差が平滑化され，NPR 相関度を単調に減少させることが可能となる．

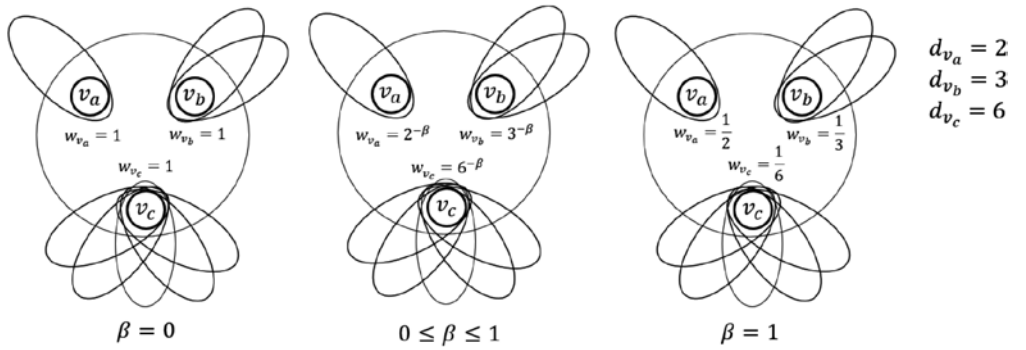

図 4　ノードの次数に応じた重み付け

これら二つのパラメータ（α,β）の強弱をユーザの探索要件に合わせてチューニングすることで，両極端な既存手法（クリーク展開とハイパーグラフ変換）の間を結び，最適な相関度を持つ PR を出力することが可能となる．

3. プロトタイプの構築

2 章の重み付けフレームワークを前提として採用し，RDB に格納されたデータをハイパーグラフに変換し，探索を実行する「特許探索プロトタイプ」を構築した．本章では，その構成について説明する．

3.1 プロトタイプの全体アーキテクチャ

開発したプロトタイプは以下の四つのフェーズで構成される（図 5）．

1）グラフ構築

指定したデータソースからデータを読み込み，ノードとするデータの基本単位と，ハイパーエッジとして結びつける共通属性を定義してグラフ構造を構築する．本システムでは，自然言語で記述されたテキストを形態素解析したものと，分類番号などを同一のグラフに統合する．

2）探索指示

ユーザが探索の起点となる単一ないし複数のノード（基準となる手持ちの特許など）をシステムに入力する．従来の検索エンジンのように特定のキーワード群を入力する必要はなく，ノードそのものを探索クエリとして扱う．ここでランダムウォークの試行回数や終了確率，前提技術で定義したエッジ・ノード重み付けパラメータ（α,β）を指定する．

3）グラフ探索処理

指定された起点ノードから，式(1)および式(2)の遷移確率に基づくランダムウォーカーがネットワーク全体を巡回し，各ノードの PPR 値を計算する．

4）結果提示

計算された PPR 値に基づき，起点ノードから関連性が高いと推定されるノード群をランキング形式でリスト化し，提示する．

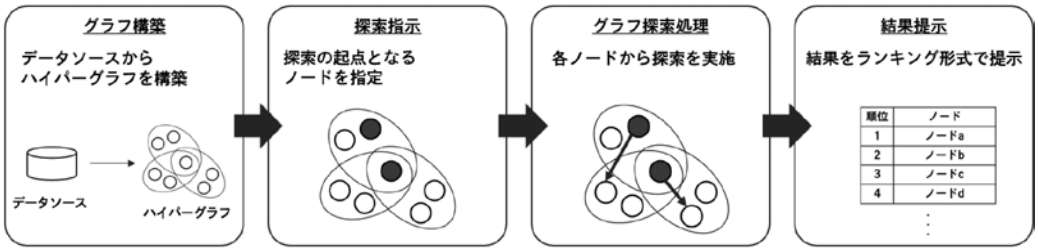

図 5　プロトタイプの全体アーキテクチャ

4. 実験

構築したプロトタイプの探索精度と有効性を定量的に評価するため，知的財産分野における特許先行技術調査業務を取り上げて実証実験を行った．

4.1　ユースケース選定の背景と目的

本プロトタイプで実装したグラフベースの探索技術は，テキストや数値ベースであれば特定のデータソースやフォーマットに特化したものではなく，様々な用途に対して汎用的に利用可能である．しかし，出力されるランキング結果は「ノード間の関係性の強さ（PPR 値）」という指標であるため，一般的なデータセットにおいては「何が正しい探索結果であるか」を客観的かつ定量的に評価（正解データを定義）することが非常に難しいという課題があった．

そこで，定量的な評価が可能なユースケースとして特許調査業務を選定した．通常，ユースケースとして商品レコメンドなどの個人によって嗜好が異なる（正解のない）ものが考えられるが，定量的な評価を行うことが難しい．特許調査においては，特定の特許からの関係性の有無を，特許庁の審査官の判断を正解として扱うことで定量的な評価を行った．

例えば企業が新たな技術について特許出願を行う際，事前に特許庁の審査官と同様の観点から先行技術調査（サーベイ）を実施することが不可欠である．出願審査において，審査官から既存特許との抵触を指摘されて拒絶理由通知書が発行され，最終的にそのまま拒絶査定となった場合，出願までに要した多大な時間と資金が浪費される結果となる．

特許探索においては，技術分野が少しでも異なれば同一の概念や技術的課題に対しても全く異なる語彙が用いられることが多く，申請者が自身の専門外の技術領域にまで精度の高い調査を及ぼすことは単純なキーワード検索では極めて困難である．本実験では，ある特許に対し審査官が引用した「引用文献」を正解ペア，つまりクエリ特許に対して関連性があると審査官に判断された特許と定義する．当プロトタイプが起点特許から探索を開始し，提示したランキングリストの上位に，この引用文献がどれだけの確率で含まれているかを測定することで，システムの探索精度を評価する．

4.2　データセットの諸条件

特許データベースより，以下の条件にて特許データを抽出し実験用データセットを構築した．なお，データセットの諸条件については，異なる手法であるが類似特許検索の研究である中山[2]らの研究条件を参考にしている．

特許データベースより，以下の条件にて特許データを取得した．

1）国際特許分類（IPC）：D（繊維・紙）
2）対象期間：全期間
3）データクレンジング：「要約」項の存在しない特許レコードを除外し，最終的な全ノード件数を 164,075 件とした．

取得した特許データの項目は以下の通りである．

1）発明の名称
2）要約
3）請求項
4）IPC
5）F ターム
6）出願人識別番号

7）代理人識別番号
8）引用文献情報

IPC（国際特許分類）とは，世界各国が共通に使用できる特許分類である．

F タームとは，特許審査のための先行技術調査を迅速に行うために開発された検索インデックスのことであり，特許庁が開発している．

出願人識別番号は出願人本人が付与される 9 桁の番号であり，代理人識別番号は弁理士などの代理人が付与される 9 桁の番号である．

中山[2]らが用いているデータ項目に対し，本取り組みでは請求項や出願人識別番号，代理人識別番号などを追加で用いている．

4. 3　ハイパーグラフの構築

グラフ構築においては，ノードとエッジを定義する必要がある．今回の実験では，ノードは特許 1 レコードとした．エッジに関しては図 6 のように二通りのエッジ生成処理を行い，同一のハイパーグラフへと統合した．

1）自然言語で記載されたテキスト項目に関する処理

特許レコードから「発明の名称」「要約」および「請求項」の自然言語テキストを抽出し，形態素解析を用いて解析，「名詞」のみを抽出する．抽出された各名詞に対し，その名詞が出現した特許レコード同士をハイパーエッジで接続する．図 6 の例であれば，特許 A の「要約」から複数の名詞を抽出し，同じ名詞を有する他の特許とハイパーエッジで接続している．

2）分類番号に関する処理

「IPC（国際特許分類）」「F ターム」「引用文献情報」「出願人識別番号」「代理人識別番号」といった項目は，そのまま属性値として扱う．名詞と同様に，各属性値がどの特許レコードに出現したかを検索し，共有するレコード同士をハイパーエッジで構築する．また，「引用関係」については引用及び被引用特許を含んだハイパーエッジとして構築する．図 6 の例であれば，特許 A と同じ IPC に属する他の特許とハイパーエッジで接続している．

これにより，テキスト内の共通単語から生じるエッジと，分類番号から生じるエッジが同一のハイパーエッジとしてグラフ内に存在することとなる．

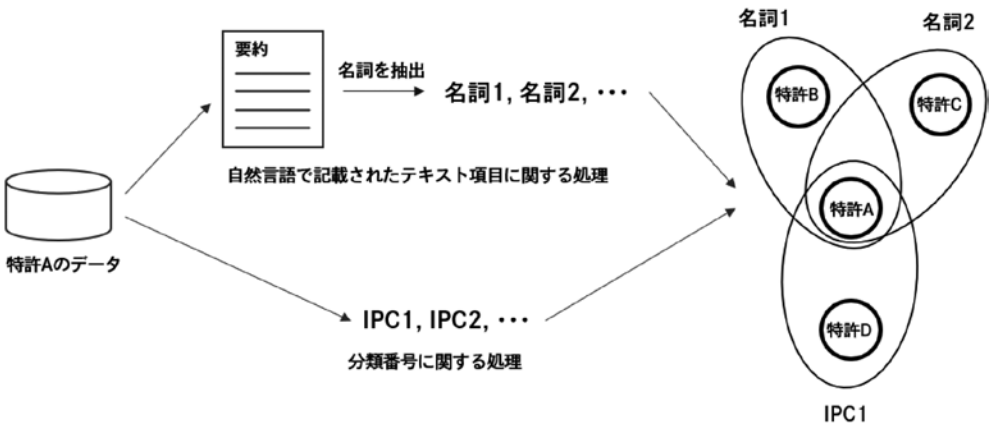

図 6　特許データからハイパーグラフの構築

また事前の検証において，上記で構築手順したグラフは重み付けが無視されるほどにノードに対する次数が過度に多くなってしまい（遷移確率がほぼ同一となり，ランダムに遷移先を選ぶことと同じとなってしまい）結果が収束せず精度が出ない，という課題が発生した．例えば一つのノードを含むエッジが 1,000,000 個存在する場合は，各エッジへの遷移確率が $\alpha = 1$ において 0.0001％となり，α を 1 以外に調整し重み付けを行ったとしても，ランダムにエッジを選び遷移することとほぼ同じとなり，収束しない結果となる．よって今回のグラフについては，エッジの大きさに上限値を設け，上限値を超過した大きさのエッジを削除した．

4. 4　実験条件と評価結果

精度評価の指標として，中山[2]らの先行研究を参考に Recall@k を採用した．Recall@k は，クエリとなる特許に対してシステムが予測出力した上位 k 個のリストの中に，実際の正解特許（審査官が提示した引用文献）が含まれている割合を算出し，最終的に全クエリにおける平均を取った指標である．特許調査においては，適合率（Precision）よりも見落としを防ぐ網羅率（Recall）が重要と考え，本取り組みではこの指標を採用している．

テスト手法として，データセット全体（164,075 件）のうち，時系列的に最も新しい 10％にあたる特許群を探索の起点（クエリ）とし，残りの過去 90％のデータ群を探索対象の母集団とした．クエリとなる 10％のデータ群については「引用関係」のハイパーエッジを削除した状態で実験を行った．探索実行時には，ランダムウォークの試行回数，終了確率，重み付けパラメータ（α, β）について，精度が最大化されるようチューニングを行った．

表 1 に提案手法の精度を記載する．（引用関係有り）と記載した手法は，4. 2 節で述べたデータ項目全てを用いてハイパーグラフを構築した場合の結果である．（引用関係無し）と記載した手法は，4. 2 節で述べたデータ項目から「引用関係」を除外してハイパーグラフを構築した際の結果である．

完全に同一のデータセットではなく用いているデータ項目も異なるため，あくまで参考値ではあるが，中山ら[2]の実験結果よりキーワードによる全文検索の精度と中山ら[2]の提案手法である機械学習ベースの類似特許検索手法による精度も併記する．

表 1　各手法の精度

No.	探索手法	k＝50	k＝100
1	本取り組みの提案手法（引用関係有り）	37.5%	48.3%
2	本取り組みの提案手法（引用関係無し）	36.8%	47.8%
3	キーワードによる全文検索	25.2%	31.0%
4	中山ら[2]の提案手法	37.1%	48.1%

注記：精度算出に用いたデータセット等が完全に同一でないため No. 3 と No. 4 は参考値

ハイパーグラフを用いた提案手法は，k＝50,100 のどちらにおいてもキーワードによる全文検索の精度を大幅に上回った．

提案手法のうち引用関係有りの結果は，中山ら[2]の提案手法と同程度の水準となった．

引用関係無しの結果においても，引用関係有りの結果や中山ら[2]の結果から大きくは劣らな

い結果となった．これは，「引用関係」のような正解のデータが無い状態であっても，ある一定の精度で正解を推測することが可能であることを示している．また，正解の存在しないような課題に対して適用した場合においても，ある程度もっともらしい結果を提示することができると考えられる．

5. 考察

得られた結果に基づき，一定の精度が得られた要因を考察する．

5.1 グラフベース探索の精度の要因について

表1に示された通り，キーワード全文検索の精度（Recall@100 で 31.0%）は他の手法に比べて低い．これは，特許文書特有の「言い換え」や「専門用語のブレ」など，技術領域ごとの専門用語を単純な文字列一致では吸収できないためと推測される．

一方，提案手法が高い精度を記録した最大の要因は「ハイパーグラフによる多面的なトポロジー」にあると考えられる．特許間の関連性は，単に明細書のテキストが似ているかという一側面的な要素だけでなく，「特定のニッチなFターム群を共有している」「同じ代理人が担当しており，記述が似ている」「同じ出願人が過去に出した一連の技術ポートフォリオに含まれる」といった，複数のメタデータの交差によって特徴づけられる．

提案手法は，名詞の共有というテキスト的側面に加え，IPC やFターム，出願人といった多様な文脈を，「ハイパーエッジ」として単一のグラフ空間に投影している．その上でランダムウォークを実行することで，熟練した審査官が経験的に行うような「複数の要素を経由して目的の特許に至る」といった連想的な推論プロセス（マルチホップな探索）を，網羅的に再現できたのではないかと考えられる．

5.2 パラメータ（α, β）が探索の精度に与えた影響

特許データベースにおいては，「G06F（計算処理）」のような極めて汎用的で範囲の広いIPC や，明細書中に頻出する「装置」「方法」といった一般名詞が無数に存在する．クリーク展開のようにこれらを単純なエッジで結合してしまうと，これらの巨大な属性が過度にランダムウォーカーを吸い込み，結果として次数に基づく一面的な結果になってしまう．

本プロトタイプで導入したエッジ重み付け（α）は，式(3)によってこのような「大きすぎる分類コード」や「汎用すぎる単語」が持つエッジの引力を数学的に減衰させる機能を果たしたと推測される．これにより，ノイズとなる汎用的な繋がりが排除され，マイナーではあるがクエリと確かな共通項（特定のニッチなFタームや専門用語の組み合わせなど）を持つ，真に関係性のある先行特許へとランダムウォーカーを効率的に到達させることに寄与したと考えられる．

5.3 プロトタイプの残課題と将来展望

本プロトタイプについて，技術的な残課題が存在する．

第一に，重み付けパラメータ（α, β）のチューニングである．例えば，EC サイトの商品レコメンドで用いる場合であれば，個人毎に探索の趣味嗜好が異なるため，結果を参照しつつ自身の好みのパラメータに調整すればよい．このような場合であればユーザにパラメータ調整を

委ねることが可能であるが，本実験のような絶対的な正解が存在し精度を求められるケースにおいては，精度が高まるようパラメータのチューニングを行う必要がある．クエリの性質，目的や対象データセットに応じて最適な（α,β）の値は動的に変動するはずであり，なんらかのチューニングのアルゴリズムが必要である．例えば，事前のデータ分布やネットワークの次数分布から，これらのパラメータを自動的に推定・調整を行うアルゴリズムが考えられる．

第二に，異種データベースの接続である．今回は特許データベースという複数の項目はあるが単一のデータベース内で探索を完結させているが，実際の調査業務においては，特許文献だけでなく，学術論文データベースなど，異なる外部の情報資源と横断的に探索を行うことが求められる．異なるデータベースから生成されたグラフ同士を「共通」もしくは「近しい」ノードやエッジを介して自動的に接続する方式（グラフ間の自動接続）を確立する必要がある．

6. お わ り に

本稿では，企業や社会全体に散在しサイロ化が進む情報資源を統合し，網羅的かつスケーラブルな情報探索を実現するためのアプローチの第一歩として，データソースからグラフを構築，重み付けされたランダムウォークの探索フレームワークを提案し，プロトタイプを構築した．知的財産分野における特許調査業務を対象とした実験を通じて，本手法が一定の探索精度を達成できることが分かった．

今後は，本実証実験の考察で明らかになったパラメータチューニングの自動化や，学術論文等を含めた異種データベースとのグラフ接続方式の確立に向けて研究を進展させる予定である．

謝辞

本取り組みは，大日本印刷株式会社と慶應義塾大学金子研究室との共同研究に基づくものである．

参考文献

［1］ 藤井洸太朗，山下剛志，金子晋丈，リレーショナルデータベースのグラフ利用のための重み付けフレームワーク，信学技報，電子情報通信学会，vol. 123，no. 342，2024年1月，https://ken.ieice.org/ken/paper/20240119wcar/

［2］ 中山樹，菊地良将，中西宏和，佐々木勇和，荒瀬由紀，鬼塚真，グラフ深層学習を用いた類似特許検索の精度向上，第16回データ工学と情報マネジメントに関するフォーラム，2024年1月，https://pub-files.atlas.jp/fs/public/deim2024/ver_4/abstract/ja/T2-B-4-02.pdf

※ 上記参考文献に示したURLのリンク先は，2026年4月1日時点での存在を確認．

執筆者紹介 植 木 達 彦（Tatsuhiko Ueki）

2014年大日本印刷(株)入社．インフラエンジニアとしてサーバやネットワークの設計，構築，保守，運用までを担当した後，2023年より先進技術のリサーチ業務，特に企業が保有するデータの利活用についてのリサーチに従事．

牧　原　幸　伸（Yukinobu Makihara）
2011年大日本印刷(株)入社．プリント基板および電装システムの設計開発業務に従事．リハビリテーション用歩行アシストロボットの研究開発を担当した後，ヘルスケア領域のデータ分析技術の研究開発を経て，2024年より先進技術のリサーチ業務に従事．博士（工学）．

到来するデータ主権に応えるLazarus AIが示すローカルLLMの可能性

How Lazarus AI Illustrates the Potential of Local LLMs in Response to the Rise of Data Sovereignty

青木岐夫

要　約　生成AIの普及に伴い，シャドーAIや自律的AIエージェントによる新たなセキュリティリスクが顕在化する中で，データを外部に送信せず企業の管理下で運用できるローカルLLMへの関心が高まっている．国内事例が示すように，ローカルLLMは実用段階にあり，継続的な運用改善を前提に，PoCから段階的に導入することが肝要である．ユニアデックスでは，抽出や精査に特化したLazarus AIを，機密性と検証可能性を担うハイブリッド運用の中核として位置づけている．

Abstract　As the adoption of generative AI brings emerging security risks like “shadow AI” and autonomous AI agents to the forefront, interest is growing in local LLMs that can operate under corporate control without transmitting data externally. Domestic use cases demonstrate that local LLMs have reached a practical stage of maturity. Therefore, a phased implementation starting with a Proof of Concept (PoC) and building upon a foundation of continuous operational improvement is essential. Uniadex positions Lazarus AI, a tool specialized in data extraction and scrutiny, as the core component of a hybrid operational model designed to ensure confidentiality and verifiability.

1. はじめに

2026年現在，企業への生成AI導入はすでに「始めるかどうか」の議論ではない．総務省の「令和7年版 情報通信白書」によれば，何らかの業務で生成AIを利用している日本企業は55.2%に達している[1]．文章作成，メール対応，業務フローの自動化にLLMが組み込まれる状況は，大企業だけでなく中堅・中小企業にも広がっている．

その一方で，普及の裏側では見えにくいひずみが蓄積している．CyeraとCybersecurity Insidersによる「2025 State of AI Data Security」調査[2]では，回答者の76%が自律的に行動するAIエージェントはセキュリティ確保が最も難しいと認識しており，70%が公開LLMへの外部プロンプトを高リスクと捉えている．さらに，同レポートは，調査対象企業の83%がAIを業務利用している一方，利用実態を可視化できている企業は13%に留まることも示している[2]．

これらは欧米企業を対象とした調査ではあるが，入力データが海外サーバーで処理されうるクラウドAIの構造は日本企業にも共通する．そのため，特に金融，医療，防衛などの規制産業では，機密データを組織の管理下に保持したまま利用できるAI基盤として，オンプレミス[*1]のデータセンターやプライベートクラウド[*2]への関心が高まっている．背景にあるのは，機密データをどこまで自社の管理下に置くかというデータ主権[*3]と，それに適したAI基盤を

どう整備するかという課題である．

本稿では，こうした背景を踏まえ，AIエージェント[*4]時代に顕在化したリスクとローカルLLMへの関心が高まる要因を整理し，実現アプローチの違いを比較する．その上で，ローカルLLMの中でも「生成」ではなく「抽出・精査」に特化したLazarus AIに着目し，同製品の技術的な特徴を紐解きながら，企業の機密データを安全に活用し，日々の意思決定を支える新たな基盤としてどのような価値をもたらすのかを検討していく．まず2章で生成AIの普及により顕在化したセキュリティリスクについて触れた後，3章でローカルLLMの必要性，4章でローカルLLMの実現アプローチを述べる．5章ではローカルLLMの段階的導入の重要性，6章ではコスト構造を説明する．7章でLazarus AIの技術的な特徴，8章でクラウドAIと併用するハイブリッド運用について述べる．

2.　AIエージェント時代が生み出した新たなセキュリティリスク

本章では，生成AIの普及に伴って顕在化したリスクを，シャドーAI，AIエージェントの自律性，そして禁止策の限界という三つの観点から整理する．

2.1 「シャドーAI」という構造的問題

シャドーAIとは，IT部門や経営層が関知・承認していないにもかかわらず，従業員が業務で独自に使用しているAIサービスを指す．かつての「シャドーIT」の現代版と言えるが，生成AIには入力データがモデルの学習に再利用され，外部へ流出するという特有のリスクが存在する．

独立行政法人情報処理推進機構（IPA）[3]や総務省[4]が警鐘を鳴らすように，従業員は利便性を優先し，組織の規定を逸脱して未承認ツールを使い続ける傾向にある．また，Gartner® も四半期ごとのEmerging Riskレポートで，シャドーAIがリスクマネジメント責任者にとって最重要関心事の一つであると述べている[5]．具体的には，企画書作成のために売上データをAIに貼り付けたり，契約書のレビューを外部サービスに依頼して機密情報を入力したりするケースが日常的に発生している．

Oktaが2026年2月に発表した調査[6]では，ITリーダーの86%がAIエージェントを「ミッションクリティカル」または「非常に重要」と位置付ける一方，データ漏洩・流出（83%）や過剰な権限付与（80%）を導入上の主要な懸念として挙げており[6]，AIの「必要性」と「危険性」のジレンマが浮き彫りになっている．「生成AIは危険だから全面禁止」という対応は，逆効果になる．NTTドコモビジネスが公開する資料[7]が端的に指摘するように，IT部門が承認するツールがなければ，従業員は個人的なアカウントで個人スマートフォン経由のAI利用を続けるだけである．禁止措置は見かけ上のリスク低減に過ぎず，実態としてはリスクの「不可視化」を招く．

2.2　AIエージェントの自律的行動がもたらすリスク

問題をさらに複雑にしているのが，AIエージェントの台頭である．単に質問に答えるチャットボットではなく，カレンダーを操作し，メールを送信し，外部APIを呼び出し，データベースにアクセスするような「行動する存在」としてのAIエージェントが，企業の業務フローに組み込まれ始めている．

トレンドマイクロが発表した「AIエージェントと脆弱性 PART 1[8]」は，この状況の危うさを具体的に示す．AIエージェントが外部のWebサイトを参照したり，外部ツールと連携したりするプロセスの中で，間接的なプロンプトインジェクション攻撃*5のリスクがある．悪意を持って設計されたWebページやドキュメントに隠された命令を読み込んだAIエージェントが，ユーザーの意図とは無関係に機密情報を外部へ送信してしまうシナリオは，もはや理論上のリスクではない．

より根本的な問題として，OWASPが公開する「2025年版LLMアプリケーションのトップ10リスク[9]」には，プロンプトインジェクション，機密情報の漏洩，サプライチェーンの脆弱性，データポイズニング*6が上位にランクされている．これらのリスクは，LLMが外部（クラウド）に存在することによって根本的に対処しにくくなる構造的な問題を抱えている．

2.3 「全面禁止」という解決策の限界

2.1節で述べた通り，「生成AIを禁止すればよい」という発想は逆効果になる．しかし，このAIの「必要性」と「危険性」のジレンマには別の解決策がある．それは，従業員が安全にAIを利用できる環境を企業自身が提供することである．すなわち，LLM基盤そのものを自社の管理下に置くというアプローチである（図1下「整備アプローチ」）．これにより，シャドーAIの動機を抑制し，従業員が承認済みのAIツールを利用しやすくなる．

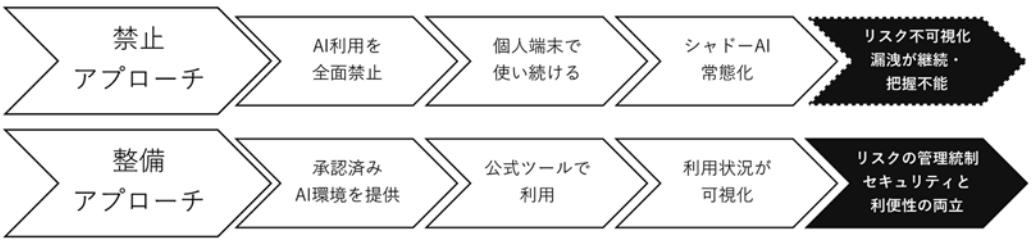

図1　生成AIの禁止措置と利用環境整備の構造的な相違
（参考文献[3][4][7]を踏まえて著者が作成）

3.　LLMの定義と普及加速の背景

本章では，ローカルLLM（Local Large Language Model）の定義と特徴，普及が加速した背景を述べる．

3.1　ローカルLLMとは何か

ローカルLLM（Local Large Language Model）とは，クラウド上の外部サーバーに常時依存せず，オンプレミス環境やプライベートクラウドなど企業の管理下で運用する大規模言語モデルを指す．NTTPCが公開する資料[10]では，クラウドLLMとの違いを「データを外部に送信することなく，マシン内で処理を完結できるため，機密情報の漏えいリスクを大幅に抑えられる」と整理している．重要なのは，適切な構成を取ることで，外部送信経路を限定または遮断しやすい点にある．

企業がChatGPT EnterpriseやGeminiなどのパブリックAPI*7を利用する場合，本質的には外部事業者が運用するAI基盤を利用していることになる．データプライバシーに関する契

約上の保証があったとしても，アーキテクチャ上，データが企業境界外の基盤を通過する構成は残る．共有インフラにおけるモデル反転攻撃（Model Inversion Attacks：攻撃者がモデルにクエリを投げて訓練データを抽出する手法）や，プロンプトインジェクションによるコンテキスト漏洩の理論的リスクは完全には消えない．これに対して，オンプレミス展開や閉域環境では，AIの知能そのものである「モデルの重み（学習データから得たパラメータ）」，利用者が入力した「推論ログ（対話履歴）」，そして社内資料などをAIが読み取れる形に変換した「ベクトル埋め込み（データの特徴量）」のすべてを，自組織の物理的な管理下に置くことができる（図2）．

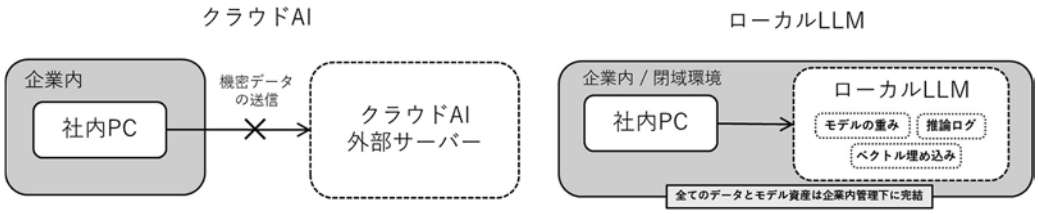

図2　クラウドAIとローカルLLMにおけるデータ処理境界の違い

3. 2　2025～2026年に普及が加速した三つの背景

ローカルLLMの普及が加速した背景は三つある．第一の背景は，モデルの軽量化・高性能化と，それを支えるハードウエア要件の低下である．2025年にはMetaのLlama系列，Alibabaの Qwen，MistralAIの各モデルが性能向上を遂げ，OpenAIもオープンソースモデル「gpt-oss」をリリースした．NTTが2025年に発表した国産軽量モデル「tsuzumi 2[11]」は，単一のGPUで動作しながら金融・医療・公共分野の知識を強化したモデルとして紹介されている．さらに，4ビットおよび8ビット量子化[*8]（GGUF，AWQフォーマット[*9]）の普及により，従来よりも少ないGPU枚数で運用できる構成が現実的になってきた．加えて，特定のGPUアーキテクチャに縛られない推論環境の多様化が進んだこと[12]により，ハードウエアの選択肢は大きく広がっている．

第二の背景は，導入ツールのGUI化・民主化である．かつてはコマンドラインに精通した一部のエンジニアに限られていたローカルLLMの運用も，2025年以降は「Ollama」や「LM Studio」といったツールの普及により，IT担当者でも試しやすい環境が整ってきた．これらのツールは，量子化モデルの変換や，Apple Siliconのユニファイドメモリ，AMD製アクセラレータなど，多様なハードウエア環境における差異の吸収にも寄与している[12][13][14]．

第三の背景は，規制環境の変化とデータ主権への関心の高まりである．欧州では「EU AI Act[*10]」の段階的な施行が進み，GDPR[*11]との整合性を確保する観点から，データがどの国・どのサーバーで処理されるかが経営レベルの論点になっている．日本でも個人情報保護や経済安全保障の観点から，海外ハイパースケーラーへの依存を見直す動きがあり，国内データセンターや顧客専有環境への関心を高める要因となっている．

4.　自社専用LLM環境（ローカルLLM）を実現する三つのアプローチの比較

企業がChatGPT等のパブリックAPIを避け，自社の統制下でLLMを運用する方法は一様

ではない．本稿では，他テナントから論理的または物理的に隔離された「自社専用のLLM環境」を広義のローカルLLMと定義し，大きく三つのアプローチに整理する．それぞれのアプローチには明確なトレードオフが存在し，どれが優れているかは組織の要件によって異なる．

4.1　アプローチA：OSS[*12]モデルの自社構築（DIY型）

LlamaやQwen，Mistralなどのオープンソースモデルを自社のGPUサーバー（オンプレミスまたはIaaS）にデプロイし，RAG[*13]（検索拡張生成）やファインチューニング[*14]で業務に最適化するアプローチである．米国のテクノロジー企業や金融機関では，このDIYアプローチが主流といわれており，モデルの重みへの直接的なアクセスを活用して差別化されたプロダクトを自社開発している．標準的な技術スタックは，Kubernetes上で稼働するLlama 3等のモデル，オーケストレーションを行うvLLMやTGI（Text Generation Inference），そしてアプリケーションロジックを担うLangChainやDify（ローコードプラットフォーム）などで構成される．

このアプローチの強みは，ライセンス費用が無料であること，自社のユースケースに合わせた深いカスタマイズができること，技術的な資産として社内に蓄積されることである．過去20年の判例ファイルで訓練された法務特化モデルや，社内のプライベートリポジトリへの完全なアクセス権を持つコーディングエージェントなど，パブリッククラウドでは実現困難なユースケースが開ける．

弱みは，高度なAI/MLエンジニア（AIや機械学習のエンジニア）による継続的な運用体制が不可欠であり，初期のモデル選定からRAGパイプラインの構築，ハルシネーション[*15]対策まで，品質を業務水準に引き上げるための工数と失敗コストが膨大になりやすいことである．近年はNVIDIA NIM（NVIDIA Inference Microservices）[15]のような推論最適化コンテナの登場でデプロイの障壁は下がりつつあるが，依然として運用難易度は高い．さらに，トレンドマイクロのレポート[8]が警告するように，公開リポジトリから取得したモデルファイル自体にマルウエア[*16]が仕込まれるリスクも現実に存在する．PoC[*17]に際してはまず7B～8Bパラメータの量子化モデルとRAGの組み合わせから始め，小さな業務課題で実証実験を行い，効果が出たら大規模モデルへ拡張する段階的アプローチが現実的である．

4.2　アプローチB：メガクラウドのプライベート環境（VPC型）

AWSやAzure，GCPなどのメガクラウドが提供する仮想プライベートクラウド（VPC）内にLLMを展開するアプローチである．「クラウド上にあるが，論理的に隔離されている」という位置づけである．近年は，NeoCloud（ネオクラウド）[16]と呼ばれる特化型GPUクラウド（CoreWeave, Nebius Group, Lambda Labs等）による「ベアメタル」インスタンス（単一の顧客に専有される物理サーバー）の提供も進んでいる．これは，オンプレミスに近い「物理的な主権（他テナントとの混在なし）」と，クラウド特有の「調達の速さと弾力性」を両立させる，第三の選択肢として注目されている．

このアプローチの強みは，自社でGPUハードウエアを保有・管理する負担がなく，スケーラビリティが高いことである．Azure OpenAI Serviceのプロビジョンドスループットや Amazon Bedrockなど，既存の企業向けクラウド契約の延長線上で導入できる場合が多い．冷却設備や電力供給の管理といった物理インフラの煩雑さを回避できる点も大きい．

弱みは，「厳密には自社の物理環境ではない」というガバナンス上の問題が残ることである．防衛関連企業や機密情報を扱う政府機関のように，高度なデータ主権要件を持つ組織にとっては，米国法（CLOUD Act[*18]）の適用範囲に含まれる可能性がある米国企業のインフラ上にデータを置くこと自体が許容できないリスクになりうる．また，継続的な API 利用料やインスタンス稼働費用がランニングコストとして積み上がる構造も無視できない．社内利用が想定以上に拡大した場合，短期間で月間予算を消化する恐れがある．

4. 3　アプローチ C：エンタープライズ特化のパッケージ型ローカル LLM

企業向けに最適化された形でハードウエアまたはソフトウエアがパッケージングされて提供されるローカル LLM ソリューションである．この領域では，設計思想の異なる複数の選択肢が存在する．例えば，Lazarus AI 社との戦略パートナーシップに基づきユニアデックス株式会社（以下，ユニアデックス）が展開するソリューション[17][18][19]は，ハルシネーションを抑制した高精度な文書処理・情報抽出と，セキュアな閉域環境での運用に軸足を置いた構成である．一方，日立ソリューションズが Allganize と提携して提供する「Alli LLM App Market[20]」は，SaaS 型からオンプレミス型まで選択できる生成 AI アプリ基盤として位置づけられている．また，NTT の「tsuzumi[11]」は，1GPU で動作可能な国産 LLM として，自社インフラや SIer 基盤へ組み込む前提で採用が進んでいる．

このアプローチの強みは，導入の確実性とコストの予測可能性である．LLM 単体ではなく，RAG に必要なベクトルデータベースや UI，権限管理，さらにはアプライアンスサーバー[*19]が一体の「プレパッケージ」として提供されるケースが多く，AI 専門エンジニアを持たない企業でも本番適用を実現しやすい．完全なオンプレミス・閉域網での稼働が可能なため，製造業の設計データや金融機関の顧客データといった機密性の高い要件に合致する．また，買い切り型やリース契約により初期コスト化しやすい点は，日本企業の稟議プロセスとも相性が良い．

弱みは，「モデルの陳腐化への対応」と「スケーラビリティの物理的限界」である．日進月歩で進化する LLM 業界において，閉域環境のパッケージ製品は最新モデルへのアップデートにタイムラグが生じやすい．また，処理リクエストが急増した際，クラウドのように即座にリソースを拡張することが難しく，独自の既存システムと深く API 連携させる際にもベンダーのアーキテクチャに依存する制約が生じやすい．

ただし，近年では知識基盤をモデル本体から分離するアーキテクチャも登場しており，業務知識の更新をモデルの再学習に依存させず，外部の知識基盤側で管理することで，モデルを頻繁に入れ替えることなく最新情報への追従を図る取り組みも進んでいる．Lazarus AI の Vector Knowledge Graph（VKG）はその一例であり，7. 4 節で詳述する．

本章で述べたローカル LLM を実現するための三つのアプローチの比較を表 1 にまとめた．

表1　ローカルLLM実現アプローチの比較

比較項目	A：OSS自社構築	B：VPC型	C：パッケージ型
初期コスト	中〜高 （GPU調達・構築費）	低	中〜高 （パッケージ・ハード費）
ランニングコスト	低 （電気代・保守費）	高 （Cloudサービス利用費）	低〜中 （保守ライセンス費）
AI/ML専門人材	高度な人材が必要	構成により必要	原則として限定的
基盤運用・保守	自社対応が中心	クラウド側が一部担当	ベンダー支援を利用可能
スケーラビリティ・弾力性	低（ハード追加が必要）	高（クラウドの弾力性）	低（ハード追加が必要）
カスタマイズ性	高	中	限定的
データ主権・機密性	極めて高い	ベンダー規約に依存	極めて高い
ハルシネーション制御	自社でパイプライン構築	ベンダー機能に依存	製品の設計・機能に依存
導入までの期間	長い	短い	短い

5.　ローカルLLMの国内先行事例と段階的導入指針

本章では，国内における先行事例を踏まえながら，ローカルLLMを実運用へつなげる際の導入上の勘所を確認する．

5.1　国内の先行事例

ローカルLLMや閉域環境での生成AIの導入は，セキュリティ要件の厳しい業界を中心に実証・商用化がすでに進んでいる．

金融分野では，あおぞら銀行がneoAIと協業し，行内オンプレミス環境に金融業務に特化した独自のLLM基盤の構築を進めている[21]．また，同行は行内規定やマニュアルをRAG（検索拡張生成）技術と組み合わせて照会できる専用アプリを導入し，回答の根拠を提示させることでハルシネーションを抑えつつ検索業務を効率化している．常陽銀行も，クラウド上の閉域環境にChatGPTを構築し全行員向けに展開しているほか，RAGを活用した「営業ソリューション検索サービス」を追加し，行員の業務効率化を推進している[22]．

医療分野では，機密性の高い患者データの取り扱いが必須となるため，オンプレミス環境でのLLM活用が不可欠である．社会医療法人祐愛会織田病院は，オプティムが提供するオンプレミス生成AI「OPTiM AIホスピタル」を導入し，電子カルテと連携させた[23]．これにより，外部へデータを出すことなく，退院時看護サマリーの作成時間を約54%短縮するという実用的な成果を上げている．

これらのケースでは，顧客情報や患者の個人情報をパブリックな外部サーバーに送信しにくいという制約があるため，オンプレミス環境や専用の閉域環境で運用可能なLLM基盤が有力な選択肢となっている．これらの事例は，厳格なデータガバナンスが求められる領域においても，セキュアな環境構築とRAGの活用によって実用的な成果を目指せることを示している．

5.2　導入における留意点

「ローカルだから無条件に安全」という思い込みは禁物である．ローカルLLMは外部送信

リスクを排除するが，内部での権限管理，物理的なサーバーセキュリティ，マルウエア感染リスクは依然として存在する．トレンドマイクロが警告するモデルファイルへの悪意あるコードの埋め込みリスク[8]は，OSS モデルを外部リポジトリから取得する際に現実に発生しうる．

また，生成結果の正確性に対する過信も危険である．RAG のチャンク*20 設計が不適切であれば的外れな文書を参照し，文書の鮮度管理が不十分であれば古い情報に基づいた回答が生成される．単に検索精度を上げるだけでなく，「根拠に基づかない回答を徹底的に排除する」といった，事実性に特化したアーキテクチャを採用するか，自社で RAG の検索品質を継続的に改善する体制を持つかが問われる．

さらに，PoC はまず小さく始めるのが妥当である．文書検索，規定照会，情報抽出のように正解と根拠を確認しやすいユースケースから着手し，アクセス制御，ログ監査，ドキュメント更新責任を並行して設計しなければ，管理下で運用するという前提が形骸化する．「導入したら終わり」ではなく，継続的な運用改善を前提に据えることが肝要である．

これら先行事例と運用上の留意点は，ローカル LLM が実用フェーズにあることを示している．一方で，汎用モデルの延長線上では解決が難しい「事実の厳格な検証」という壁も明らかになった．

6. ローカル LLM 導入のコスト構造の考察

本章では，クラウド API とオンプレミス運用の費用構造を対比しながら，ローカル LLM 導入時に見るべきコストの全体像を整理する．

6. 1 「使うたびに払う」か「買って持つ」か

クラウド AI は「初期費用なしで今すぐ使える」という手軽さが魅力であった．しかし業務利用が本格化するにつれて，問題が顕在化してきた．クラウド API は処理した文字数（トークン数）に応じて課金される従量制のため，社員全員が日常的に使い始めたり，大量の文書を自動処理するシステムに組み込んだりすると，利用料が青天井で積み上がる構造にある．利用量が少ないうちはクラウドに軍配が上がる．1 日あたり 1000 件程度の問い合わせであれば，初期費用が不要なクラウドが明らかに安価である．しかし，保険金請求の自動処理や社内文書の一括解析など，大量のドキュメントを扱うような業務にクラウド API を使い続けると，その費用は現実的ではなくなる．

一方，自社サーバーで LLM を動かす場合の目安を示す．3 章で述べた OpenAI のオープンウェイトモデル「gpt-oss」を例にとると，gpt-oss:20b（20B パラメータ）は NVIDIA RTX 4090 を 1 ～ 2 枚搭載したワークステーションでの運用が現実的であり，日常的な文書検索・FAQ 対応用途の候補となる．より高い回答品質を求める場合は gpt-oss:120b（120B パラメータ）が選択肢となるが，こちらはより大きな GPU メモリを持つサーバー構成が不可欠になる．いずれの構成でも，機器調達の初期費用を除けば追加の処理コストは主として電力と運用保守費である．国内インフラ企業の試算でも，安定稼働時はオンプレミスが 3 年間で 30 ～ 50％のコスト削減になると報告されており[24]，利用頻度が高く，問い合わせ件数や処理文書量が大きい業務では，従量課金型のクラウド API よりオンプレミスの方が総コストで有利に転じる可能性がある（図 3）．

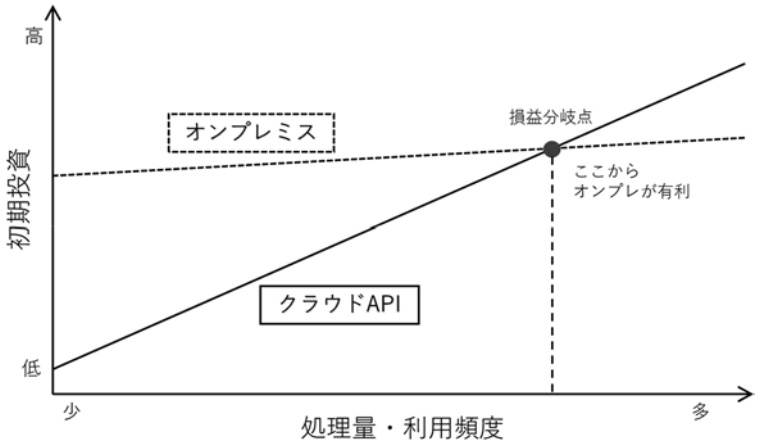

図3　クラウドAPIとオンプレミスLLMのコスト構造と損益分岐点

6.2　見落とされがちなコスト

ハードウエアの費用に加えて，自社運用では電力や冷却コストもかかる．高性能GPUを多数収容する高密度ラック（40kW超など）では液冷化の設備改修を要するケースもあり，電気代の高い日本では無視できない負担となる．また，システムの維持にはAI/MLエンジニアが不可欠であり，その人件費やSIerへの委託費も計上すべきである．

クラウドAPIはこれら運用の手間やコストが利用料に含まれるため比較しにくい．しかし，大量データの送受信には別途「データ転送料（Egress料金）」が発生し，従量制ゆえに予算が立てにくい点に注意すべきである．裏を返せば，固定費での予算管理を好む日本企業にとって，コストが読みやすいオンプレミスの方が社内稟議を通しやすい側面もある[25]．

6.3　「買ったGPUがすぐ古くなる」という懸念について

「高額なGPUを購入しても短期間で陳腐化する」という不安は，導入をためらわせる典型的な理由である．しかし，新規学習と推論用途では求められるハードウエアの更新頻度が異なる．推論に限定すれば，AWQ等の4bit/8bitへの「量子化技術」によりメモリ使用量を劇的に削減することができる．実際に海外事例では，量子化等により推論のリソース要件を最大70%削減しつつ品質を維持した報告もある[26]．こうした技術進化が旧世代ハードウエアのライフサイクルを延ばし，長期間の実運用に耐えられるようにしている．

7.　エンタープライズ特化型ローカルLLMの技術的特徴―Lazarus AIという選択肢

5章で確認した通り，企業が機密資産を知能化する上で，モデルの「表現力」以上に「事実の正確性」が成否を分ける．また前章では，クラウドAPIとオンプレミス運用のコスト構造と判断軸を整理した．本章では，これらを踏まえ，事実抽出の正確性と閉域運用を両立するソリューションとしてLazarus AIを取り上げる．

7.1　セキュリティ設計と日本展開―ユニアデックスとの戦略的提携

Lazarus AIは，文書処理と情報抽出に特化した企業向けAIである．保険・金融・政府機関などの高機密領域での利用を想定し，顧客専有のクローズド環境で処理を完結させる構成や，

参照元提示による検証可能性を特徴とする[19].

防衛産業や重要インフラにおいて求められる，インターネットから完全に遮断されたエアギャップ環境*21 や，強い閉域性を持つネットワーク環境への展開にも対応できる.

国内では，ユニアデックスが 2025 年 7 月に戦略パートナー契約を締結し，オンプレミス向け LLM ソリューションの提供を開始した[17]．これにより，RikAI2（7.3 節）や VKG（7.4 節）を中心とした構成を，日本企業の閉域環境へ展開する体制が整っている.

7.2　ATLS プラットフォーム――「AI を指揮するための AI」

Lazarus AI の技術的な中核を担うのが，ATLS と呼ばれるオーケストレーション・プラットフォームである[27]．ATLS は，タスクに応じて最適な AI モデルを選択・配分する「AI を指揮するための AI」として機能し，RikAI2 をはじめとする複数の専門モデルを統合的に制御する.

米国国防総省（DoD）などの情報分析事例では，膨大なマルチソース・データの相関分析において，人間が約 8 週間を要するタスクを数分へ短縮し，大幅なコスト削減を実現したとされる[19]．なお，2026 年 3 月時点では政府・防衛分野での実績が先行しており，民間向け展開のロードマップは今後順次公開される見込みである.

日本企業にとって ATLS が示唆するのは，「単一のモデルですべてを解決する」のではなく，「複数の専門モデルを協調動作させる」というエージェンティックなアプローチの有効性である．ここで重要なのは，2 章で提起した「AI エージェントが自律的に外部と通信し，機密情報を漏洩させるリスク」に対して，ATLS が明確な解決策を提示している点である．ATLS は「AI を指揮・監視する AI」として機能し，エージェントごとの権限を最小限に制限して，実行可能なアクションやデータアクセスの境界を厳密に統制する.

このように，暴走リスクを抑え込みながら複数の AI を安全に連携させ，実業務のプロセスに適合させるアプローチこそが，これからのエンタープライズ AI 運用における重要な鍵となる．この思想の核となるのが，次節で述べるフラッグシップモデル RikAI2 である.

7.3 「生成」ではなく「抽出・精査」という設計思想― RikAI2 の技術的革新

企業業務（法務判断，医療記録の参照，財務規定の照会など）における LLM 活用では，事実と異なる情報を生成するハルシネーションが致命的リスクとなる.

これに対して Lazarus AI は「We optimize for truthful reasoning」という思想を掲げ，一般的な生成型 AI よりも，文書からの正確な抽出と検証可能性を重視する[17]．その中核モデル「RikAI2」の技術的革新は，LLM と Vision Transformer（ViT）*22 を統合し，抽出型 AI*23（Extractive AI）としての動作原理を確立している点にある[28].

汎用的なマルチモーダルモデル*24 は視覚情報に対しても「生成（推測）」を行うためハルシネーションのリスクがある．一方で，RikAI2 は文書を視覚情報として直接解釈し，それを「厳格な事実抽出」にのみ用いる．そこに真の独自性がある（図 4）．RikAI2 では，手書き文字や複雑な表・グラフ・図面等のレイアウト構造を正確に把握でき，引用元の確実な明示（事後検証）を実現できる（図 5）．派生モデルとして構造化データ抽出に特化した「RikAI2-Extract」も提供されている.

図 4　汎用 LLM と RikAI2 における文書処理アーキテクチャの比較

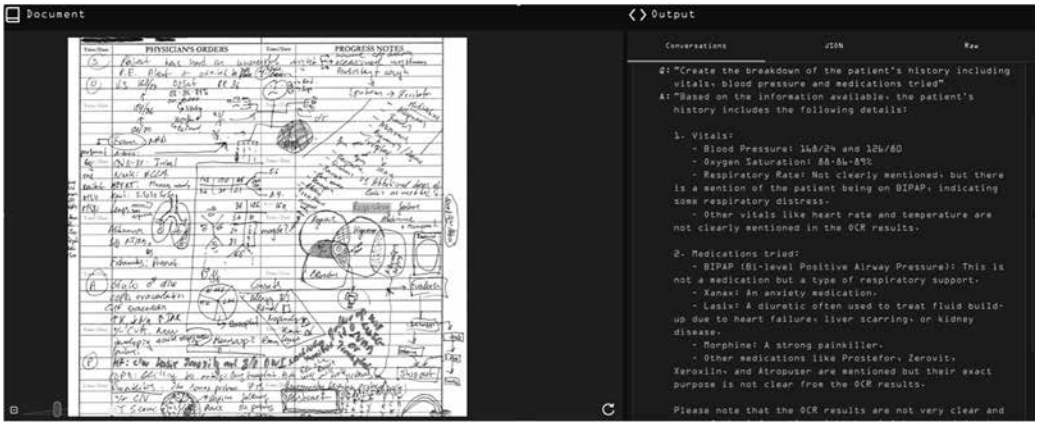

図 5　RikAI2 による手書き医療カルテの読み取り事例[19]

7. 4　Vector Knowledge Graph アーキテクチャ

RikAI2 の抽出精度を支える知識基盤が「Vector Knowledge Graph（VKG）」アーキテクチャである[19]．企業のドメイン知識*25 をモデルに組み込む際，ファインチューニングは再学習コストが高く，一般的な RAG（文書をチャンク分割するベクトル検索）では文書間の論理的関係や階層構造が失われやすい．加えて，OSS などを組み合わせた一般的な RAG 構成では，検索用の埋め込みモデルと生成用の LLM が異なるケースが多く，両者の意味や表現のズレが精度低下の要因となることがある．

VKG はベクトルデータベース*26 と知識グラフ*27 のハイブリッド構造により，検索の柔軟性を維持しつつ文書間の関係性を保持し，複雑な文脈理解を実現する．さらに，Lazarus AI のアーキテクチャはベクトル化から情報抽出に至るプロセスが統合的に設計されているとみられ，検索空間と生成モデル間の意味的なズレが生じにくく，より効率的で高精度な処理が期待できる．内部ドキュメントをインポートするだけで知識ベースがリアルタイムに更新されるため，再学習を伴わない運用が実現できる．これにより，新しいドキュメントの追加は知識グラフの拡張として処理され，業務マニュアルや規定集の更新にも即座に追従できる（図 6）．

以上を踏まえると，Lazarus AI は汎用的な生成性能よりも，規定集，契約書，設計書，報告書などの「根拠を示しながら読み解く・抽出する」業務に特化したローカル LLM である．閉域運用，参照元提示，文書構造理解の統合に同社の差別化要因がある．ATLS を含む全体アーキテクチャの国内展開にはまだ流動的な部分もあるが，出力の透明性と説明可能性を備えた RikAI2 と VKG を軸とする構成は，現時点でもセキュアな文書処理基盤として高い導入価値を持つ．

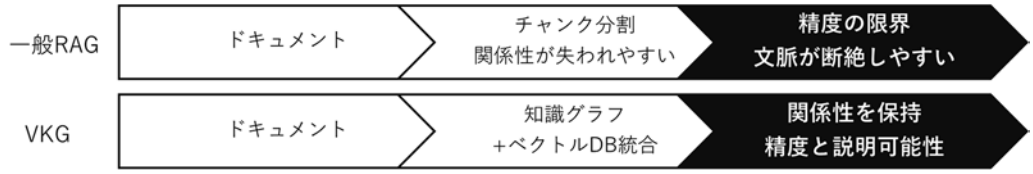

図 6　一般的な RAG と Vector Knowledge Graph の知識構造の違い

8.　Lazarus AI の位置づけとハイブリッド運用

本章では，Lazarus AI を全面置換のための製品としてではなく，クラウド AI と併用するハイブリッド構成の中での位置づけ，ガバナンス[*28]を維持しつつ安全に運用すべきかを考察する．

8. 1　ハイブリッド構成の中での位置づけとガバナンス

ここまで，データ主権を保護し，厳格な事実抽出を実現するローカル LLM としての Lazarus AI の特性を詳述してきた．しかし，実際のエンタープライズ環境において，すべての AI タスクをローカル LLM のみで完結させることは，インフラのコストや処理の汎用性を考慮すると必ずしも現実的ではない．実際の現場で求められるのは，機密性の高いデータの処理や厳格な事実抽出には Lazarus AI を配置し，一般的な文章の要約やアイデア出しにはクラウド LLM を利用するといった，適材適所の「ハイブリッド構成」である．これが企業における生成 AI 活用の現実解であり，Lazarus AI が担うべき中核的な位置づけとなる．

しかし，このデータ主権とリスク許容度に基づく運用を安全に行うためには，単に「重要なデータはローカルで処理する」というネットワークの境界防御だけでは不十分である．データが「いま，どこで，どのように処理されているか」をシステム内部で常に把握できる状態（可観測性[*29]：Observability）を作り，国際基準に沿った統合的なガバナンスを効かせることが不可欠となる．具体的には，ISO/IEC 42001[29]が求める AI ライフサイクル全体での透明性の確保や，NIST AI 600-1[30]が推奨する継続的な監視体制の構築である．同時に，OWASP Top 10 for LLM Applications[9]が警告するプロンプトインジェクションや，意図せぬデータの流出といった固有リスクに対しても，システム全体で一貫した防御ポリシーを適用しなければならない．

8. 2　AI 統合管理基盤とドメイン特化型導入の展望

こうしたガバナンス要件を満たしつつ，安全なハイブリッド運用を実現する現実的な手段が，AI ゲートウェイや統合監視ソリューションなどの可観測性基盤の導入である．これらを活用することで，ユーザーの入力に機密情報が含まれる場合には自動で Lazarus AI へ処理を振り分けるセマンティック・ルーティング[*30]といった動的なトラフィック制御ができるようになる．導入にあたっては，ゲートウェイが処理のボトルネックとならないようパフォーマンスを確保しつつ，自律型 AI[*31]（Agentic AI）に対する権限付与を最小限に留め，NIST が求める継続的な監視の原則をクラウド・ローカル双方に適用することが考えられる．

これらクラウドとローカルを横断する複雑なネットワーク境界の制御や，統合的な可観測性の確保は，AI モデル単体の知識だけでは実現できない．IT インフラ全体を俯瞰し，ネットワークやセキュリティに精通したパートナーが，企業ごとの要件に合わせてシステム全体を最適化することで，初めて安全なハイブリッド運用が成立する．これらのアーキテクチャ要件を踏ま

えた上で，現実のハイブリッド構成において重要なのは，6章で述べたLazarus AIが万能基盤ではなく，機密性と検証可能性を要する領域を担う中核として位置づけられる点である．Gartner® は，2027年までに企業で利用される生成AIモデルの過半数が業務・ドメイン特化型モデル（特定の業界や業務機能に特化したもの）になると予測している．これは2024年時点の1%から増加となる[31]．この潮流において，文書理解と検証可能性を重視したLazarus AIは，既存のクラウドAIを全面置換するのではなく，限定領域から導入するローカルLLMの有力な選択肢となる．

9. おわりに

本稿では，生成AIの急速な普及を背景にセキュリティとデータ主権の観点からローカルLLMが求められる背景を整理し，Lazarus AIの技術的特徴と企業導入上の位置づけを考察した．

同製品の意義は，単に閉域環境で運用できる点にとどまらない．RikAI2による文書理解・情報抽出とVKGによる知識基盤，そしてATLSによる複数モデル協調の構想を通じて，汎用LLMが持つ「表現の多様性」よりも「推論の正確性と検証可能性」を優先した点にある．とりわけ厳格なガバナンスが求められる規制領域において，この設計思想は実運用上の重要な意味を持つ．汎用的な生成AIが持つ利便性を享受しつつも，自社のコアコンピタンス[*32]となる機密データは，確実な制御下で安全に処理・抽出できなければならない．Lazarus AIの抽出・検証機能に，自社環境に適したインフラ統合を掛け合わせることは，企業がデータ主権を保護しながらAIを業務適用するための有力なアプローチとなる．

さらに将来を見据えれば，この堅牢な推論基盤は，自律的に業務を遂行するAgentic AI時代における「マルチエージェント協調」の重要な布石となる．例えば，クラウド上の汎用エージェントが収集した外部市場データを，オンプレミス環境のLazarus AIが社内の機密ドキュメントと安全に照合・精査するといった，セキュリティ境界を越えた連携が実現できる．事実に基づく正確性を担保した環境があってこそ，AIは単なる「回答者」から信頼できる「エージェント」へと飛躍できるだろう．

＊1　クラウドを使わず，自社の建物内に設置したサーバーでシステムを運用する形態．
＊2　特定の企業・組織が専用に利用するクラウド環境．パブリッククラウドとは異なり，他社とインフラを共有しない．
＊3　自社のデータをどの国・どの組織が管理・処理するかを自ら決定できる権利・状態．
＊4　メール送信・ファイル操作・外部サービスの呼び出しなど，人に代わって自律的に作業を実行するAI．
＊5　AIへの入力（プロンプト）に悪意ある命令を紛れ込ませ，意図しない動作をさせる攻撃手法．
＊6　AIの学習データに意図的に誤情報を混入させ，モデルの動作を狂わせる攻撃．
＊7　不特定多数の企業・開発者が利用できる公開されたAIサービスの接続口．ChatGPTやGeminiなどが該当し，インターネット経由でデータを送受信する．
＊8　AIモデルのパラメータのビット数を削減し，精度をほぼ保ちながらメモリ消費・計算コストを削減する技術．
＊9　LLMの量子化モデルの代表的なファイル形式．メモリを節約しながら，LLMをローカル環境で動かすために使われる．
＊10　EUが定めたAIに関する包括的な規制法．AIのリスクレベルに応じて義務や禁止事項を定めており，段階的に施行されている．
＊11　EUの個人情報保護法（General Data Protection Regulation）．個人データの収集・処理・移転に関するルールを定めており，違反には高額の制裁金が科される．AI活用においても

データの取り扱い方法に影響する．

＊12 オープンソースソフトウエアのこと．ソースコードが公開されており，誰でも無償で利用・改変・再配布できるソフトウエアの総称．本稿では Llama や Qwen，Mistral などの公開された AI モデルを指す．

＊13 社内文書などを検索して取得した情報をもとに AI が回答を生成する手法．事実性の向上に有効．

＊14 汎用 LLM モデルに特定業務のデータを追加学習させ，専門領域に最適化すること．

＊15 AI が事実と異なる情報を，もっともらしく生成してしまう現象．

＊16 悪意を持って作られた不正なソフトウエアの総称．ウイルス・ランサムウエアなどが含まれる．文中では OSS の LLM モデルのファイル自体に仕込まれるリスクとして言及されている．

＊17 新しい技術やシステムが実際に機能するかを小規模で検証する「概念実証」のこと．本格導入前に効果や課題を確認するための試験的な取り組み．

＊18 米国企業に対して，海外サーバー上のデータも米政府の要求に応じて開示を義務付ける米国法．外資クラウド利用のリスク要因．

＊19 特定の用途に特化して，ハードウエアとソフトウエアがあらかじめ一体化された状態で提供されるサーバー機器．

＊20 RAG で長い文書を扱う際に，小さなブロック（チャンク）に切り分けること，およびその分割方法の設計．分割が不適切だと文脈が失われ，的外れな回答が生成される原因になる．

＊21 インターネットや外部ネットワークから物理的に完全に隔離された環境．防衛産業や重要インフラ，高度な機密情報を扱う組織で求められる，最も強固な閉域運用形態．

＊22 画像を細かなパッチに分割し，文書のように解析する視覚 AI モデル．RikAI2 の基盤技術

＊23 文書の内容をもとに新たな文章を「生成」するのではなく，文書中に書かれている事実をそのまま「抜き出す」ことに特化した AI．根拠が明示できるため，ハルシネーションが起きにくい．

＊24 テキスト・画像・音声など複数の種類のデータを統合的に扱える，特定用途に限定しない幅広い目的向けの AI モデル．GPT-4o や Gemini などが該当する．

＊25 特定の分野，業界，または業務に関する専門的な知識のこと．

＊26 テキストや画像などを数値化（ベクトル）して保存し，単語の完全一致ではなく「意味の類似性」で高速検索を可能にする AI 向けのデータベース．

＊27 データや概念を「点」とし，その関係性を「線」で結んでネットワーク状に構造化したもの．AI が単なる単語の検索だけでなく，情報同士のつながりや複雑な文脈を理解するための基盤技術．

＊28 AI や機密データを安全かつ適正に管理・運用するための社内ルールや統制（アクセス制御，監査，情報漏洩対策など）の仕組み．

＊29 システムの外部出力から内部の状態や動作を把握する能力．AI 運用においては，モデルの回答精度，処理遅延，トークン消費量，エラーなどをリアルタイムで監視・追跡し，問題を迅速に特定する仕組み．

＊30 入力されたテキストの意味を理解して，最適な処理先やモデルに振り分ける仕組み．

＊31 目標に基づき，自律的に計画・実行・ツール利用を行う AI エージェント群の総称．

＊32 他社には真似できない，顧客に独自の価値を提供する企業の中核的な能力のこと．

参考文献

［1］ 総務省，令和 7 年版 情報通信白書，2025 年 7 月，
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/nd112220.html

［2］ Cyera，Cybersecurity Insiders，2025 State of AI Data Security Report，2025，
https://www.cyera.com/research-labs/2025-state-of-ai-data-security-report

［3］ 独立行政法人情報処理推進機構，「テキスト生成 AI の導入・運用ガイドライン」，
独立行政法人情報処理推進機構，
https://www.ipa.go.jp/jinzai/ics/core_human_resource/final_project/2024/f55m8k0000003spo-att/f55m8k0000003svn.pdf

［4］ 総務省，AI 利活用ガイドライン，
https://www.soumu.go.jp/main_content/000624438.pdf

［5］ Gartner®，Emerging Risk Deep Dive: Shadow AI，Ben Fisher et al.，18 August 2025
https://www.gartner.com/en/documents/6714034（Gartner Subscription Required）
GARTNER is a trademark of Gartner, Inc. and its affiliates.

［6］ Okta，Enterprise Buyer Survey: AI Agent Security，2026 年 2 月，
https://www.okta.com/newsroom/articles/enterprise-buyer-survey-ai-agent-security/

［7］ NTT ドコモビジネス，「生成 AI が情報漏えいを引き起こす？「シャドー AI」のリスクと対策を解説！」，https://www.ntt.com/bizon/shadow-ai.html

［8］ トレンドマイクロ，「AI エージェントと脆弱性 PART 1」，
https://www.trendmicro.com/ja_jp/research/25/f/unveiling-ai-agent-vulnerabilities-

part-i-introduction-to-ai-agent-vulnerabilities.html
[9] OWASP, Top 10 for LLM Applications 2025, https://genai.owasp.org/llm-top-10/
[10] NTTPCコミュニケーションズ，セキュア×高速！ローカルLLMが変える企業AI活用とデータガバナンス最前線，2025年12月，
https://www.nttpc.co.jp/gpu/article/knowledge22_local_llm.html
[11] NTT，NTT版大規模言語モデル「tsuzumi 2」，2025年10月
https://www.rd.ntt/research/LLM_tsuzumi.html
[12] ITmedia PC USER，M4 Mac miniで「gpt-oss」は動く？ 動作が確認できたローカルLLMは……，2025年9月，
https://www.itmedia.co.jp/pcuser/articles/2509/30/news035.html
[13] Ollama，AMD support，2024年3月，https://ollama.com/blog/amd-preview
[14] PC Watch，Ryzen AI Max+ 395を「ビデオメモリ128GB積んだ鬼GPU」にしてしまう方法，2025年10月，
https://pc.watch.impress.co.jp/docs/column/nishikawa/2054963.html
[15] NVIDIA，NVIDIA NIMマイクロサービス，
https://www.nvidia.com/ja-jp/ai-data-science/products/nim-microservices/
[16] NTTドコモビジネス，「ネオクラウドとは？」，
https://www.ntt.com/business/services/xmanaged/lp/column/neocloud.html
[17] ユニアデックス株式会社，Lazarus AIとの協業によるオンプレミス向けLLMソリューション提供，2025年7月，
https://www.uniadex.co.jp/news/2025/20250717_lazarus-ai.html；
[18] ユニアデックス株式会社，Lazarus AI製品情報，
https://solution.uniadex.co.jp/service/product/lazarus-ai.html
[19] Lazarus AI，Technology，https://www.lazarusai.com/technology
[20] 日立ソリューションズ，Alli LLM App Market，
https://www.hitachi-solutions.co.jp/allganize/
[21] PR TIMES，あおぞら銀行 x neoAI オンプレミス型次世代AI基盤構築に向けて，金融・行内特化LLMを開発，2025年4月17日，
https://prtimes.jp/main/html/rd/p/000000026.000109048.html
[22] PR TIMES，【常陽銀行】生成AI「ChatGPT」のバージョンアップおよびRAGを活用した「営業ソリューション検索サービス」の取り扱い開始について，2025年9月18日，https://prtimes.jp/main/html/rd/p/000000033.000081984.html
[23] オプティム，オンプレミス生成AIサービス「OPTiM AIホスピタル」によって退院時看護サマリー作成にかかる時間を54.2%削減―社会医療法人祐愛会織田病院様，2025年10月6日，https://www.optim.co.jp/media/cat-case/aih_251006-01
[24] Ragate，オンプレミスLLM構築の実践：セキュアな閉域環境での大規模言語モデル運用術，https://www.ragate.co.jp/media/developer_blog/xooise_gx7xb
[25] マクニカ，クラウドでは難しい社外秘データの活用に向けた生成AI ～NVIDIAを例にローカルLLMについて考える～，
https://www.macnica.co.jp/business/semiconductor/articles/nvidia/148236/
[26] Reinforz BizDev，「生成AI原価の真実と収益最大化戦略：GPUコスト時代の単価設計と粗利管理」，https://bizdev.reinforz.co.jp/4557
[27] Lazarus AI，ATLS for Department of Defense Information Analysis，
https://www.lazarusai.com/government/atls-for-department-of-defense-information-analysis
[28] アルゴスサービスジャパン株式会社，「ハルシネーション（幻覚）」を最低限に抑制できるLazarus RikAI2，https://note.com/argossj/n/n5b33813b0b75
[29] ISO, ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system, https://www.iso.org/standard/81230.html
[30] NIST, NIST AI 600-1: Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile,
https://www.nist.gov/itl/ai-risk-management-framework
[31] Gartner®，Press Release, Gartner Forecasts Worldwide End-User Spending on GenAI Models to Total $14.2 Billion in 2025，2025年7月10日，
https://www.gartner.com/en/newsroom/press-releases/2025-07-10-gartner-forecasts-worldwide-end-user-spending-on-generative-ai-models-to-total-us-dollars-14-billion-in-2025

※　上記の参考文献に示したURLのリンク先は，2026年5月14日時点での存在を確認.

執筆者紹介　青　木　岐　夫（Michio Aoki）

2021年ユニアデックス(株)中途入社．AI・DX分野における新規サービスの開発・導入支援業務に携わる．現在はLazarus AIの開発・導入支援を主業務とし，生成AI関連の最新技術動向を継続的に調査・研究している．JDLA認定 E資格保有.

プレス成形における摩耗予測とワレ予兆の比較に基づく生成 AI によるひずみ波形解析支援の可能性と限界

Possibilities and Limitations of Generative AI-Assisted Strain Waveform Analysis Based on a Comparison of Punch Wear Prediction and Crack Precursor Detection in Press Forming

長　島　正　貴

要　約　実生産ラインの金型に設置したひずみ・荷重センサーの時系列データに対し，生成 AI，とりわけ大規模言語モデル（Large Language Model，LLM）を解析支援に活用することを試みた．これにより，非 IT 技術者でも特徴量設計，可視化，異常検知設計に参加しやすい手順の整備を目指している．ピアスパンチ摩耗予測では，位相整列と区間化に基づく特徴量設計が有効であり，LLM は仮説整理，コード生成，可視化テンプレート化に寄与した．一方，絞り加工におけるワレ予兆では，前処理や可視化の初速向上には有効であったが，発生位相の自律的特定には限界があった．これらの比較を通じて，現時点の実用解は，人間が現象理解と評価観点を設計し，LLM が手順化，再現化，説明補助を担う役割分担にあることを示す．

Abstract　This paper describes an attempt to apply generative AI, in particular large language models (LLMs), to support the analysis of time-series data acquired from strain and load sensors installed on dies in an actual production line. The aim is to establish procedures that allow even non-IT engineers to participate in feature design, visualization, and anomaly detection design. In pierce punch wear prediction, feature design based on phase alignment and segmentation proved effective, and the LLM contributed to hypothesis organization, code generation, and visualization templating. In contrast, for crack precursor detection in deep drawing, the LLM was effective in accelerating the early stages of preprocessing and visualization, but it had limitations in autonomously identifying the phase at which the anomaly emerges. Through this comparison, we show that the practical solution at present lies in a division of roles in which humans establish the understanding of the phenomena and design the evaluation perspectives, while the LLM handles proceduralization, reproducibility, and explanatory support.

1. は　じ　め　に

製造現場における品質安定化は，単に不良率を下げるための活動ではなく，設備停止，後工程での手戻り，部材廃棄，保全負荷の増大を抑える基盤である．とりわけプレス成形では，金型の状態，パンチの摩耗，材料条件，潤滑状態，段取り変更など，複数の要因が品質へ重畳的に影響するため，不具合が顕在化する前に兆候を把握できる仕組みへの期待が高い．一方で，品質異常の兆候は必ずしも目視点検だけで安定して捉えられるわけではなく，加工中の荷重やひずみの変化を継続的に観測することが求められる．

2017 年頃から，実生産ラインの金型や周辺設備にセンサーを設置し，高頻度に取得した時系列データを品質監視や保全判断へ結び付ける取り組みが進んでいる．ひずみ・荷重波形には，

加工中の変形挙動や破断挙動が反映されるため，工程物理に即した特徴量を設計できれば，従来の点検結果だけでは見えにくい変化を定量的に追跡できる可能性がある．しかし，この可能性を実務へ落とし込むには，データ解析の専門知識と，前処理・可視化・モデル化を実装するITスキルの双方が求められる．この二重の要求が，現場でのデータ活用を難しくしてきた．

従来の解析では，工程理解のある技術者が，どの区間に注目すべきか，どのピークや立上り，積分値，時間差が意味を持つかを考え，そのうえでIT技術者がコードを書く構図になることが多かった．この構図では，分析の狙いが口頭で伝達される過程で曖昧になりやすく，コードと可視化の実装が属人化しやすい．また，試行錯誤のたびにグラフや特徴抽出を作り直さなければならず，現場レビューへ持ち込むまでに時間が掛かる．その結果，良い仮説があっても，検証速度が追い付かず，解析結果が継続的な運用改善に結び付かない場合もあった．

この課題に対し，本稿では，大規模言語モデルを解析支援に用いる意義を，単なるコード自動生成にとどめずに捉える．具体的には，自然言語による仮説整理，手順のテンプレート化，可視化コードの生成，レビュー観点の整理，再現性の確保という観点にまで広げて考える．LLMは，工程知識そのものを代替するものではないが，現場技術者が持つ暗黙知を自然言語で表現し，それを解析手順へ変換する媒介として有効である可能性がある．また，生成された手順やコードを固定化することで，同じ入力に対して同じ出力を得やすくなり，技術継承や横展開の基盤にもなり得る．

本稿では，この考え方を二つの事例で検討する．第一は，ピアスパンチ摩耗予測である．これは，連続的かつ累積的に進行する劣化現象であり，位相整列と区間化を行うことで，物理的意味を持つ特徴量を比較的安定して定義しやすい例である．第二は，絞り加工におけるワレ予兆である．これは，突発的かつ条件依存的に発生し，どの位相で異常の兆候が現れるかを適切に捉えること自体が難しい例である．この二事例を対比することで，LLMが有効に働く条件と，なお人間の専門知識が不可欠な条件を整理する．

本稿の目的は，LLMを用いれば何でも自律的に解けると主張することではない．むしろ，現時点で現場に適用できる役割分担を明らかにし，非IT技術者でも扱える解析手順をどのように構築すべきかを示すことにある．2章では，まず対象データと現場課題を整理し，3章ではLLMによる波形解析支援プロセスを説明する．その後，4章で成功例としてのピアスパンチ摩耗予測を，5章で課題例としての絞り加工ワレ予兆を詳述し，6章で成功例と課題例の比較と考察を行い，7章で今後の展望を述べる．

2. 対象データと現場課題

本検討において対象とするデータは，実生産ラインの金型に設置したひずみセンサーおよび荷重センサーから取得される時系列データである．これらのデータは，1ストローク中の変形挙動，せん断，破断，材料解放後の戻りなどを連続的に含んでおり，加工現象との対応付けができれば豊富な情報源となる．一方で，実ラインのデータは，研究室環境で整然と取得したデータとは性質が異なる．高頻度で大容量であるだけでなく，材料ロット，板厚ばらつき，温度，潤滑状態，段取り条件，立上げ時と安定時の差など，多数の要因で分布が変動しやすい．金型のどこにセンサーを設け，1ストロークでどのような時系列データが得られるかを図1に示す．図1のように，1ストローク中の変形から材料解放後の戻りまでの挙動が連続的に記録される．

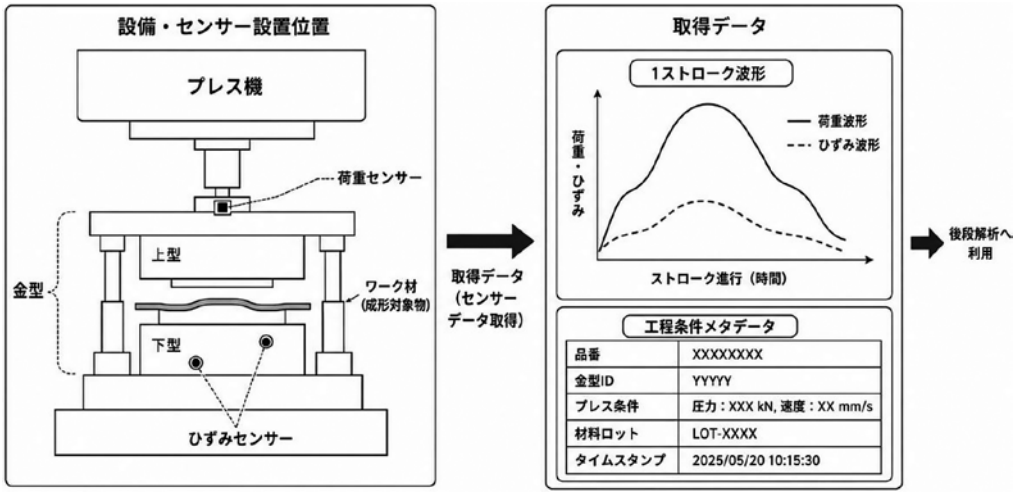

図 1　プレス成形ラインにおけるセンサー設置位置と取得データ

さらに，現場データには，ノイズ，基線ずれ，同期ずれ，ドリフトが含まれる．同じ工程であってもストロークごとに立上り時刻がわずかにずれるため，単純な時刻比較だけでは意味のある差が見えにくい．また，ひずみ波形では，センサーの取り付け位置や剛性差により，絶対値の比較が難しい場合もある．荷重波形でも，立上り前後の基線が変動すると，ピーク値や積分値の解釈に影響が出る．このため，解析の第一歩はモデル化ではなく，どのような前処理を施し，どの位相を基準に整列させるかを決めることである．これらの乱れが波形に与える影響を図 2 に示す．図 2 は，基準波形に対してノイズ重畳，基線ずれ，同期ずれ，ドリフトがそれぞれどのように波形を変化させるかを模式的に示したものである．

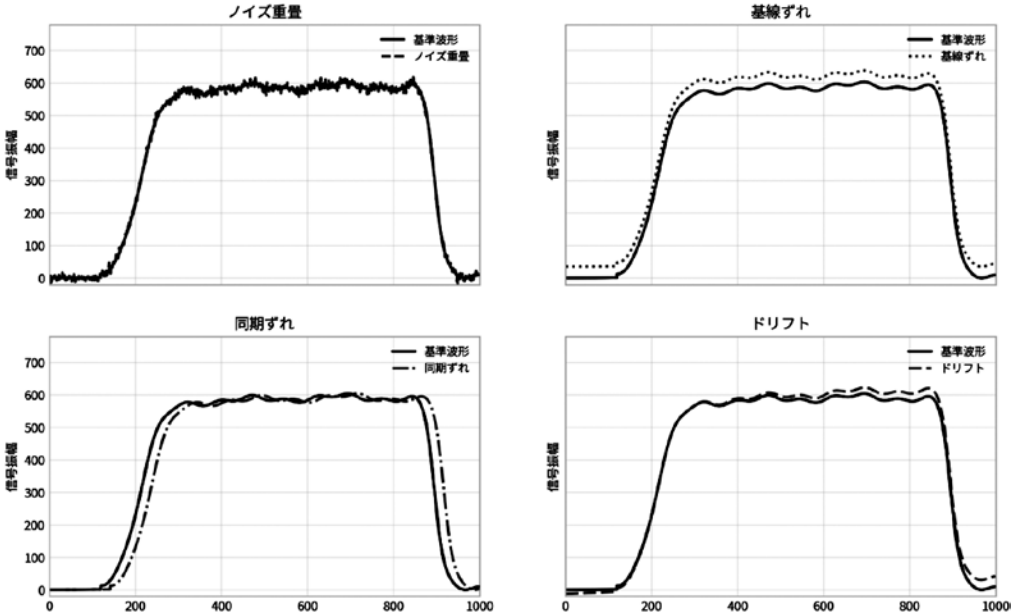

図 2　実生産データに含まれるノイズ，基線ずれ，同期ずれ，ドリフト

従来の解析では，まず波形を目視し，どの区間に注目すべきかを技術者が決めなければならなかった．そのうえで，区間分割，ピーク検出，立上り傾き算出，積分値計算，区間統計量算出，周波数帯エネルギー算出などの処理を個別に実装し，結果をグラフへ重ね合わせて評価する流れが一般的であった．加えて，しきい値設計や簡易判定ルールの作成も，過去の経験や試行錯誤に依存しやすい．これらの処理は，個別には難解でなくても，一連の作業を再現可能な形で維持するには手間を要していた．

この従来方式の課題は，第一に属人性の高さである．ある技術者が有効と判断した特徴量や区間定義が，明文化されないまま運用されると，異動や世代交代で作業が継続できなくなる．第二に，再現性不足である．同じ課題に対しても，分析者ごとに前処理の細部やグラフ表現が異なると，結論の比較が難しくなる．第三に，実装負荷とレビュー負荷である．現場技術者が見たい図を，その都度 IT 技術者がコード化し，解釈をすり合わせる作業は，開発初期には特に負荷が高い．第四に，非 IT 技術者にとっての心理的障壁である．これは，良い仮説を持っていたとしても，自分でコードを書けない場合は検証に参加しにくいといった心理面での課題である．

本検討で目指すのは，こうした断絶を小さくすることである．すなわち，自然言語で仮説や注目区間，見たいグラフ，確認したい観点を表現し，LLM がそれを前処理コード，特徴量抽出コード，可視化テンプレートへ変換する．その結果を人間がレビューし，物理的妥当性や運用妥当性を確認したうえで，プロンプト，コード，図表テンプレートを固定化する．この流れにより，非 IT 技術者にとっては解析の入口に立ちやすくなり，IT 技術者にとってはゼロからの実装と比べ作業負荷が軽減するため，レビューと整備に集中しやすくなる．

もっとも，この仕組みは，人間の判断を不要にするものではない．前処理の有無や位相整列の基準によって，本来捉えるべき差が見えなくなったり，反対に実体のない差が見かけ上現れたりするため，物理現象を踏まえた評価軸は不可欠である．したがって，LLM の導入価値は，判断の代行ではなく，仮説形成から再現可能な検証手順への橋渡しにある．3 章では，この位置付けを明確にするために，LLM による波形解析支援プロセスについて整理する．

3. LLM による波形解析支援プロセス

本章では，2 章で述べた課題に対し，波形解析の工程構成と，各工程における人間と LLM の役割分担について述べる．

3. 1 全体像

LLM による波形解析支援プロセスの狙いは，波形解析に求められる思考と実装を，自然言語を介して接続することである．従来は，現場技術者の頭の中にある着眼点が，口頭や断片的なメモとして共有されるにとどまり，そこからコードとグラフを作る段階で情報が欠落しやすかった．LLM を介在させることで，「どの現象を見たいのか」「どの区間を比較したいのか」「何を異常とみなしたいのか」を文章として整理しやすくなり，その文章から前処理や可視化の叩き台を素早く生成できる．図 3 に示した全体図の通り，このプロセスではデータ取り込みから運用展開までを九つの工程に分け，各工程で人間と LLM が役割を分担する．工程間の戻り矢印は，レビュー結果を踏まえた修正の反復を表す．本稿では，この一連の工程を概念図としてではなく，実務でたどる手順として位置付ける．

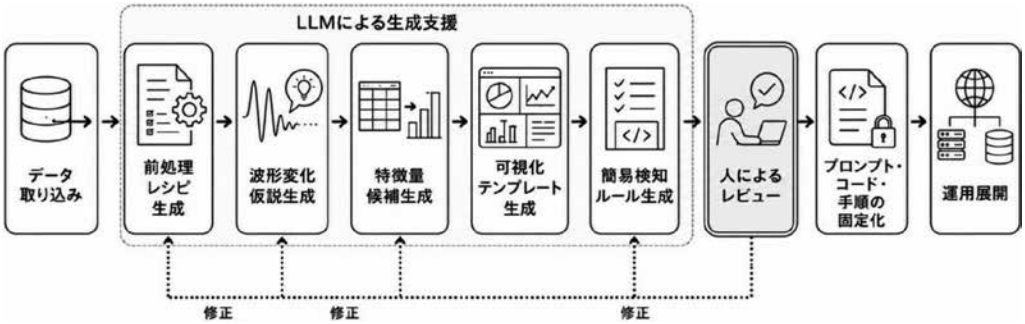

図 3　LLM による波形解析支援プロセス

3. 2　解析工程と LLM の支援内容

図 3 の九つの工程における人間と LLM の役割分担を表 1 に示す．以降の章では，各工程を表 1 の番号で参照する．

表 1　波形解析支援プロセスの九工程と役割分担

工程	人間が決めること	LLM が担う支援
①データ取り込み	必要な情報の範囲，1 ストロークの切り出し条件，保持すべきメタデータ	確認項目のチェックリスト生成
②前処理レシピの生成	各処理を入れる目的，補正の要否	前処理候補の列挙とコード断片への変換
③波形変化仮説の生成	採用する仮説の選定	仮説候補の整理，注目すべき変化点と比較区間の列挙
④特徴量候補の生成	説明しやすさと運用性による絞り込み	抽出式とグラフ化方法の提案
⑤可視化テンプレートの生成	比較軸と図の読み方の確定	凡例，注釈，色分け，レイアウトの整備
⑥簡易検知ルールの生成	しきい値と警報段階の決定	ルールの叩き台の文章化とコード化
⑦人によるレビュー	物理的妥当性，現場解釈，保全判断との整合	レビュー観点の整理，説明文の草案作成
⑧プロンプト・コード・手順の固定化	資産として残す範囲の判断	前処理条件，特徴量定義，レビュー観点の文書化
⑨運用展開	導入段階とフェイルセーフ設計	運用フロー図，説明資料，レビューシートの草案作成

工程②で重要なのは，前処理を単なる定型処理として流すのではなく，何のためにその処理を入れるのかを明示することである．たとえば，基線補正はピーク値の絶対比較を容易にするため，位相整列はストローク間で注目区間を同じ意味の時刻に揃えるため，平滑化は高周波ノイズに起因する疑似的な変化点を抑えるために行う．

工程④では，特徴量を増やし過ぎず，説明しやすく点検誘導や警報に結び付けやすいものを優先すべきである．

工程⑤の可視化テンプレートが固定化されると，分析者が変わっても同じ見え方で議論でき

るため，再現性だけでなくレビュー効率も高まる．

工程⑧では，LLM の出力をその場限りの試作で終わらせず，再利用できる形で保存する．自然言語で残されたプロンプトは，暗黙知を形式知へ変換する役割も果たす．

3. 3　人間が担う判断領域

九工程のうち，工程⑦のレビューは本プロセスの成否を左右する位置にある．LLM が生成した特徴量が，統計的には差を示していても，工程上の意味が薄い場合は採用すべきでない．反対に，わずかな差でも，現場では重要な兆候である場合がある．この選別は人間の役割である．また，レビュー時に「なぜその区間を見るのか」「なぜその特徴量を採るのか」を説明できることが，現場導入の条件となる．

また，工程⑨の運用展開においても，完全自動判定を急ぐよりも，段階的警報や点検誘導など，人間が介在する形で導入する方が安全である．すなわち，LLM の役割は，発想支援，手順化支援，実装支援，可視化支援，説明文生成支援，レビュー観点整理支援に分かれるのに対して，最終判断，物理的妥当性確認，フェイルセーフ設計は，人間が担うべきである．

3. 4　本プロセスを事例に適用する観点

以降の章では，本プロセスを性質の異なる二つの事例へ適用し，工程ごとに支援の効き方がどう変わるかを見る．4 章では，連続的・累積的に進行するピアスパンチ摩耗予測を扱い，工程②から⑨までがどのように連結したかを述べる．5 章では，突発的・条件依存的な絞り加工ワレ予兆を扱い，同じ工程列のどこで支援が有効に働き，どこで限界が現れたかを述べる．6 章では，両事例を工程ごとに対比する．

この対比で着目するのは，工程③の波形変化仮説の生成と工程④の特徴量候補の生成，すなわち「どの位相の何を見るか」を決める工程である．この二工程を人間の側で明確に定義できるかどうかが，工程⑤以降を LLM へ委ねられる範囲を決めると考えられる．

4.　事例研究①：ピアスパンチ摩耗予測

本章では，一つ目の事例として，ピアス工程でのパンチ摩耗予測への LLM 適用について述べる．

4. 1　背景

ピアス工程では，板材に穴を開けるためにパンチとダイが繰り返し作用する．このとき，パンチ先端の摩耗やクリアランスの変化が進むと，加工荷重の立上り方や破断に至るまでの応答が徐々に変化する可能性がある．摩耗が進んだ状態を見逃すと，穴品質の悪化やバリ増大，さらには設備停止や突発交換の原因となり得る．一方で，交換時期を早め過ぎれば，まだ使用できる部品を交換することになり，保全コストが増える．したがって，現場では，過剰保全と事後保全の間で適切な判断支援が求められる．

この判断を経験だけに頼った場合，技能差によるばらつきを避けることは困難である．摩耗は突発破壊ではなく，連続的に進む現象であるため，波形上にも単発ではなく緩やかな変化として現れることが期待される．この種の現象は，連続・累積的な傾向を捉える特徴量と相性が良い．そこで本事例では，ひずみ・荷重波形を 1 ストローク単位で比較し，どの区間に摩耗進

行の兆候が表れるかを検討した.

4. 2　波形の捉え方

ピアス工程の 1 ストロークは，概念的には「変形」「せん断」「破断」の区間に分けて捉えることができる．変形区間では，材料が弾性変形から塑性変形へ移る過程が主であり，波形の立上りが現れる．せん断区間では，材料除去へ向かう負荷が支配的となる．破断区間では，材料が分離することで応答が急激に変化する．このように区間を意味付けすることで，単一のピーク値を見るよりも，現象に沿った比較が行える．この区間分割の考え方を図 4 に示す．図 4 のように，立上りが現れる変形区間，材料除去へ向かう負荷が支配的なせん断区間，応答が急変する破断区間に分けて 1 ストロークを捉える．

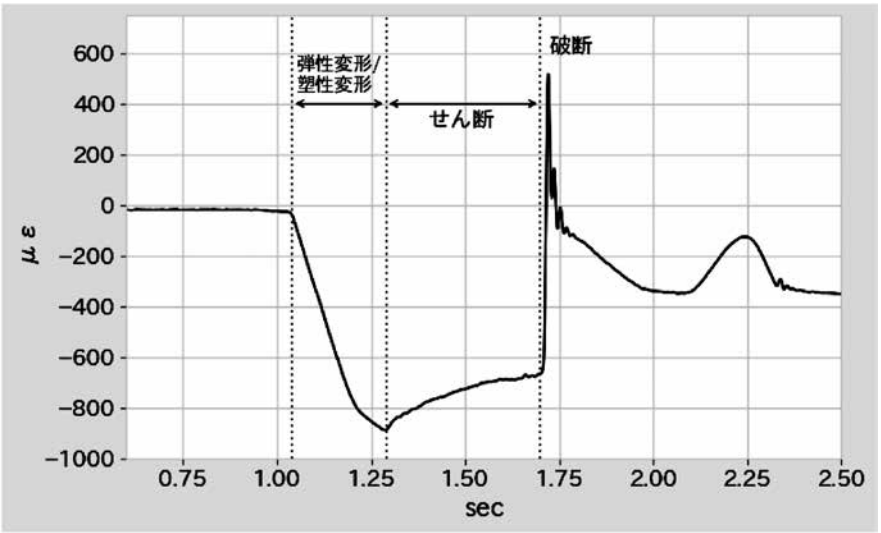

図 4　ピアス工程における 1 ストロークの区間分割

ここで重要なのは，絶対時刻で比較するのではなく，立上り開始や特定イベントを基準に位相整列することである．ストローク間でイベント時刻にわずかなずれがある場合，整列前の比較では，実際には同じ現象を別の時刻として扱ってしまい，特徴量のばらつきが実態以上に増える．

位相整列後に区間ごとの特徴量を見ることで，摩耗進行がどの区間に現れやすいかを整理しやすくなる．たとえば，ピーク値だけでなく，立上り傾き，せん断区間の平均値，破断完了までの時間差などを候補として並べると，現場技術者は，各特徴量が加工現象のどの変化に対応しているかを検討し，どの指標であれば工程の言葉で説明できるかを議論しやすい．たとえば，破断完了までの時間差であれば，クリアランス変化に伴って材料が分離するまでの応答が変わったものとして説明できる．この説明可能性は，後の運用で警報や点検誘導を受け入れてもらうために重要である．

4. 3　LLM による仮説生成

本節は，3 章の工程③（波形変化仮説の生成）および工程④（特徴量候補の生成）に対応する．

LLM への入力は，特定の特徴量を指定しない工程課題そのものとした．具体的には，「パンチの摩耗，劣化を予測し，工具の交換時期を最適化するための方策を教えてください」という問いである．波形上の着目点や，摩耗がどの機序で波形へ現れるかは，この時点では与えていない．

得られた出力は，センシング，特徴量設計，判定という段階に整理された方策の一覧であった．このうち特徴量設計の段階では，荷重から導かれる指標としてピーク荷重，破断完了までの時間，せん断エネルギー（荷重の時間積分）が挙げられ，あわせて打音の実効値や高周波成分比といった振動系指標，バリ高さなどの品質系指標，累積ストローク数といったカウンタ系指標が候補として列挙された．

このうち，人間の検討にとって有効であったのは，破断完了までの時間である．荷重開始点と破断点の時間差に着目する仮説は，人間が事前に提示したものではなく，工程課題を与えた問いに対して LLM が分析仮説の候補として示したものである．人間はこの候補に対し，パンチ摩耗が進むとパンチとダイの見かけ上のクリアランスや切れ味が変化し，材料が分離するまでの応答時間が変わり得るという物理的解釈を与えられることを確認したうえで，工程④の絞り込みとして分析対象に採用した．一方，振動系や品質系の指標は，本事例で取得しているデータの範囲では検証できないため採用していない．

その後，「荷重の立上り開始から急峻な荷重低下が完了するまでの時間を測り，ストローク間で並べて比較したい」と自然言語で追加の指示を伝えることで，LLM は荷重開始点と破断点の定義，イベント検出のしきい値設定や注釈付きグラフの草案を返した．この考え方は，摩耗が徐々に進むなら，応答の変化も連続的に進むという人間側の直感と整合していた．このように，LLM は採用された仮説をもとに，イベント検出コードや注釈付きグラフの草案を生成し，候補特徴量の比較を進めるうえでの初速を高めた．

ここで重要なのは，LLM が自ら物理機構を理解して真の特徴量を発見したのではないという点である．出力された候補は，工具摩耗の監視において一般に知られる指標の列挙であり，そのうちどれが本工程のデータで意味を持つかまでは示されていない．LLM の価値は，候補を幅広く並べたうえで，人間が選んだ仮説を具体的な処理手順，抽出コード，可視化案へ素早く展開できる点にある．この役割分担が明確であるため，生成された結果の妥当性もレビューしやすい．摩耗進行に伴う破断完了時刻の変化を図 5 に示す．図 5 は，新品に近い状態と摩耗進行状態とで，破断完了までの時間差 Δt が生じる様子を模式的に表したものである．

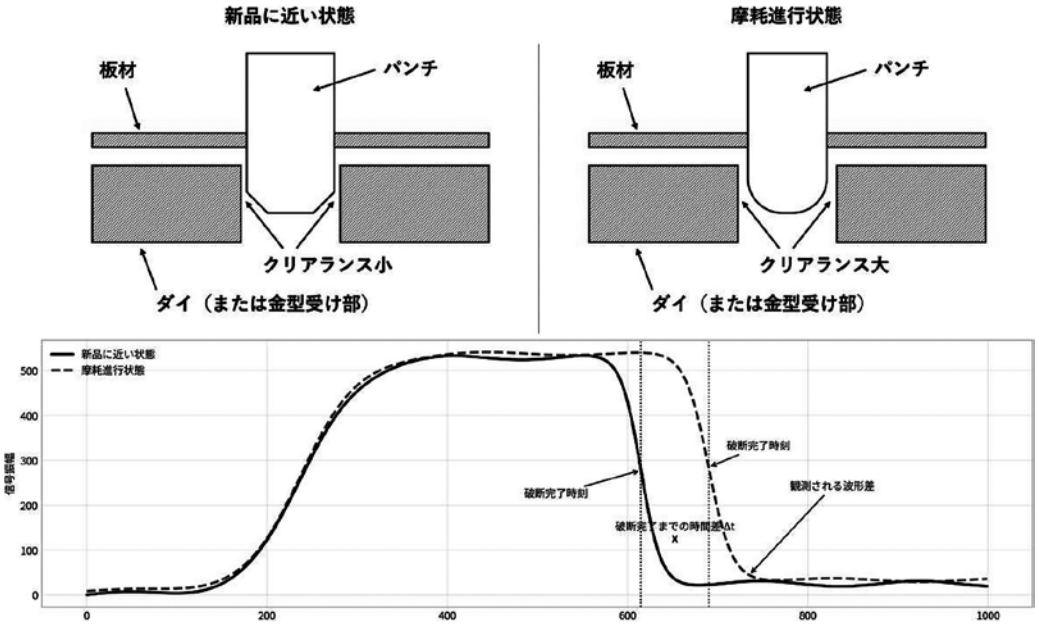

図 5　パンチ摩耗に伴う破断時間変化

4. 4　前処理・特徴量抽出・可視化

本節は，3 章の工程②（前処理レシピの生成），工程④（特徴量候補の生成），工程⑤（可視化テンプレートの生成）に対応する．

実務上は，自然言語の指示から LLM に前処理と可視化コードの草案を生成させる形が有効であった．たとえば「摩耗の進行を見たい」「破断完了までの時間差を比較したい」「区間を注釈付きで重ね描きしたい」といった指示である．これにより，最初のグラフが出るまでの時間が短くなり，現場技術者は早い段階で波形を確認できる．波形が期待と異なれば，どの前処理が過剰か，どの整列基準が妥当かを人間が修正し，再度 LLM へ指示することで反復を進められる．

この反復過程で重要なのは，毎回異なるコードを一から生成して終わらせないことである．有効と判断した前処理条件，イベント定義，特徴量の式，グラフレイアウトは，工程⑧としてプロンプトとコードの双方で固定化すべきである．これは，LLM 活用の本質を「自動化」より「再現化」に置く考え方である．

可視化については，区間注釈付きグラフが特に有効であった．工程区間を注釈した波形と，特徴量の時系列トレンドを並べて示すことで，現場技術者は単なる数値変化ではなく，「どの加工フェーズで何が変わっているのか」を把握しやすい．この形式は，異常判定の根拠説明にも適しており，後の点検判断にも接続しやすい．LLM は，注釈位置，凡例，色分け，複数グラフのレイアウトを揃えるうえでも有効であった．あわせて，図の説明文や特徴量の意義の草案を自然言語で併記できるという点で，コードを読まない関係者でも内容を理解しやすい．

4. 5　モデル化と運用

本節は，3 章の工程⑥（簡易検知ルールの生成），工程⑦（人によるレビュー），工程⑨（運

用展開）に対応する．

本事例では，交換判断を支援するという観点から，分類問題として扱う考え方が取りやすい．破断完了までの時間に関わる指標は，変化方向が解釈しやすく，現場での説明責任を果たしやすい．

また，摩耗の進行が連続的であるという前提に立つと，単調性制約を持つ考え方は有用である．すなわち，摩耗が進むほど指標がある方向へ変化するという制約を置くことで，判定結果の解釈一貫性を保ちやすい．

さらに，区間境界の微調整や区間再定義により，特徴量のロバスト性を高められる場合がある．LLM は，境界条件の違いによる可視化比較を素早く試すのに役立つが，最終的にどの定義を採用するかは，人間が工程の意味と運用性を踏まえて決めなければならない．現場運用では，単一の自動判定で交換を決めるのではなく，注意喚起，点検誘導，人間による確認，最終交換判断を段階的に実施するフェイルセーフ設計が現実的である．こうした段階的な運用の流れを図 6 に示す．図 6 は，特徴量の抽出結果を部品交換の判断候補として提示した後，注意喚起と点検誘導を経て人間が確認し，最終的な交換判断を行う運用の流れを示している．部品交換は不要と判断された場合には継続運転と経過観察に戻ることで，過剰保全と事後保全のいずれにも偏らない判断を支える構成としている．

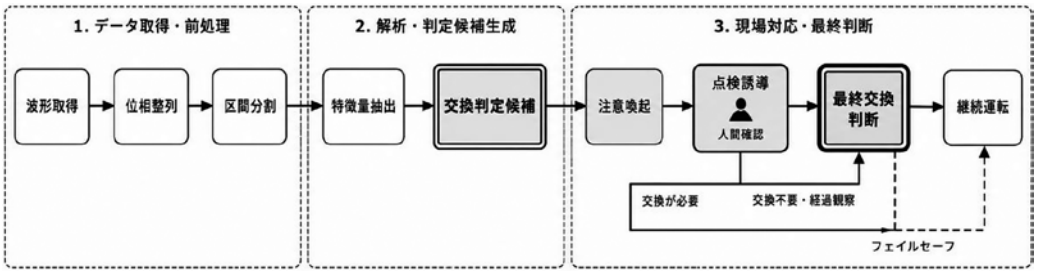

図 6　ピアスパンチ摩耗予測の運用フローと段階的警報

この意味で，本事例における LLM の価値は，モデル性能の最大化そのものより，仮説から説明文までを一貫した手順として整備し，現場判断を支える資料を素早く用意できる点にある．

5.　事例研究②：絞り加工におけるワレ予兆

本章では，二つ目の事例として，絞り加工工程でのワレ予兆検出への LLM 適用について述べる．

5. 1　背景

絞り加工におけるワレは，製品品質に直接影響する重大な不具合である．一度ワレが発生すると，当該製品の不良化だけでなく，後工程での手戻りや選別負荷の増加につながる．また，ワレが発生する条件は，材料状態，潤滑，しわ押さえ条件，金型状態など，複数要因の組み合わせに依存しやすく，原因の切り分けも容易ではない．このため，ワレが発生した後に要因分析を行うだけでなく，加工中の波形から予兆を捉えられるかどうかは，現場で大きな関心を集めている．

しかし，ワレ予兆検出は，4 章の摩耗予測とは性質が大きく異なる．摩耗予測では，劣化が徐々に進むため，ストロークをまたいで比較したときに，ある方向への変化を捉えやすかった．これに対し，ワレは突発的であり，同じ条件に見える加工でも，ある境界で急に発生する場合がある．また，波形全体の差よりも，最大荷重直後や破断に近い特定位相の局所的な形状変化が本質的な要因である可能性もある．したがって，「どこを見るべきか」を外すと，いくら多くの特徴量を計算しても，肝心の兆候を捉えにくい．

本事例は，LLM の弱点を示すだけではなく，人間の現象理解がどの段階で求められるかを考えるうえで重要である．以下では，まず LLM が有効であった点を整理し，次に苦手であった点と，指示で改善できた範囲，それでも残る本質的な課題を順に述べる．3 章の工程との対応は，各節の冒頭に示す．

5. 2　LLM が有効だった点

本節で述べる支援は，工程②（前処理レシピの生成），工程④（特徴量候補の生成）のうち候補の列挙，工程⑤（可視化テンプレートの生成），および工程⑦（人によるレビュー）の観点整理にあたる．

ワレ予兆の解析でも，LLM は初期処理の立上げを大きく支援した．具体的には，基線補正，平滑化，位相整列，グラフ描画の叩き台を短時間で生成できる点が有効であった．実生産波形では，基線のわずかなずれやノイズが，局所的な傾き変化の観察を妨げる．このため，どの程度平滑化するか，どの基準で波形を揃えるか，正常波形とワレ発生波形をどう重ねるかといった初歩的だが重要な処理を素早く回せることは，大きな利点である．

また，区間統計量，ピーク値，立上り傾き，積分値，位相差分など，時系列解析で一般に候補となる特徴量を列挙し，それぞれの抽出コードや比較図を用意する点でも LLM は役立った．これにより，分析者は「何が使えそうか」をゼロから考えるのではなく，候補一覧を俯瞰しながら，どれが現象理解に合うかを選別できる．特に，プロットの体裁を揃えたうえで正常・異常を比較できることは，試行錯誤の速度を高めた．

さらに，LLM は説明文やレビュー観点の整理にも有効であった．「最大荷重近傍を拡大して比較する」「正常・ワレ波形の差がどの区間で顕著かをコメントする」といった指示を，図の作成とコメント草案へ展開できる点は，複数人でレビューする際に有用であった．

5. 3　LLM が苦手だった点

本節で問題となるのは，工程③（波形変化仮説の生成）と工程④（特徴量候補の生成）における重要位相の同定，すなわち 3. 4 節で述べた「どの位相の何を見るか」を決める工程である．

一方で，本事例の核心は，LLM が一般的な前処理やコード生成を得意としても，現象特有の「形の変化」を自律的に捉えることは難しかった点にある．「最大荷重直後の早期下降」のような局所形状は，人間がグラフを見れば違和感として把握しやすい．しかし，LLM へ一般的な異常点抽出を依頼しただけでは，この部分を必ずしも優先して特徴量化してくれない．むしろ，末期の減衰や全体形状の違いに引きずられ，重要位相を外す場合があった．

典型的には，正常データの減衰末期を「下降異常」とみなしてしまう場合があった．これは，傾きの大きい箇所を機械的に拾うと，工程上あまり重要でない終盤の変化を過大評価するためである．反対に，ワレ発生データでは，最大荷重直後から破断にかけての区間こそ重要である

にもかかわらず，そこを外して，より後ろの時刻に注目してしまうことがあった．このような取り違えは，グラフ上では明らかであっても，現象理解を持たないまま汎用的な特徴量探索を行うと起きやすい．この取り違えの例と再指示後の改善を図 7 に示す．図 7 では，重要位相を外した抽出と，注目範囲を明示して再指示した後の抽出とを対比しており，同じデータでも着目点の与え方によって結果が変わることがわかる．

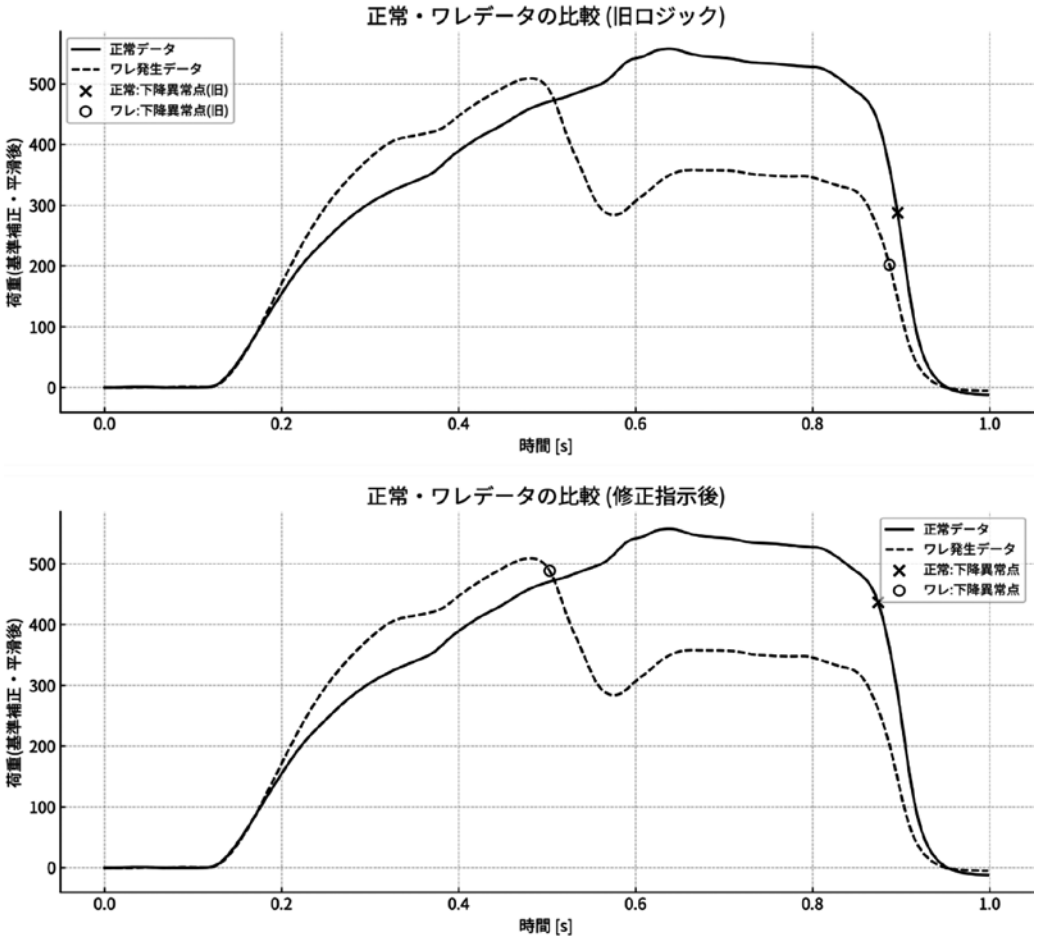

図 7　LLM の見落としと再指示後の改善

ここで問題となるのは，LLM がコード生成や可視化を得意とすることと，物理機構理解に基づいて重要イベントを同定することが別の能力である点である．前者は，指示された処理を整った形で実装する能力であり，後者は，どの時間帯に何が起きているかを工程知識と結び付けて判断する能力である．本事例では，まさに後者が成否を分ける要因となった．

5. 4　指示による改善とその限界

本節の説明は，工程③と工程④を人間が先に確定させたうえで，特徴量の抽出と工程⑤の可視化を LLM へ委ねた場合にあたる．

もっとも，LLM が全く改善できなかったわけではない．最大荷重前後の時間窓を明示し，

その範囲で傾きのしきい値，曲率の符号反転点，注目範囲を指定して再指示すると，所望に近い抽出や図示は可能であった．たとえば，「最大荷重直後の一定範囲で最初に強い下降が現れる点を候補とする」「下降開始と曲率変化を併記する」といった条件を与えると，図示結果は人間の見方に近づいた．

このことは，人間が注目位相を先に定めさえすれば，工程④の抽出と工程⑤の可視化については LLM が依然として有効であることを示している．

ただし，この改善は本質的には「人間が答えに近い注目範囲を先に与えている」ことに依存する．したがって，これは自律検出の達成ではなく，指示の精緻化による半自動化である．もし別の条件や別の不具合に対して同じ処理を適用する場合，どの範囲を指定すべきかを再び人間が考えなければならない．この点は，摩耗予測と比べて，テンプレート化だけでは吸収しにくい難しさである．

5.5　本質的な課題

本質的な課題は工程③に残る．すなわち，ワレの発生タイミングをデータから正しく推定し，その理解に基づいた特徴量を設計することが困難なため，工程⑥以降の検知ルール化と運用展開まですすめられなかった．単に波形全体の差を見つけるだけでは不十分であり，最大荷重，下降開始，破断候補時刻，その間の変化率や曲率変化などを，イベントとして扱うべきである．つまり，ワレ予兆では「どの位相で何が起きるか」を捉えるイベント中心の見方が重要である．

この問題は，LLM 単体では解きにくい．少なくとも本事例で試した範囲では，外部知識や明示的なアルゴリズムを与えない状態の LLM は，どの微分次数を見るべきか，どの変化点が物理的に重要かを安定して判断しにくかった．確立したアルゴリズムとの組み合わせによる対処については，7.1 節で述べる．

本事例が示すのは，LLM の失敗そのものより，解析対象の性質によって求められる支援形態が変わるという事実である．この点は，6 章で摩耗予測と比較して整理する．

6.　成功例と課題例の比較・考察

4 章と 5 章で述べた二つの事例を，3 章で述べた九つの工程に沿って比較した結果を表 2 に示す．支援の効き方は，全工程で一様に異なるわけではない．工程①，②，⑤では両事例に大きな差がなく，差は工程③および工程④に集中している．そして，この差が工程⑥以降の到達点と，工程⑦におけるレビュー負荷の違いにつながった．

なぜ工程③と工程④で差が生じたのかについて，本稿の二事例を比較した範囲では，次の三点が影響したと考えられる．

第一に，現象の性質である．摩耗は累積的に進むため，複数ストロークを時系列に並べれば変化の方向が現れる．したがって，仮説の当否をデータ全体で検証でき，仮説が外れていれば比較の過程で気付ける．一方，ワレは特定のストロークで突発的に生じるため，比較すべき対象が「ワレが生じたストロークの内部のどの位相か」となり，検証の単位が波形内部の局所へ移る．この場合，仮説が外れていること自体に気付きにくい．

第二に，特徴量設計の容易さである．摩耗予測では，変形・せん断・破断という区間が工程物理から定まるため，特徴量設計を「区間の切り方」と「区間内の統計量」の組合せへ分解で

表 2　工程別に見た摩耗予測とワレ予兆の比較

工程	ピアスパンチ摩耗予測	絞り加工ワレ予兆
①データ取り込み	切り出し条件の整理から開始．両事例で差はない	同左
②前処理レシピの生成	位相整列・基線補正の候補生成が有効	同様に有効．初速向上に寄与
③波形変化仮説の生成	工程課題の問いから破断完了時間の候補が得られた	重要位相を外し，減衰末期を過大評価した
④特徴量候補の生成	単調に変化する指標へ絞り込めた	候補列挙は可能だが核心の局所形状に届かない
⑤可視化テンプレートの生成	区間注釈付きグラフが定着	正常・ワレ比較図の体裁統一に有効
⑥簡易検知ルールの生成	単調性を前提とした判定の考え方まで整理	着目位相が未確定のため未到達
⑦人によるレビュー	区間定義と妥当性の確認が中心	注目位相と評価観点の設計自体が必要
⑧プロンプト・コード・手順の固定化	前処理・特徴量・図を資産化できた	条件が変わると再設計が必要で資産化しにくい
⑨運用展開	段階的警報を含む運用フローを設計できた	追加検証が必要

きた．この形であれば，LLM が得意とする一般的な候補列挙がそのまま使える．これに対しワレ予兆で着目すべきものは，最大荷重直後の下降の速さという局所形状であり，区間統計量の組合せとしては表現しにくい．汎用的な候補列挙が核心に届かなかったのは，このためである．

第三に，評価観点の定義しやすさである．摩耗予測では，摩耗が進むほど指標が一方向へ動くという単調性を評価基準に置けるため，特徴量の良否をデータから判定できた．ワレ予兆では，ワレの発生タイミングそのものが推定対象であり，正解とすべき時刻が確定しない．このため，抽出結果が妥当かどうかは人間がグラフを見て判断するほかなく，工程⑦の負荷が増し，反復の回転も遅くなった．

これら三点は独立しているのではなく，現象の性質が特徴量設計を分解できるかどうかを決め，それが評価観点を定義できるかどうかを決めるという連鎖関係にある．LLM の支援が有効に働くかどうかは，この連鎖の起点である現象の性質に強く依存すると考えられる．

一方で，LLM が両事例に共通して効果を発揮した点もある．それは解析作業の初速向上と再現性向上である．自然言語で仮説や処理要求を与えると，前処理コード，可視化コード，説明文を短時間で生成できるため，現場技術者と IT 技術者の間にある実装の壁を下げられる．特に，プロンプト，コード，図の形式を固定化できる点は，解析業務を属人的な試作から再利用可能な資産へ変えるうえで重要である．

したがって，現時点の実用解は，LLM に完全自律検出を期待することではない．人間が現象理解に基づいて注目範囲と評価観点を設計し，妥当性を確認する一方，LLM は暗黙知を形式知へ変換し，手順化・可視化・再現化を通じてレビューできる形へ整える．この構図であれ

ば，LLM の長所を活かしつつ，誤った自律判断のリスクを抑えられる．LLM が有効な作業と人間によるレビューを要する作業の対応を，本稿の事例で該当した場面とあわせて表3に示す．

表 3　LLM が有効な作業と人間によるレビューを要する作業

項目	LLM が有効な作業	人間によるレビューを要する作業	本稿の二事例で該当した場面
仮説整理	候補列挙，観点整理，説明文草案	採用仮説の選定，工程整合確認	摩耗：候補から破断完了時間を選定 ワレ：重要位相の取り違え
前処理	コード生成，処理候補比較	過補正の有無，整列基準の妥当性判断	両事例：位相整列と基線補正の基準決定
特徴量設計	汎用候補の提案，抽出実装	物理的意味付け，採否判断	摩耗：振動系・品質系の候補を不採用と判断
可視化	グラフ作成，注釈整備，テンプレート化	図の読み方，重要位相の確認	ワレ：注目範囲の明示により抽出が改善
運用化	手順書草案，レビューシート草案	警報設計，フェイルセーフ，現場導入判断	摩耗：段階的警報を含む運用フローの設計

今後，LLM がアルゴリズムや知識ベースと連携することで，より高度な自律支援へ近づく余地はある．ただし，解析対象の現象差に応じて，人間がどの観点を設計し，どこまで機械へ委ねるかを明示することは，引き続き実運用に耐える仕組みの条件となる．

7.　今後の課題と展望

本章では，6 章の比較・考察結果を踏まえ，今後の LLM の適用の方向性と評価観点について述べる．

7.1　LLM ×アルゴリズムの連携

6 章で述べたとおり，本プロセスの難所は工程③と工程④，すなわち「どの位相の何を見るか」を決める段階にあった．3.1 節で示した図 3 のプロセスでは，この段階を人間と LLM の対話によって決めている．しかし 5 章のワレ予兆のように，人間が注目位相を先に与えなければ進まない場合，候補の絞り込みが人間の当たりの付け方に依存し，別の不具合へ適用するたびに同じ検討を繰り返すことになる．

そこで，今後の第一の方向性は，LLM を「説明可能な司令塔」と位置付け，実際の数値処理は確立されたアルゴリズムへ委ねる構成である．この連携の構成を図 8 に示す．図 8 の「3. LLM による説明支援」に示したように，LLM は仮説生成，手法選択，結果解釈，再探索方針の提示を担い，「2. アルゴリズム解析」に示した変化点検出，微分・曲率，包絡線，位相整理，クラスタリング，異常検知などの計算は専用アルゴリズムが担う．この分担により，LLM の言語能力と，数値処理の安定性を両立しやすくなる．

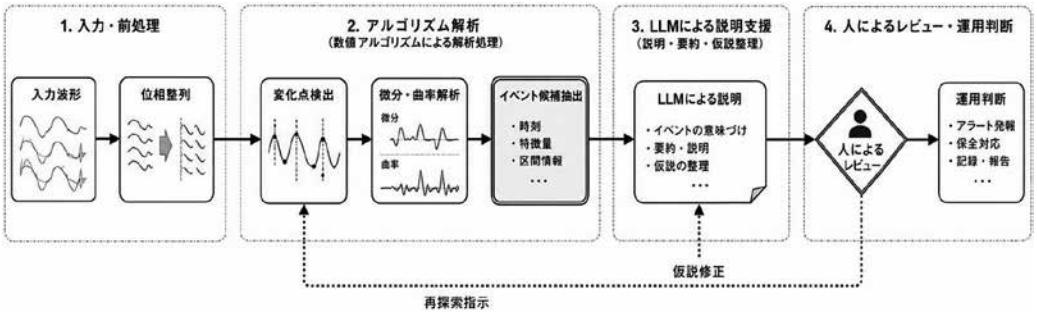

図 8　LLM ×アルゴリズム連携によるイベント抽出フロー

図 3 のプロセスとの違いは，工程③と工程④の入口を数値処理へ移す点にある．図 3 では特徴量候補を言語的な列挙から出発させていたのに対し，図 8 では変化点検出などのアルゴリズムが候補イベントを先に提示し，LLM はその結果を説明して次に見るべき箇所を提案する役割へ移る．これにより，人間が注目位相を事前に特定できない場合でも，候補の提示までは機械側で担える構成となる．

特にワレ予兆では，出力を単なる異常スコアではなく，イベント時系列として整理することが重要である．たとえば，最大荷重時刻，破断候補時刻，複数の変化点，それぞれの信頼度，根拠となる波形特徴をイベントとして明示すれば，人間はどこに着目して判断すべきかを理解しやすくなる．LLM は，これらのイベントの説明と相互関係の整理に適している．この形であれば，完全自動判定よりも，レビューと意思決定を支える説明可能な支援へ近づく．

7. 2　RAG と手順テンプレート

第二の方向性は，RAG（Retrieval-Augmented Generation）と手順テンプレートの整備である．RAG は，LLM が外部知識ベースを検索し，回答や提案に参照情報を取り込む構成である．なお，本稿の二事例では RAG を用いた検証は実施しておらず，以下は 6 章で明らかになった課題に対する今後の方向性として述べる．

RAG に期待するのは，工程③の課題への対処である．4. 3 節で述べたように，外部知識を与えない LLM の出力は，工具摩耗の監視で一般に知られる指標の列挙にとどまり，そのうちどれが当該工程で意味を持つかは示されなかった．また 5. 3 節では，工程物理を参照しないまま汎用的な特徴量探索を行った結果，重要位相を外している．成形理論，作業標準，過去解析事例，トラブル事例，保全ルールを外部知識として参照させることで，こうした一般論への偏りを減らし，工程物理と整合した候補を提示しやすくなると考えられる．この知識参照と資産化の流れを図 9 に示す．図 9 のように，LLM が RAG によって外部知識を参照して工程物理と整合した手順を生成し，得られたプロンプト，コード，テンプレートを再利用可能な資産として蓄積する．

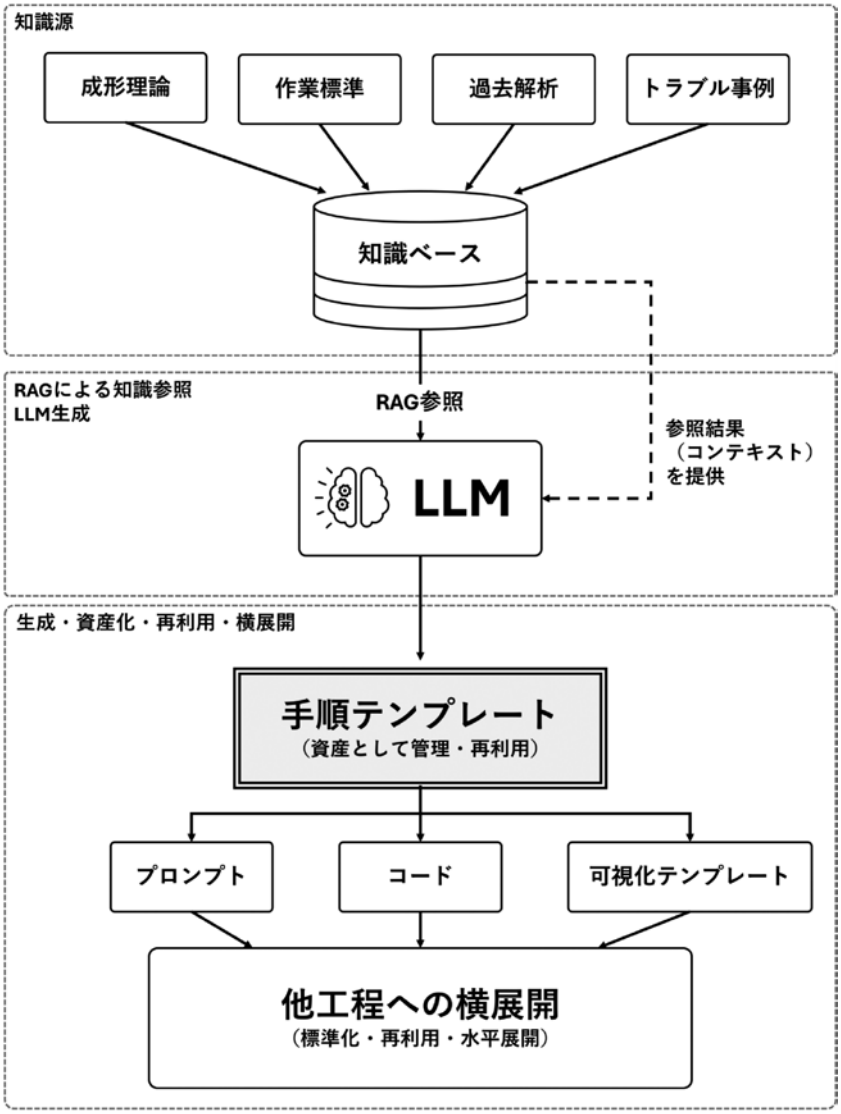

図 9　RAG と手順テンプレートによる知識参照・資産化フロー

一方，手順テンプレートの資産化が対応するのは，6 章の表 2 の工程⑧に示した課題である．ワレ予兆では，条件が変わるたびに注目範囲を設計し直す必要があり，成果が資産として蓄積しにくかった．テンプレートには，イベント定義，注目区間定義，候補特徴量，可視化方法，検証観点，誤検知・見逃し確認，現場レビュー観点を含める．テンプレート化の利点は，分析者が変わっても最低限見るべき項目がぶれにくい点にある．プロンプト，コード，判断基準をまとめて資産化すれば，他工程への横展開や技術継承にもつながる．

7. 3　実運用に向けた評価観点

実運用を見据えると，評価指標は単一では足りない．表 4 に，今後整理すべき主な評価観点を示す．検知精度だけでなく，発生タイミング誤差，リードタイム，誤検知，見逃し，説明可能性，現場レビュー負荷，再現性，メンテナンス性，他工程への横展開性といった観点を併せ

て評価すべきである．

表 4　実運用に向けた評価観点

評価観点	ねらい	本稿での評価状況
発生タイミング誤差	重要イベントをどれだけ正しい位相で捉えられるかを確認する	今後の課題（正解時刻が定まらず未評価）
リードタイム	異常が顕在化する前にどれだけ早く兆候を捉えられるかを確認する	今後の課題
誤検知	不要な警報が現場負荷を増やさないかを確認する	部分的（5.3 節で誤抽出の事例を確認）
見逃し	重大不具合を取り逃さないかを確認する	部分的（5.3 節で重要位相の見落としを確認）
説明可能性	判定根拠を現場へ説明できるかを確認する	定性的に確認（4.2 節，4.3 節）
現場レビュー負荷	導入後の確認作業が過大でないかを確認する	定性的に確認（6 章の工程⑦の対比）
再現性	分析者や時期が変わっても同じ手順が再現できるかを確認する	部分的（摩耗予測は工程⑧により向上）
メンテナンス性	条件変更時にテンプレートやルールを更新しやすいかを確認する	今後の課題
横展開性	他工程や類似不具合へ適用しやすいかを確認する	今後の課題

これらの観点を満たすためには，モデル単体の性能だけでなく，運用フロー全体を設計することが求められる．すなわち，どのイベントを誰が確認し，どの段階で点検や条件見直しにつなげるかを含めて設計しなければならない．LLM は，この運用設計を文書化し，レビューできる手順へ整える支援役としても活用できる．

8. お わ り に

本稿では，実生産ラインの金型センサーデータ解析に対し，LLM を解析支援に用いる枠組みを，ピアスパンチ摩耗予測と絞り加工ワレ予兆の二つの事例で検討した．

二事例の比較から得られた最も重要な知見は，LLM を自律判断の代替として位置付けるのではなく，解析プロセスの再現化と標準化を担う支援役として位置付けるべきである，という点である．本稿の範囲では，LLM が工程物理を理解して重要位相を自ら同定することはなかった．摩耗予測で有効であった特徴量も，LLM が列挙した一般的な候補の中から，人間が工程上の解釈可能性を根拠に選び取ったものである．一方，選ばれた仮説を前処理，特徴量抽出，可視化，説明文へ展開し，それらを固定化して再現可能な手順へ整える作業では，LLM は両事例を通じて一貫して有効であった．すなわち，LLM の価値は判断の自動化ではなく，判断に至る過程の標準化にある．

この位置付けに立てば，特徴量設計，可視化，コード化，検証手順の標準化を進めやすくなり，非 IT 技術者でも解析に参加しやすくなる．固定化したプロンプトやテンプレートは，再

現性に加えて技術継承の足場となる．

ただし，支援がどこまで届くかは対象現象の性質に依存する．ピアスパンチ摩耗予測では，連続的な劣化現象を区間化して捉える枠組みと親和性が高く，LLM は仮説整理と実装支援に有効であった．これに対し，絞り加工ワレ予兆では，発生位相の自律的特定が依然として課題であり，現時点では人間が注目範囲と評価観点を設計し，LLM が実装，可視化，再現化を担う役割分担が実用的である．

今後は，LLM ×アルゴリズム連携，RAG，手順テンプレートを組み合わせ，説明可能で再現性のある異常検知支援へ発展させることが重要である．その際には，数値性能だけでなく，発生タイミング誤差，レビュー負荷，フェイルセーフ，横展開性まで含めた実運用評価が求められる．本稿で示した整理が，生成 AI を過度に誇張することなく，現場に根差した形で活用していくための一助となれば幸いである．

執筆者紹介　長　島　正　貴（Masaki Nagashima）
2010 年ユニアデックス（株）入社．Unisys 系メインフレームの基本ソフトウェア保守業務に従事する．2018 年より岐阜大学 地域連携スマート金型研究センターに参加．主にプレス機械のセンサーデータ分析と成形品の品質判定についての研究に取り組む．

農業経営支援サイトAgriweBにおける生成AI×RAGのPoCと本番導入の実践報告

A Practical Report on PoC and Production Deployment of Generative AI × RAG for AgriweB

齋藤広樹

要　約　本稿では，農業経営支援サイトAgriweBに生成AIとRAGを適用した対話型AI「経営アシストAI」について，PoCから本番導入までの設計と実装を報告する．曖昧な相談を起点に論点を整理し，出典提示付きで専門コンテンツへ誘導する対話設計（追加質問・類似質問を含む）により，従来手法では難しい課題の具体化と深掘りを支援した．PoCで有効性と運用成立性を確認し，本番では認証連携とセキュアな分離構成，UI実装により継続提供可能なサービスとして実装した．生成AIを課題整理の伴走役として位置づけることが，業務やサービス適用においては有効である．

Abstract　This paper reports the design and implementation of "Management Assist AI," a dialogue-based AI system that applies generative AI and Retrieval-Augmented Generation (RAG) to the agricultural business support platform AgriweB, from proof of concept (PoC) to production deployment. Starting from vague consultations, the system organizes key discussion points through dialogue and guides users to expert-supervised content with explicit source citations, including follow-up questions and suggestions for similar questions. Thus, it provides support for the clarification and in-depth exploration of issues that are difficult for conventional methods. Through the PoC, we confirmed the effectiveness of this approach and its operational feasibility. In production, we implemented authentication integration, a secure, isolated architecture, and a user interface to enable the continuous provision of service. The case study indicates that positioning generative AI as a companion for structuring issues—rather than an agent that delivers definitive answers—is effective for practical business and service applications.

1. はじめに

近年，生成AIは文章理解・要約・対話といった領域で高い性能を示し，業務やサービスへの適用が進んでいる．特に，大規模言語モデル[*1]（Large Language Model：以下，LLM）は，ユーザーの曖昧な相談に対して文脈を踏まえた応答や追加質問を生成できる点が特徴であり，従来の検索エンジンやFAQでは対応が難しかった課題に対する新たな手段として注目されている．

一方で，LLMは学習時点以降の情報や特定ドメインの専門知識に対する制約を有しており，根拠を示さない回答は業務サービスとしての説明責任を満たしにくい．こうした課題に対し，LLMの外部の信頼できる情報源を検索し，その結果を基に回答を生成する検索拡張生成[*2]（Retrieval-Augmented Generation：以下，RAG）は，知識の最新性と回答の信頼性を両立する手法として位置づけられる．

農業分野においては，経営環境や制度が複雑かつ変化しやすく，農業経営者が抱える課題も

明確に言語化されていない場合が多い．農業経営支援サイト AgriweB[*3] には，専門家が監修した多様なコンテンツが蓄積されている一方で，情報量の多さから，ユーザーが自らの状況に即した情報へ到達するまでに一定の負荷を要するという課題があった．

本稿では，AgriweB において，生成 AI と RAG を組み合わせた対話型 AI の概念実証（Proof of Concept：以下，PoC）を通じて業務サービスとしての有効性を検証し，「経営アシスト AI」[*4] を本番導入した取り組みを報告する．特に，課題が曖昧なユーザーに対する対話設計，専門コンテンツとの接続方法，出典提示による信頼性担保，および業務サービスとして成立させるためのシステム構成や運用設計に焦点を当てる．まず 2 章で背景と課題認識を整理して，3 章から 6 章で PoC の目的，設計，実装，検証結果を述べる．7 章では本番導入の構成および PoC の結果を基に実施した対応策について述べる．

なお，本稿では，LLM の性能比較や学習手法の詳細な比較検証，農業制度・補助金制度等の内容解説，UI/UX 設計の一般原則の体系的議論，セキュリティの実装詳細や設定値については記載の対象外とする．

2. 背景と課題認識

本章では，農業経営支援における情報探索の現状と，本プロジェクトが解決すべき課題について整理する．まず，農業経営者が直面している情報収集の困難さと，既存の支援サイトである AgriweB が抱えていた課題を述べる．次に，従来の FAQ や検索手法の限界を指摘する．その上で，生成 AI および RAG を採用した技術的背景と妥当性について，関連研究や他分野の事例を交えて整理し，本稿で扱う対話型 AI[*5] の位置づけと方向性を示す．

2. 1 農業経営における情報探索の課題

農業経営は，労働力不足，市場環境の変化，補助金・制度の多様化などにより，経営判断に必要な情報が増加し，意思決定の難度が高まっている．そのため，経営者が必要情報へ迅速に到達し，状況に応じて判断の質を高める支援の重要性が増している．

しかし現場では，情報が複数の情報源に散在し，ユーザーが「どこを見ればよいか分からない」「情報量が多く読み切れない」「自分の状況に合う判断材料に到達できない」といった状態に陥りやすい．特に課題が曖昧な段階では，検索する文章自体が定まらず，必要情報への入口を作りにくい．

2. 2 AgriweB の概要・ユーザー特性と課題

AgriweB は，農林中央金庫グループの株式会社 AgriweB（以下，AgriweB 社）が運営する農業経営支援サイトである．AgriweB では，農業経営に関する基礎的な事項を，ユーザーに理解しやすく体系的に整理した情報コンテンツ「基礎知識」を掲載している．さらに，農業経営者の課題意識に寄り添い，経営や技術，制度等に関する知見を分かりやすく解説した記事コンテンツ「コラム」などもある．AgriweB の会員登録者は 3 万名を超えている[1]．ユーザーは経営者，現場担当者，新規就農者など多様であり，抱える課題も経営計画，販路開拓，資金計画など幅広い．

このようなユーザーに対して AgriweB は専門家が監修した多様なコンテンツを提供しているが，課題は大きく二つに整理できる．

第一に，サイト構成およびコンテンツ閲覧に関する課題である．AgriweB には専門的な情報が蓄積されている一方で，情報量が多く，テキスト中心の構成であるため，ユーザーが必要な情報を探し出し，要点を把握するまでに一定の時間と負荷を要する．特に，ユーザーが複数の記事を比較しながら自身の状況に合う情報を判断する場合，情報探索の負担が大きくなる．

第二に，ユーザー自身の課題認識に関する課題である．農業経営に関する悩みは，経営計画，販路開拓，資金計画など複数の論点が絡み合うことが多く，ユーザーが最初から自身の課題を明確に言語化できるとは限らない．そのため，検索語や質問文を適切に作ることが難しく，結果として必要な情報への入口を見つけにくい場合がある．

以上のように，AgriweB における課題は，専門コンテンツを十分に活用するための情報探索上の課題と，ユーザーが自身の課題を明確化しにくいという課題の二面から捉えられる．本 PoC では，これらの課題に対して，生成 AI と RAG を組み合わせることで，ユーザーの曖昧な相談を起点に論点を整理し，関連する専門コンテンツへ誘導することを目指した．

2. 3　先行手法（FAQ/検索）における課題と，生成 AI 活用に向けた知識付与手法

近年，自治体等において，住民向け FAQ チャットボットの導入や活用が進んでいる[2]．FAQ 型 AI[*6] や検索型 AI[*7] は，定型的な問い合わせ対応には有効である一方，ユーザーが課題を適切に言語化できることを前提とする．そのため，課題が曖昧な相談では，質問が定まらず，対話が単発で終わりやすい．

また，生成 AI に業務ドメイン固有の知識を適切に組み込む手法としては，モデルの再学習[*8]（Fine-Tuning　以下，ファインチューニング）と LLM が有していない外部知識を活用する RAG が主要な選択肢となる．Microsoft が 2024 年に発表した研究では，複数の LLM と知識集約タスク[*9] を用いた比較実験により，新規知識や更新頻度の高い情報に対してはファインチューニングよりも RAG が一貫して高い性能を示すことが報告されている[3]．

さらに，医療分野では富士通の「AI 問診支援システム[4]」のように，RAG を活用して医療ガイドラインや論文を根拠とした回答を生成する事例がある．金融分野でも三菱 UFJ 信託銀行の「AI 規制 FAQ[5]」のように，法令・規制文書を検索し根拠付きで FAQ 回答を生成するシステムが実用化されている．

このような背景から，AgriweB においても，コンテンツ閲覧時の課題を解決するために，生成 AI と RAG を組み合わせた対話型 AI を採用することを検討して PoC を実施した．従来の検索型 AI や FAQ 型 AI は，ユーザーが明確な質問文や検索語を入力することを前提とするが，農業経営に関する相談では，ユーザー自身が課題を十分に整理できていない場合がある．そのため，ユーザーの曖昧な相談を起点に，対話を通じて論点を整理し，必要な情報へ段階的に導くために対話型 AI が有効であると考えた．

3.　目的と仮説（PoC）

本 PoC は 2023 年 12 月から 2024 年 3 月に実施した．PoC の目的は，農業経営者からの相談に対して「課題の具体化」および「課題の深掘り」を支援する対話体験を成立させることである．そのため，AgriweB の情報と LLM を組み合わせて，ユーザーの曖昧な相談から関連情報提示や対話を通じて深掘りできる仕組みを構築した．具体的には，AgriweB のサイト上のコンテンツ「基礎知識」「コラム」と，LLM のモデルである GPT-3.5[*10] の一般知識を併用した．

PoC では，ユーザーが「課題が分からない/言語化できない」状態から出発することを前提に，対話を通じて課題を深掘りすることで明確化し，必要情報へ導く価値を狙った．従来の検索や FAQ が「質問が明確であること」を暗黙の前提にしていたのに対して，本設計は「質問が曖昧であること」を入力条件として取り込んだ．

その上で，以下の三つを組み合わせることで，従来の FAQ 型 AI では困難であった「課題の具体化」を実現できるという仮説を立てた．

1）ペルソナ設計とシナリオ設計

2）RAG を用いた AgriweB 固有情報の参照

3）追加質問・類似質問の生成によるフォローアップ

4.　提供価値とサービス設計（PoC）

本章では，3 章で立てた仮説に基づき，PoC において想定したユーザー体験と提供価値，およびそのサービス設計について述べる．本 PoC では，農業経営者が抱える悩みとして，新たな販路開拓や事業拡大，計画書作成，事業承継など多岐にわたる課題が存在することを踏まえ，ペルソナ設計およびそれに基づくシナリオ設計を行った．具体的には，「事業・販路拡大を考えている農業経営者」をペルソナとして設定し，課題が曖昧な状態から相談を開始するユーザーを想定している．

農業経営という専門性の高いドメインにおいて，本システムでは，ユーザーの相談に対して対話を通じて課題を深掘りしつつ，RAG により参照した AgriweB コンテンツを出典として提示する構成とした．さらに，追加質問や類似質問の提示により，ユーザーの思考を継続的に支援する仕組みを備えている．これらの設計背景と，PoC で検証対象とした機能および導線について整理する．

4. 1　ユーザー課題と提供価値（PoC で狙った価値）

農業経営の分野では，事業計画・資金計画・販路開拓といった経営情報に加え，栽培技術，病害虫，気象，市況など生産・販売判断に関わる多様な専門情報を扱う必要があり，さらに制度や市場環境の変化に伴い内容が更新される領域においては，モデル内部への知識固定化よりも，信頼性のある既存コンテンツを柔軟に参照・活用できる構成が適している．そのため，本事例では AgriweB が有する専門コンテンツを最大限活用するため，RAG を中核とした対話型 AI による生成 AI アーキテクチャを採用した．

対話を通じてユーザーの課題やニーズを具体化し，信頼性のある専門コンテンツへ誘導することで，FAQ 型 AI/検索型 AI などの従来型サービスでは実現するのが難しかった価値を提供する．

4. 2　PoC で検証した機能と導線（PoC 範囲）

PoC では，生成 AI と RAG を組み合わせた対話型 AI を実装した．ユーザーの相談を受けて課題を深掘りし，参照した AgriweB コンテンツを出典として提示し，フォローアップとして追加質問・類似質問を提示する．主な機能は以下の表 1 の通りである．

表 1　各アプローチにおける主な機能

No	アプローチ	機能	説明	評価観点
1	ユーザーの曖昧な悩みを AI が段階的に具体化する.	チャットボット形式（対話 UI）	ユーザーの入力に対して追加の質問を返し，課題の明確化を支援する.	対話が継続するか/ユーザーが課題を明確化できるか
2	回答の根拠を明示し，信頼性を担保する.	RAG（検索 + 生成）/出典提示（URL 提示）	AgriweB コンテンツを検索し，参照箇所および記事 URL を回答と併せて提示する.	出典が常に提示されるか/参照元と回答の整合が取れているか
3	課題を多角的に掘り下げ，次の行動につなげる.	追加質問・類似質問	回答後に，深掘りのための追加質問・類似質問を提示し，次の問いを作る.	深掘りが促進されるか/導線（クリック等）が機能するか

5.　システム構成と実装（PoC）

本章では，4 章で示したサービス設計を実現するための技術的アプローチと実装内容について述べる．PoC 環境として採用したクラウド基盤および LLM の選定，ならびに対話制御を担うアプリケーション基盤の役割を整理する．特に，本システムの核となる RAG については，検索精度と回答の信頼性を両立させるためのデータ前処理やプロンプト設計，さらにユーザーの課題の深掘りを支援する機能の実装について示す.

5. 1　技術的アプローチ（PoC）

システムを構築するために PoC 用に Microsoft Azure（以下，Azure）環境を利用してプロトタイプ環境を構築した．PoC 環境であり一般公開しないためアクセス制限（IP 制限）をかけて，セキュリティを確保した．また，LLM は Azure のリソースの Azure OpenAI Service の GPT-3.5 を使用した.

5. 1. 1　チャットボット形式

対話型 AI の UI および会話制御には，BIPROGY 株式会社（以下，BIPROGY）が提供している対話型 AI アプリケーション基盤（RinzaTalk[*11]）を利用し，チャットボット形式のプロトタイプを構築した．RinzaTalk を利用することで，UI 開発や会話制御機能の個別実装に工数・コストをかけることなく，RAG による回答生成や追加質問・類似質問生成といった検証対象機能の実装・テストに注力することができた.

5. 1. 2　RAG 設計（データ前処理，ハイブリッド検索，プロンプト設計）

RAG に投入するコンテンツのデータ前処理では，AgriweB のコンテンツを分割・タグ付けし，検索精度を高めるためのハイブリッド検索[*12] を実装した．プロンプト設計では，AI に「農業経営者に寄り添うアドバイザー」としての役割を与えて，AI が誤った回答をしないように命令することで，出典提示や回答形式を制御した．これにより，回答の信頼性と一貫性を担保し，ユーザーが根拠を確認できるよう設計した.

図 1 に示す回答生成フローは，ユーザー入力における質問から RAG による検索，生成 AI による回答生成，出典提示までの一連の流れを示している．表 2 は図 1 の各処理ステップを対

応付けて説明したものである．AgriweB の専門コンテンツを検索可能な形式にするため，AgriweB 社から受領したコンテンツに対して前処理を実施した．具体的には，HTML タグ等のメタデータを除去し，本文テキストを検索投入用に整形した．メタデータを除去した検索用データを 800 文字単位でチャンク化[*13] し，本文テキストに対して Embedding[*14] を生成し，検索インデックスへ格納した．RAG におけるベクトル検索用の Embedding 生成には，Azure OpenAI Service が提供する text-embedding-ada-002 を用いた．これにより，ユーザー入力との意味的類似度に基づく検索ができるようになり，キーワード検索と組み合わせたハイブリッド検索の一要素として利用した．

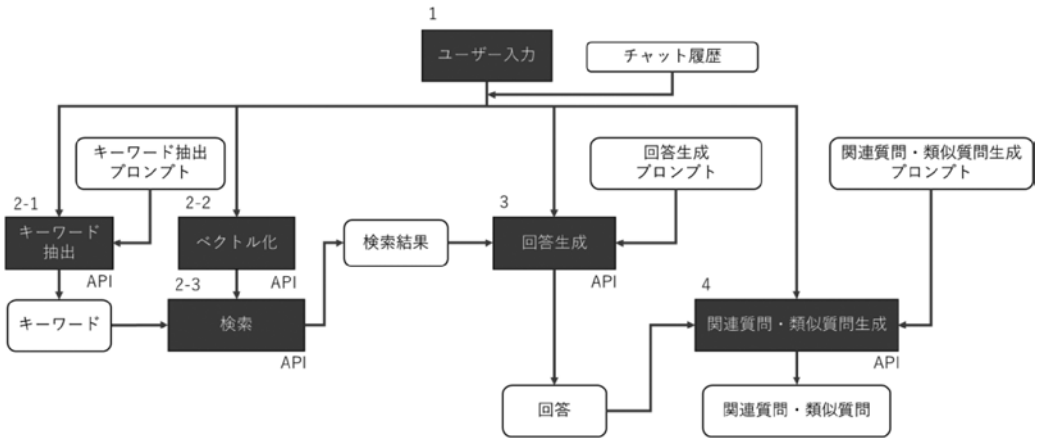

図 1　回答生成フロー

表 2　回答フローの説明

No	機能	概要
1	ユーザー入力	ユーザーがチャットアプリに質問文を入力し送信する．
2-1	キーワード抽出	No1 のユーザー入力からキーワードを抽出する．
2-2	ベクトル化[*15]	No1 のユーザー入力をベクトル化する．
2-3	検索	No2-1 で抽出したキーワードと No2-2 でベクトル化した値をインプットにハイブリッド検索し，検索結果を出典提示する．
3	回答生成	No2-3 で検索した結果と No1 のユーザー入力から回答を生成する．
4	追加質問・類似質問生成	No3 で生成した回答と No1 のユーザー入力から追加質問と類似質問を生成する．

5. 1. 3　出典提示

出典提示機能では，RAG の検索結果のヒット時に記事内の文章だけでなく，記事の URL もセットで返すことで参考文献として URL を提示することができる．

一方で，生成 AI の回答には，ハルシネーション[*16] のリスクがある．本システムではプロンプト設計や出力制御により，ハルシネーションが発生しづらい状況を作った．RAG による出典提示を必須とし，回答ごとに参照元記事を明示することで，ユーザーが一次情報を直接確認できるようにし，信頼性を担保した．また，ユーザーが出典を直接確認できる UI 設計とし，

説明責任を果たす仕組みを導入した．図 2 の枠内が出典提示機能であり，参考文献というかたちでユーザーが根拠情報を直接確認できる設計となっている．これにより，AI の回答の信頼性が担保される．

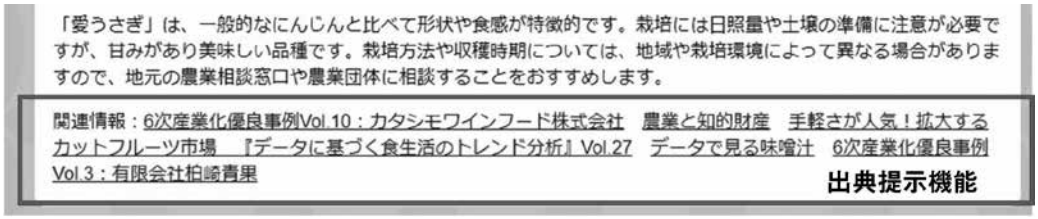

図 2　出典提示

5. 1. 4　追加質問・類似質問

追加質問・類似質問機能はユーザーが課題を深掘りできるよう，AI が追加で質問する追加質問機能や，類似の質問を提示する類似質問機能を実装した．AI は対話履歴を参照し，ユーザーの状況に応じて適切なフォローアップを行う．これにより，ユーザーが自らの課題を解決できる体験を設計した．図 3 の中央部に表示されている追加質問例が，ユーザーの課題の深掘りを促す設計となっている．

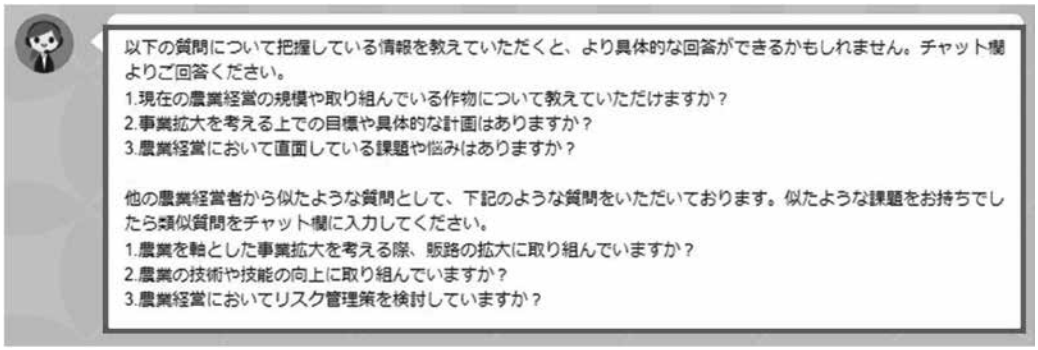

図 3　追加質問・類似質問

6.　検証と評価

本章では，AgriweB における生成 AI × RAG 活用の PoC を通じて実施した検証と評価について述べる．本 PoC では，定量的な性能評価ではなく，本番導入可否を判断するための定性的な観点での検証を主眼とした．

まず，6. 1 節では検証方法について整理し，シナリオ検証およびフィールド検証の観点から評価の枠組みを示す．次に，6. 2 節では各検証において得られた結果について述べ，対話体験の有効性および課題を明らかにする．最後に，6. 3 節ではこれらの結果を総合的に評価し，本番導入に向けた課題と位置づけについて整理する．

6. 1　検証方法

本 PoC における評価は，効果測定を目的とした定量評価ではなく，本番導入可否を判断す

るための論点整理として実施した．評価は，関係者による定例の打ち合わせの中で，都度デモ環境を用いた試行とフィードバックを繰り返す形で進めた．なお，本 PoC では定性評価における体系的な記録は作成していない．このため，本章で述べる結果は，当時の関係者間の合意に基づく定性的判断である点を本稿の制約として明記する．生成 AI および RAG を農業経営支援サービス「AgriweB」に適用するにあたり，実利用を想定した対話体験の妥当性を確認することを主眼に検証を行った．検証は以下の表 3 のとおり，二つの内容で実施した．

表 3　検証方法

検証内容	確認方法	期待値
シナリオ検証/本番導入を見据えた評価	事前に定義したユーザーペルソナおよび相談シナリオに基づき，想定される質問に対する応答内容，ならびに追加質問・類似質問による対話の連続性を確認した．	ユーザーの課題具体化につながること 出典提示と回答内容の整合が取れていること ハルシネーションが発生せず，ユーザーが理解できる日本語で回答が返ってくること
フィールド検証	RAG 構成のもと，実サービスに近い UI および導線で試行を行い，関係者によるレビューを通じて定性的な評価を実施した．	関係者が問題なくサービスを利用できること

6. 2　検証結果

本節では，6. 1 節で示した検証方法に基づき，シナリオ検証およびフィールド検証の結果について述べる．各検証において，ユーザーの課題具体化支援の有効性，回答内容の信頼性および出典との整合性，ならびに導線や利用体験上の課題について確認した結果を整理する．

6. 2. 1　シナリオ検証/本番導入を見据えた評価

シナリオ検証/本番導入を見据えた評価を通じて，ユーザーが抱える漠然とした悩みや関心事に対して，生成 AI が対話を通じて論点を整理し，経営課題としての切り口を提示できることを確認した．特に，単発の Q&A に留まらず，追加質問や類似質問を提示することで，ユーザーの課題を深掘りする導線が機能する点が確認された．

一方で，回答内容の具体性は，検索対象データの粒度や前処理状況に大きく影響を受けることが明らかとなり，RAG に投入するデータ設計の重要性が再認識された．検索対象データを一律に増やせば回答品質が向上するわけではないことも分かった．当初はコンテンツの一つである「一問一答」を検索対象データに含め「基礎知識」「コラム」「一問一答」の三つのコンテンツを使用して検証を行っていた．しかし，「一問一答」の情報は具体的かつピンポイントな回答を含むため，本 PoC が目的とした「曖昧な相談を起点に論点を広げる」用途では，検索結果のノイズとなる可能性があった．本件への対応として，「一問一答」の情報を検索対象から除外して検証したところ，回答が改善した．このことから，RAG に投入するデータは，単に量を確保するだけでなく，想定するユーザー体験や対話シナリオに応じて，粒度や対象範囲を設計することが重要であると考えられる．

また，生成 AI の回答には，参照した AgriweB コンテンツの出典を明示する構成とした．この結果，回答内容が AgriweB の「基礎知識」や「コラム」と自然に接続され，出典提示と

回答内容の整合が取れていることと，既存サービスの信頼性を損なうことなく AI を補助的に活用できることを確認した．

これにより，生成 AI が独立した知識提供者として振る舞うのではなく，AgriweB が保有する情報資産への入り口として機能する構成が有効であることが示唆された．

また，ハルシネーションの発生を抑えるプロンプト設計により，検証範囲においてハルシネーションは確認されず，ユーザーが理解できる日本語で回答が返ることが確認された．

そして，PoC 期間中に設計したプロンプト設計，回答生成ロジック，出典提示ならびに追加質問・類似質問生成の仕組みについては，基本構造を維持したまま本番環境へ適用できるとの判断に至った．

6.2.2　フィールド検証

フィールド検証においては，実際の利用を想定した試行を通じて関係者からフィードバックを収集した．その結果，回答内容の充実度や文章量については概ね好評であり，機能として利用したいという意見が得られるなど，一定の有用性が確認された．

一方で，回答生成に時間を要する点や，追加質問・類似質問の位置づけがユーザーにとって分かりにくく，次にどのような行動を取ればよいかが不明確となるケースがあることが指摘された．また，フォローアップ質問の内容自体には納得感があるものの，それを受けてどのように入力すればよいか，またその結果どのような応答が得られるかがユーザーに直感的に理解されにくいという課題が確認された．

加えて，生成 AI の特性上，不適切な回答の生成やそれに伴うリスクに対する懸念も指摘されており，過度な信頼や誤解を防ぐための利用上の説明やガイドの必要性が認識された．

以上より，生成 AI に対して過度な万能性を期待させないための説明や，ユーザーの期待値を適切にコントロールする設計が，本番導入に向けた重要な課題として整理された．

6.3　評価のまとめ

本 PoC で実施したシナリオ検証およびフィールド検証の結果を踏まえ，以下の通り評価する．まず，ユーザーの曖昧な相談に対して対話を通じて課題を具体化し，論点を整理する機能については，回答内容の充実度や文章量に対する評価からも，一定の有効性が確認された．

また，出典提示と回答内容の整合性についても問題なく，既存の AgriweB コンテンツへの導線として機能することが確認された．

一方で，追加質問および類似質問の位置づけがユーザーにとって分かりにくく，次の行動が不明確となる点や，回答生成までの待ち時間が UX 上の阻害要因となり得る点など，インターフェースおよび体験設計に関する課題が明らかとなった．さらに，生成 AI の特性上，不適切回答への懸念が残ることから，ユーザーに過度な期待や誤解を与えないための説明やガイドの整備を含め，利用時の期待値コントロールが不可欠である．

以上の結果より，本 PoC の構成は，農業経営支援において，「正解を提示する AI」ではなく，「課題を整理し，次の行動につなげる伴走役」として，一定の有効性を有することが確認された．よって，PoC で抽出した課題に対して改善を講じること（7.2.5 項で詳述）を前提に，本番導入を進めることにした．

7. 本番導入

本章では，PoC の検証結果および課題を踏まえ，AgriweB における生成 AI × RAG の本番導入に向けた取り組みについて述べる．6 章で整理した評価結果をもとに，サービスとして継続的に提供するために必要となるシステム構成，認証連携，UI 設計，および運用面の対応について整理し，PoC から本番環境へ移行する上での設計方針と実装内容を示す．

7.1 本番導入における位置づけ

PoC を通じて，AgriweB に蓄積された情報と生成 AI を組み合わせることで，ユーザーの課題整理や思考支援に一定の有用性が確認できた．一方，本番導入では，サービスとして継続的に提供することを前提に，セキュリティや運用を含めた対応が求められる．PoC で検証した機能や設計を最大限活用しつつ，本番利用に向けていくつかの対応を行った．本番導入する上でのシステム名称は「経営アシスト AI」とした．

7.2 本番導入において実施した主な対応

本節では，PoC で得た検証結果および抽出した課題を踏まえ，本番導入にあたり実施した主な対応について述べる．具体的には，システム構成や認証連携，UI 設計，および運用面の観点から，本番環境として安定的かつ継続的にサービス提供を行うために実施した対応内容を整理する．

7.2.1 本番導入・運用における役割分担

本番導入にあたっては，システム開発および運用を円滑に進めるため，関係者間で役割分担および責任範囲を明確に定義した．本システムにおける主な役割分担は以下の通りである．

1）BIPROGY

BIPROGY は，経営アシスト AI を構成する生成 AI 機能の設計・開発，Azure 上での基盤構築，UI 実装，ならびに RAG やプロンプトを含む回答生成ロジックの実装を担当した．また，本番環境の構築にあたっては，運用実績のある構成を迅速に適用するため，生成 AI 導入のための参照アーキテクチャ/環境構築テンプレート（Azure OpenAI Service スターターセット Plus[*17] 以下，スターターセット）を利用した．

2）AgriweB の開発ベンダー

AgriweB の開発ベンダーは，経営アシスト AI との認証連携の開発および AgriweB-経営アシスト AI の画面遷移の設計を担当した．

3）AgriweB 社

AgriweB 社は，経営アシスト AI のテストと，RAG へのコンテンツデータ投入作業を含む本番導入後の運用作業を担当した．

PoC での協業関係を踏まえつつ，本番導入・運用においては役割分担を明確化することで，安定的な開発および運用を実現した．図 4 に本番導入の開発スケジュールを記載する．

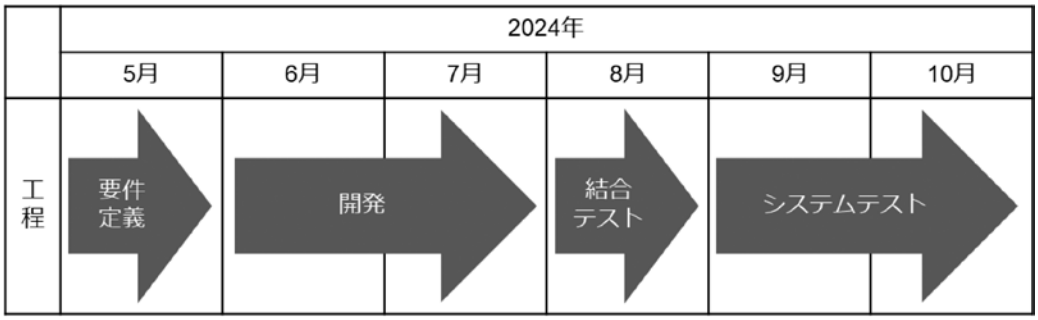

図 4　開発スケジュール

7. 2. 2　全体構成（AgriweB との分離，接続方式，認証連携）

本番導入では，AgriweB と経営アシスト AI を分離した構成とし，セキュリティと運用性を両立させている．AgriweB 側は会員管理・認証・専門コンテンツの提供を担い，経営アシスト AI 側は Azure 上に構築し，RAG による回答生成を含むチャットボット形式のチャットアプリ UI を担当する．両者は API 連携により接続され，ユーザー認証情報を利用する設計とした．ネットワーク分離や認証連携により，外部公開サービスとしての信頼性と責任分界を担保している（図 5）．

AgriweB と生成 AI サービスは物理的・論理的に分離されており，API 連携によって認証情報のみを安全に受け渡している．分離したことにより，経営アシスト AIのアプリのバージョンアップを行い，デプロイを実施する場合には AgriweB 本体をサービス停止することなく，バージョンアップを実施することができる．

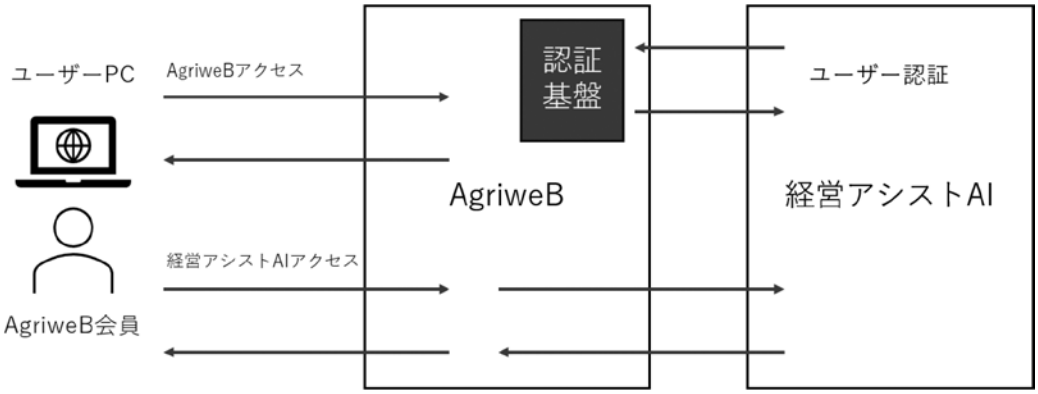

図 5　システム概要図

本番開発では，PoC 期間中に検証・調整した以下の要素を基本的にそのまま活用した．

1）回答生成プロセス

2）RAG による検索・生成方式

3）追加質問・類似質問生成機能

これにより，PoC で確認された有用性を損なうことなく，本番環境へ移行することができた．また，プロンプト設計の大幅な再設計を行わない方針としたことで，本番導入後の品質変動リ

スクを抑制している．

本番導入にあたっては，AgriweB の会員認証基盤と連携したユーザー認証を実装した．これにより，経営アシスト AI では新たなユーザー登録が不要となり，既存の AgriweB 認証方式を用いて，正規の会員のみがサービスを利用できる構成とした．

この設計により，以下の 4 点を実現した．

1）認証情報の重複管理を回避

2）不正利用リスクの低減

3）既存サービスとの一貫性確保

4）ユーザーの別サービスの会員登録作業の省略

7. 2. 3　UI 画面の本番実装

PoC では主に機能検証を目的としたデモ画面を用いていたが，本番導入に際しては，ユーザーおよび管理者の双方に配慮した図 6 のような UI 画面を構築した．具体的には，スターターセットに含まれる標準 UI をベースとし，AgriweB のユーザーの利用を想定したカスタマイズを行った．これにより，UI 開発工数を抑制しつつ，運用実績のある構成を採用することで，安定性を重視した実装とした．

図 6　経営アシスト AI のチャット画面

7. 2. 4　専門家相談の導線の設計

追加質問・類似質問によってユーザーのフォローアップを行った後に具体化した課題を専門家に相談する，という導線を設計した．課題を具体化した後，ユーザーはその課題を解決するためにそのまま対話を繰り返すこともできるし，AgriweB が提供している「経営アシストチャット」機能で専門家に相談することもできる．本番導入の際にユーザーが簡単にこの機能を利用できるよう，UI 上にその機能に遷移できるボタンを作成した．

7. 2. 5　PoC で抽出したフィードバックと本番導入での対応

PoC では，対話体験の受容性に加え，サービス提供に不可欠な運用上の課題を抽出した．本番導入にあたっては，PoC で得られた示唆を踏まえ，課題ごとに対応方針を整理した．表 4 に，PoC で抽出した課題，原因仮説，本番導入での対応を示す．

表 4　PoC で抽出した課題と本番での対応

PoC で抽出した課題	原因仮説	本番導入での対応
追加質問・類似質問の位置づけがユーザーに伝わりにくく，次の行動が不明確になる点が課題として挙げられた．	UI 上で「なぜ次の質問が提示されるか」の意図が説明されていない．	サービス案内の画面に追加質問・類似質問の概要を記載した．
回答待ち時間が UX 上の阻害要因となり得る．	回答生成中に何も表示されないため，ユーザーからするとシステムが動いているかどうか不安になる．	回答生成中はローディングアニメーションを表示して，ユーザー体験を阻害しない工夫を施した．
不適切回答が懸念される．	LLM の性質上，入力に応じて不適切回答の生成が起こり得る．	回答を鵜呑みにしない様に利用規約に記載した．

7. 3　本章のまとめ

本章では，AgriweB ×生成 AI システムの本番導入にあたり，実施した具体的な開発・運用対応について述べた．PoC で検証した成果を最大限活用しつつ，認証連携，セキュアな Azure 基盤構築，UI 実装といった本番特有の課題に対応することで，サービスとして提供できる生成 AI システムを実現した．

8. お　わ　り　に

本稿では，農業経営支援サイト AgriweB における生成 AI × RAG 活用の実践と設計上の勘所について，PoC から本番導入までの知見を整理した．課題が漠然としやすいユーザーに対し，対話型 AI による課題の具体化，専門コンテンツへの誘導，出典提示による信頼性担保，フォローアップ導線の設計，セキュアな運用構成など，業務適用に不可欠な要素を明らかにした．

最後に，本取り組みの関係者，執筆にあたって指導，協力頂いた方々に深く感謝するとともに御礼して本稿の結びとする．

＊ 1　大規模言語モデル：大量のテキストデータを用いて学習し，自然言語の理解や生成を行う深層学習モデル．
＊ 2　検索拡張生成：外部データを検索し，その検索結果を入力として大規模言語モデルによるテキスト生成を行う手法
＊ 3　AgriweB：農業経営者向けに，経営・技術・制度等に関する情報を提供する農業経営支援サイト．
＊ 4　経営アシスト AI：農業経営支援サイト AgriweB において提供される，大規模言語モデルを用いて農業経営に関する情報提供や意思決定支援を行う生成 AI 機能．
＊ 5　対話型 AI：ユーザーとの自然言語による対話を通じて，質問応答や情報提供を行う AI システム．
＊ 6　FAQ 型 AI：事前に蓄積された FAQ や文書情報を基に，ユーザーからの質問に対して自動で回答を生成・提示する AI システム．

＊ 7　検索型 AI：ユーザーの入力に応じて既存の文書やデータベースを検索し，該当する情報を提示することに特化した AI システム．
＊ 8　再学習：事前に学習済みのモデルに対し，特定の目的やドメインのデータを用いて追加学習を行い，性能や挙動を調整する手法．
＊ 9　知識集約タスク：外部知識や専門的情報を参照・統合しなければ適切な回答や判断が困難なタスク．
＊10　GPT-3.5：OpenAI が開発した大規模言語モデルの一種で，文章生成や質問応答などの自然言語処理を行うモデル．PoC 当時は最新のモデルであった．
＊11　RinzaTalk：BIPROGY が提供する対話型 AI ソリューションで，大規模言語モデルを活用し，業務における質問応答や情報提供を支援する．
＊12　ハイブリッド検索：キーワード検索（全文検索）と意味ベース検索（ベクトル検索）を組み合わせて，関連性の高い情報を取得する検索手法．
＊13　チャンク化：長い文章を一定の長さごとに分割し，検索や生成処理に適した単位にすること．
＊14　Embedding：テキストの意味内容を高次元ベクトルとして表現したもの（埋め込み）．ベクトル間の類似度に基づく意味検索に用いる．
＊15　ベクトル化：テキストの意味内容を数値ベクトルとして表現し，類似度に基づく検索を可能とする処理．
＊16　ハルシネーション：生成 AI が，学習データや与えられた情報に基づかず，事実と異なる内容や根拠のない情報を生成してしまう現象．
＊17　Azure OpenAI Service スターターセット Plus：BIPROGY が提供する，Azure OpenAI Service の利用を前提に，生成 AI 導入の初期検討から環境構築・活用支援までを支援する導入支援サービス．

参考文献

［1］ 株式会社 AgriweB，「アグリウェブの使い方」，AgriweB Web サイト，https://www.agriweb.jp/merit
［2］ 渋谷区，「渋谷区生成 AI チャットボット」，渋谷区 Web サイト，https://www.city.shibuya.tokyo.jp/kusei/kocho/ai-chatbot/ai_chatbot.html
［3］ Oded Ovadia, Menachem Brief, Moshik Mishaeli, Oren Elisha, “Fine-Tuning or Retrieval? Comparing Knowledge Injection in LLMs,” Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), Association for Computational Linguistics, 2024.
［4］ 富士通，「協業事例インタビュー」，富士通 Web サイト，https://global.fujitsu/ja-jp/about/activities/venture/interview/precision
［5］ 三菱 UFJ 信託銀行，「金融市場取引業務における生成 AI を活用した社内問い合わせ対応の自動化実現について」，ニュースリリース（PDF），https://www.tr.mufg.jp/ippan/release/pdf_mutb/240802_1.pdf

※　上記参考文献に示した URL のリンク先は，2026 年 6 月 11 日時点での存在を確認．

執筆者紹介

齋　藤　広　樹（Hiroki Saito）
2024 年，BIPROGY 株式会社に経験者採用で入社．入社以前は，Web アプリケーションの開発・保守，心臓 MRI 画像を対象とした心臓領域推定 AI モデルの開発，スマートフォンアプリケーション開発などに従事．BIPROGY 入社後は，生成 AI の導入，経理システムの基盤更改，営業店システム開発などを担当している．

生成AIを用いた探究学習の支援を目的としたAIコンシェルジュ

An AI Concierge Designed to Support Inquiry-Based Learning Using Generative AI

高 場 雄 太

要　約　生徒児童が探究学習を実施する上では，「自分で学習を進める自信がない」「探究学習でどんなことを調べたらいいのかわからない」といった課題がある．これらの課題解決のために，小学生でも使えるAI機能，「問」の質を高める機能，教材からの効率よい情報収集が可能な機能，教員用のログ確認機能を実装したAIコンシェルジュを開発した．

AIコンシェルジュを小学生に使用してもらい，アンケートと担当教員へのヒアリングを行った．結果から，小学生でも使えるAI機能，「問」の質を高める機能，教材からの効率よい情報収集が可能な機能については課題解決への効果を確認することができた．一方で，教員用のログ確認機能については課題が残る結果となった．

Abstract　Inquiry-based learning presents several challenges for students, including a lack of confidence in pursuing independent study and difficulty identifying appropriate research topics. To address these issues, we developed an AI concierge system incorporating four key features: an AI interaction function accessible to elementary school students, a function for improving the quality of research questions, a function for retrieving information from learning materials efficiently, and a log-monitoring function for teachers.

The system was evaluated through its use by elementary school students, followed by a questionnaire survey and interviews with the supervising teachers. The results confirmed that the AI interaction, question quality enhancement, and information retrieval functions were effective in addressing the identified challenges. However, the teacher-facing log-monitoring function revealed remaining issues that require further improvement.

1. はじめに

教育現場では，生徒児童ひとりひとりに端末を与える「1人1台端末」環境が促進されている．この取り組みの目的は，蓄積された教育データを活用することで，生徒児童への個別最適化された学びと，グループワークなどの協働的な学びの一体化[1]といった授業改善を行うことである．また，授業準備や成績処理などの校務負担を軽減し，学校における働き方改革を行うことも目的として挙げられている．小中学校においても，一律の指導や知識偏重を見直し，個に応じた学びや授業を実現するための探究学習の導入など，学びの在り方に変化が起きている．

一方で，生徒児童にとっては「自分で学習を進める自信がない」「探究学習でどんなことを調べたらいいのかわからない」といった課題もあり[2]，生徒児童が自律的に学ぶための「学び方」を身に付けるための伴走が求められている．

BIPROGY株式会社（以下，BIPROGY）は，東京学芸大学の高橋純先生や宮城教育大学の板垣翔大先生とともに，生成AIを活用し生徒児童と伴走するAIを目指した「AIコンシェルジュ」の機能検討を行った．本稿では，このAIコンシェルジュの効果を確認するために，佐川町立斗賀野小学校で行った実証実験の内容を報告する．まず2章ではAIコンシェルジュの

概要について，3章ではそのシステム構成や機能の詳細について説明する．4章では今回行った実証実験の内容と結果を述べ，5章では今後に向けた課題と改善点について考察する．

2. AIコンシェルジュ概要

本章では，AIコンシェルジュを開発するきっかけとなった課題と，実装した機能の概要について説明する．

2.1 探究学習における課題

文部科学省は「生きる力」の育成に向け，知識習得に加えて学びを自ら調整し，課題発見や対話を通じて考えを深める「主体的・対話的で深い学び」を求めている．その実現手法の一つが探究学習であり，教師主導の知識伝達ではなく，子ども自身が興味に基づきテーマを設定して調べ学習に取り組むことで，課題に向き合う力と学び方そのものを身につけることを目的とする．総合学習や各教科学習の中で，探究学習は既に実践されている．

一方で，探究学習には課題も指摘されている．第一に，インターネット情報の単純な収集や転記に留まり，探究の形式のみが整う一方で学びの内実が伴わない，いわゆる学びの空洞化が生じる事例が見られる．第二に，学校が設定したテーマやグループでの探究が優先される傾向があり，児童や生徒個人の興味や関心を十分に反映しづらいという構造的な課題がある．第三に，探究学習の評価観点や評価方法が明確化されておらず，教員の負担増加や支援体制の不足も継続的な検討課題として残されている[3]．加えて，生徒児童における課題として，学習のコツが掴めず上手く自分で進められないことや，効率的に調べ学習を進められるようにしたいが一人ではやり方が分からないことが挙げられる．

また，教員における課題として，探究学習を単なる「調べもの」ではない深い学びにしたいが，ひとりひとりに丁寧な指導を行うことが難しいことや，生徒児童の情報リテラシー面でインターネットを使用するにはまだ不安があるといったことが挙げられる．

最後に，教育委員会や管理職における課題として，指導要領方針を汲みGIGA端末[*1]を有効活用したいが，現場の教職員達の負荷無く授業を行うことが困難であることや，保護者説明含め安心安全なサービスを提供するための人員が不足していることが挙げられる．

これらの課題を解決するために，BIPROGYは「セキュリティや情報の信憑性など安心安全面に配慮した，子どもが自分の学びに合わせて使えるコンシェルジュ（以下，AIコンシェルジュ）」の開発を目指した．

2.2 AIコンシェルジュの四つの機能

探究学習は「課題の設定 → 情報の収集 → 整理・分析 → まとめ・表現」というサイクルで進められる．本研究ではこのうち前半の二段階，すなわち課題の設定と情報の収集に着目して支援機能を検討した．その理由は，これら前段階の質が不十分であれば，後段の整理・分析やまとめ・表現の質向上には結びつかないと考えられるためである．特に，まとめ・表現の段階，すなわちレポート作成や成果物の作成そのものは，生成AIやインターネット情報の単純な転用に依存せず，児童・生徒自身の力で行われるべき部分であり，これは前節の「学びの空洞化」という探究学習の課題とも直結する．

本研究では，教育向けに調整した生成AIの活用を，成果物作成の代替としてではなく，テー

マ設定や情報収集の過程における教育的な見方・考え方の習得，すなわち事象を調査する方法そのものを身につけることを支援する目的で，以下の四つの機能を考えた．

1）小学生でも使える AI 機能

生成 AI のサービスによっては，年齢制限のため小学生では利用できない場合がある．また，個人情報や機微な情報の扱いの面で，子どもに自由に使わせるのは不安といった懸念もある．これらに対応するため，文部科学省による生成 AI ガイドライン[4]に準拠し，フィルタリングなどを用いて安全に利用できる AI を提供することを「小学生でも使える AI 機能」と定義した．

2）「問」の質を高める機能

特に探究学習のシーンにおいて，「テーマが選べない」や「表面的な調べ学習になり問が深まらない」といった問題があった．これらに対し，学習指導要領の「見方・考え方」に基づき，問の生成・深堀を支援する機能を，「「問」の質を高める機能」と定義した．

3）教材からの効率よい情報収集が可能な機能

GIGA 端末を活用しつつも，インターネットの情報だけではなく，教科書や図書などの質のよい情報源からも学んで欲しいという教員からの要望と，教材のどこに欲しい知識や情報があるか素早く検索し繰り返し学習したいという生徒児童からの要望に基づく機能を，「教材からの効率よい情報収集が可能な機能」と定義した．

4）教員用のログ確認機能

生徒児童が AI の危険な使い方をしないように制御やモニタリングをしたい，また，生徒児童の学習の足跡を確認して，1 人 1 人の成長を支援したいといった要望に基づく教員向けの管理機能および設定機能を，「教員用のログ確認機能」と定義した．

3.　システム構成

本章では，今回設計した AI コンシェルジュのシステムの全体構成と，そのうちの AI 部分である AI エージェントの機能詳細について説明する．

3. 1　全体構成

以下の図 1 に今回設計したシステムの全体構成を示す．システムは，Microsoft Azure 上に構築した．AI 部分は Prompt flow[*2] を，フロントエンド部分は streamlit[*3] を用いて実装した．また，AI 部分については，OpenAI 社の大規模言語モデル（LLM）を搭載した簡素な AI エージェントを実装している．

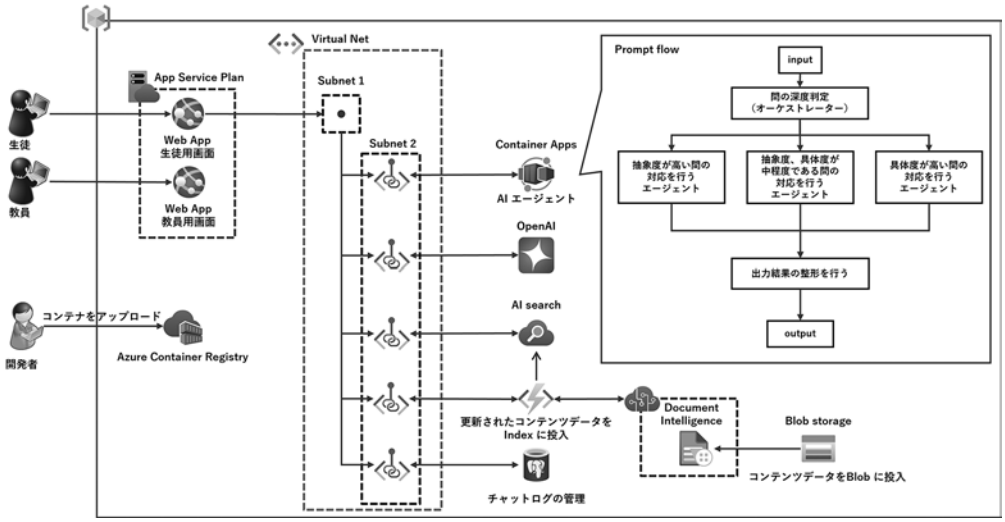

図 1　システム構成図

3. 2　AI エージェントについて

AI エージェントとは，AI を利用して特定のタスクやサービスを自動化するソフトウエアである．2.2 節で述べた「2)「問」の質を高める機能」を実現するためには，以下のフローが求められる．

1)　生徒児童から AI コンシェルジュへの問の質を判断する
2)　判断した問の質に合わせた，新たなる問を生成する
3)　生成した問を生徒児童にわかりやすいように簡潔にまとめ出力する

このようなフローを実現するためには複数の AI を要する．今回の場合，具体的には以下のような AI が求められる．

- 生徒児童からの問の質を判断する AI
- 判断した問の質に応じて対応する AI

このうち，判断した問の質に応じて対応する AI については，判断する問の質の数だけの AI を要する．今回実装したシステムでは，判断する問の質は三つとし，問の具体度に応じて，低いものから順に step 1 ~ step 3 に分類した．step 1 ~ step 3 と判断される問の具体例を表 1 に示す．

表 1　問の具体例

Step	問の種類	問の具体例
step 1	抽象度の高い問	• 戦いについて知りたい • 武将について
step 2	抽象度，具体度共に中程度の問	• 織田信長について知りたい • 大イチョウの樹って？ • 高度経済成長期とは？ • 戦国時代が気になる
step 3	具体度の高い問	• 織田信長の生まれた年は？ • 大イチョウの樹のある場所は？ • 戦国時代の有名な武将は誰？

表 1 に示した step 1 ~ step 3 において，より問の具体度を高められるような回答をする以下四つのエージェントを作成した．

- 問が step 1 ~ step 3 のどれに該当するかを判断するエージェント
  - ➢ 以下，step check エージェントとする
- step 1 の問に対して，問が具体的になるような問を作成するエージェント
  - ➢ 以下，step 1 エージェントとする
- step 2 の問に対して，問の具体度をより高める問を作成するエージェント
  - ➢ 以下，step 2 エージェントとする
- step 3 の問に対して，問の発想を広げる問を作成するエージェント
  - ➢ 以下，step 3 エージェントとする

また，各エージェントからの返答を整形する部分については，Python で実装した．各エージェントを組み合わせたフローを図 2 に示す．

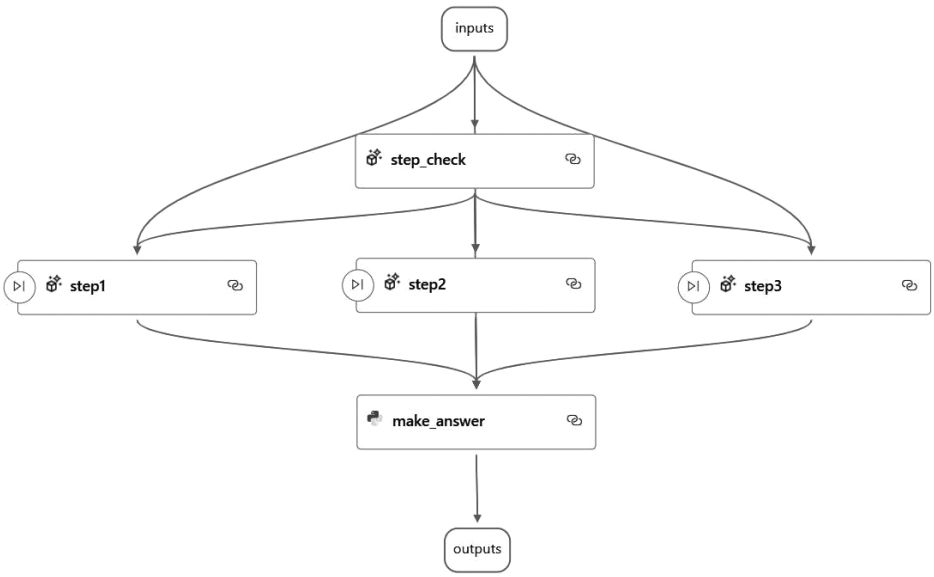

図 2　エージェント処理のフロー図

step check エージェントでは，生徒児童の問が step 1 ~ step 3 のどれに該当するかだけでなく，学習に関係のない入力や，攻撃的な入力に該当するかについても判断を行い，そのような入力があった際には，エラーを発生させるように設計し，2.2 節の「1）小学生でも使える AI 機能」を実現した．

step 1 エージェントは，生徒児童の問を受け取り，生徒児童が次に step 2 か step 3 の問を考えられるような問を生成し，step 2 エージェントでは，step 3 の問を考えられるような問を生成するように設計した．また，step 3 エージェントは生徒児童の問からさらに疑問を広げるような問を生成するように設計した．

これらの問の生成には，教科書や図書データを取り込んだ Retrieval Augmented Generation（RAG）を使用している．RAG とは，外部の知識ベースから関連情報を検索し，それを LLM にプロンプトとして与えることで，回答生成を補助する技術である．これにより，LLM

は未知の情報にも対応でき，ハルシネーションのリスクも低減できる．RAG を組み込むことによって，探究学習の「情報収集」段階をサポートするほか，自分が知りたい，見直したい知識や情報を生徒児童が既存の教材から検索しやすくなる．RAG によって，普段の予習・復習の効率や教材活用率を向上させるとともに，教科書および図書の何ページを参照すれば回答が得られるのかを，エージェントが生成した問と合わせて示すことで，2.2 節の「3）教材からの効率よい情報収集が可能な機能」を実現した．また，引用元が明記されるので，教員の「生徒児童がよくわからない情報源からコピー＆ペーストしてきているかもしれない」という不安を軽減できた．

3.3　機能詳細

各エージェントに対する入力は以下の引数をとる．

- 生徒児童の学年
- 対象教科
- 問の具体度を上げる時に注目する観点
- 生徒児童からの問

この中で，「問の具体度を上げる時に注目する観点」は学習指導要領に沿って選定した．例えば，社会科では「時間」「空間」「人物」の三つの観点に沿って問の具体度を上げるようにしている．生徒児童からの問が step 1 と判断される「戦いについて知りたい」のような抽象度の高い問であった場合には，時間に関する具体度を上げる問である「戦国時代の戦いについて調べてみよう」や，空間に関する具体度を上げる問である「日本の戦いの歴史に出てくる重要な戦場はどこだろうか？」，人物についての具体度を上げる問である「有名な武将たちがどのように戦ったのかを調査してみよう」が生成される．これらの問の内容を盛り込み，最終的に生徒児童へ向けたアウトプットを step 1 エージェントが生成し，生徒児童の問の具体度を上げることを目指している．以下の図 3 に実際の step 1 の対応例を示す．

図 3　step 1 対応例

AI コンシェルジュの生徒児童向けの工夫点として，生徒児童からの問に直接解答となる文

章を生成しないというものがある．通常の LLM では，「織田信長の生まれた年は？」のような解答が存在する問を入力すると，その解答が生成される．ただ，今回のような学習のサポートという観点では，解答そのものが生成されてしまうと，探究学習が表面的な調べ学習になり問が深まらないと考えた．そこで，問に対する解答を生成するのではなく，あくまで問に関連する問を生成するプロンプトを設計した．

さらに，AI コンシェルジュを用いて学習指導要領の見方・考え方に沿った問を繰り返すことで，生徒児童による主体的な見方・考え方の習得や，生徒児童自身にはない視点の引き出しの蓄積などができる．この問を繰り返すことで，文部科学省が目指す探究学習におけるゴールである「変化の激しい時代において自分で課題解決をはかるための考え方や取り組みのスキルの獲得」のための訓練となる．その問の繰り返しの中で，step 3 と AI が判定するような具体度の高い問が生徒児童から入力された場合には，その問が「問の具体度を上げる時に注目する観点」のうちどれに近いかを推論し，その観点に基づいた問を生成し，より深い学びにつながるようなプロンプトを設計した．これらのプロンプトにより，2.2 節の「2)「問」の質を高める機能」を実現した．

また，教員用の確認機能として，会話ログおよび生徒児童による NG ワードの入力を可視化できるダッシュボードを実装した．このダッシュボードにより，2.2 節の「4）教員用のログ確認機能」を実現した．ダッシュボード機能の実際の例を以下の図 4 に示す．

図 4　ダッシュボード機能

4.　実証実験

本章では，高知県にある佐川町立斗賀野小学校の 6 年生に，実際にシステムを使用してもらった実証実験について紹介する．

4. 1　実証実験概要

今回の実証実験では，佐川町立斗賀野小学校の社会科の探究学習「私たちの生きる社会とお金の顔」という題材において，6 年生 21 名に実際に AI コンシェルジュを使用してもらい，使用感などについてアンケート調査を実施した．

実証実験を行う授業の流れは，導入（10 分），展開（60 分），まとめ（10 分）の形となっており，主に展開部で AI コンシェルジュを使用した．導入部では，これまでの学習の振り返りと何について調べるかを確認した．展開部では，坂本龍馬，徳川家康，与謝野晶子，杉田玄白の 4 人の功績や社会に与えた影響について，図書館で借りてきた本やタブレット，AI コンシェルジュを使いつつ調べ学習を行った．また，調べ学習だけでなく，4 人のうち 3 人を一万円札，五千円札，千円札の肖像として選ぶならそれぞれ誰を採用するかとその理由，残りの 1 人を採用しなかった理由についても児童に考えてもらった．その後，調べ学習の結果をロイロノート[*4] に記録させた．まとめ部では，ロイロノートを活用して他の児童の考えを見られるようにし，その考えについての感想と振り返り，AI コンシェルジュについてのアンケート調査を実施した．アンケートの設問を以下に示す．

問 1　AI コンシェルジュは，操作しやすかったですか．（選択式）
問 2　調べたいこと（テーマ）を決めやすくなりましたか．（選択式）
問 3　AI の返事は，考えるヒントになりましたか．（選択式）
問 4　AI を使って，考えが広がったり変わったりしましたか．（選択式）
問 5　また授業で AI コンシェルジュを使ってみたいと思いましたか．（選択式）
問 6　AI コンシェルジュを使ってみて，「なるほど」「おもしろい」「ちょっと変だな」と思ったことがあれば，自由に書いてください．（自由記述形式）

選択式の設問については「1. とてもそう思う 2. そう思う 3. あまりそう思わない 4. ぜんぜんそう思わない」から選んでもらい，自由記述形式の設問については自由に回答を記載してもらった．

また，授業の終了後に授業の担当教員にヒアリングを行い，普段の探究学習と AI コンシェルジュを使用した場合での相違点について調査した．

4. 2　結果

「AI コンシェルジュは，操作しやすかったですか．」という問 1 についてのアンケート結果を以下の図 5 に示す．

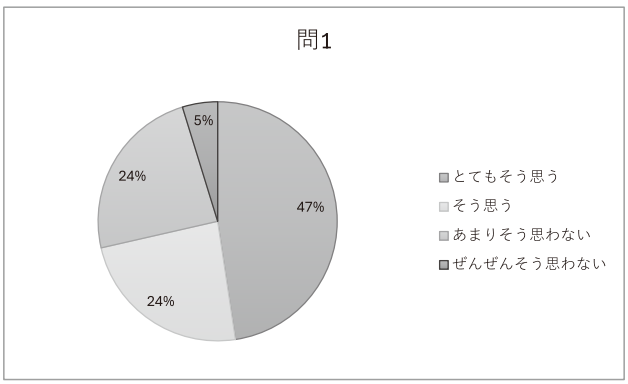

図 5　問 1 のアンケート結果

問 1 の結果として「とてもそう思う」と「そう思う」の合計が 71%となり，児童が AI コンシェルジュを操作するにあたり UI/UX の観点において大きな障害はなかったことがわかる．

「調べたいこと（テーマ）を決めやすくなりましたか.」という問 2 についてのアンケート結果を以下の図 6 に示す．

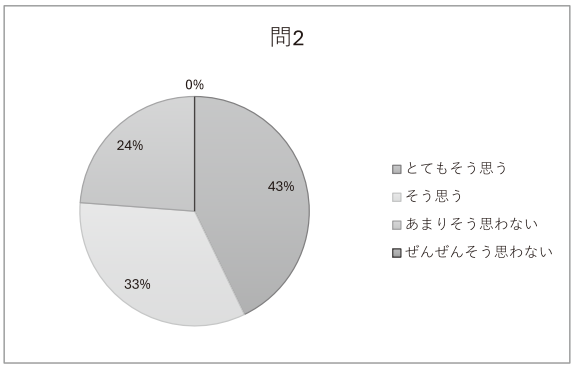

図 6　問 2 のアンケート結果

問 2 の結果として「とてもそう思う」と「そう思う」の合計が 77%，「ぜんぜんそう思わない」を選択した児童は 0 人だった．ここから，AI コンシェルジュは児童が調べたいことを決める際に補助としての役割を果たしていることがわかる．

「AI の返事は，考えるヒントになりましたか.」という問 3 と「AI を使って，考えが広がったり変わったりしましたか.」という問 4 についてのアンケート結果を以下の図 7 に示す．

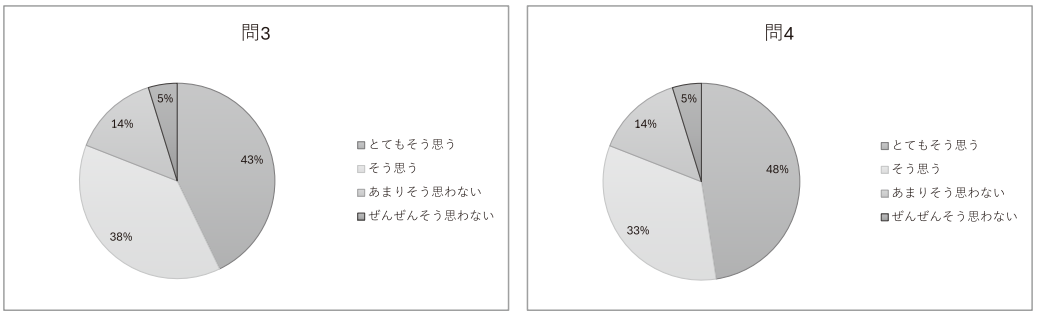

図 7　問 3，問 4 のアンケート結果

問 3，問 4 共に「とてもそう思う」と「そう思う」の合計が 71%となっており，AI コンシェルジュが生成する問は児童の考えのヒント，あるいは考えを広げるきっかけとなることがわかる．

「また授業で AI コンシェルジュを使ってみたいと思いましたか.」という問 5 についてのアンケート結果を以下の図 8 に示す．

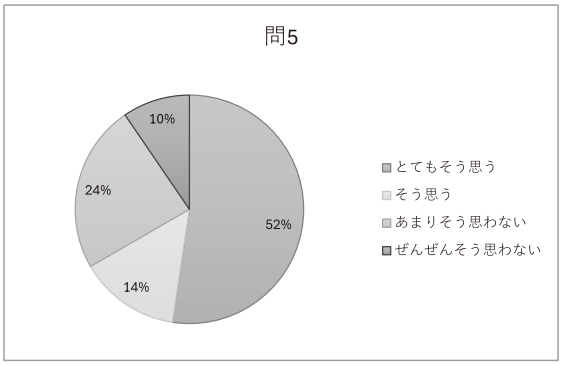

図 8　問 5 のアンケート結果

問 5 の結果として「とてもそう思う」と「そう思う」の合計が 66％となっており，半数以上の児童がもう一度 AI コンシェルジュを使用したいと考えていることがわかる．しかし，「あまりそう思わない」，「ぜんぜんそう思わない」を選択した児童が 34％となっており，考えを広げるきっかけとはなってはいるが，継続利用については消極的である児童がいることもわかる．

「AI コンシェルジュを使ってみて，「なるほど」「おもしろい」「ちょっと変だな」と思ったことがあれば，自由に書いてください．」という問 6 の記述から一部抜粋したものを以下に示す．

- 自分にはなかった考え方のヒントをくれて面白いなと思った
- 調べたら教科書も出てくるからよりわかりやすかった
- AI コンシェルジュは「ほぼ〜を調べてみてください」と教科書に載っている資料だけを出してくるから，もっと他のこと出して欲しいです
- 「歴史」について詳しい事を調べる事は出来たけど，同じ人の違う質問をしたら同じ資料しか出なかった

これらの自由記述から，考えを広げるきっかけとなったことや，教科書データを使用することで，わかりやすい問の生成ができていることがわかる．しかし，資料の数が少ないなどの意見も出ており，探究学習における伴走を支援するには，教科書データと数冊の図書データだけでは RAG に使用するデータとして足りていないことがわかる．

授業後に行ったヒアリングでは，担当教員から以下のような意見が得られた．

- 児童から「何を調べたら良いか」という質問が AI コンシェルジュを使用していない時と比較して大きく減ったため，負担軽減となった．
- 児童から使い方や，AI コンシェルジュに何を聞いたら良いのかというような質問がくると思っていたがなかったので，子供の思考にあっていると思った．
- 教科書データをもとに問を生成するため，ハルシネーションや教育的に不適切な文章が生成されないため安心して使用できる．
- 学校独自で使用している教材を読み込ませることができると嬉しい．
- 机間指導[*5]がメインとなるため，ダッシュボード機能についてはあまり有効に活用できなかった．

ヒアリング結果から，教員から見ても AI コンシェルジュが児童の伴走として役立っていることがわかる．また，教員の負荷軽減にも有効であることや，AI コンシェルジュの仕様につ

いて安心できることが示されている一方で，ダッシュボード機能などの児童の学習における見取り部分については，改善が求められていることがわかる．

5. 考察

4 章で述べた実証実験の結果について考察する．「AI コンシェルジュは，操作しやすかったですか.」という問 1 のアンケート結果と教員からのヒアリング結果から，児童が AI コンシェルジュを操作するにあたり大きな障害はなかったといえる．一方で，「また授業で AI コンシェルジュを使ってみたいと思いましたか.」という問 5 のアンケート結果からは，授業で再度 AI コンシェルジュを利用したくないと考えている児童が 3 割程度いることがわかる．この原因としては，「AI コンシェルジュを使ってみて，「なるほど」「おもしろい」「ちょっと変だな」と思ったことがあれば，自由に書いてください.」という問 6 の自由記述への回答にあるように，コンテンツが不足していることや，答えを教えない AI というのが児童にとってストレスとなってしまっている可能性が考えられる．「教科書が出てくることで理解が進みやすい」などの自由記述もあったため，AI コンシェルジュの「3）教材からの効率よい情報収集が可能な機能」については現時点で有効性が示唆されつつも，コンテンツの数を増やすことでより効果が表れる機能であると考えられる．

「調べたいこと（テーマ）を決めやすくなりましたか.」という問 2 への回答として「ぜんぜんそう思わない」を選択した児童は 0 人だったことから，AI コンシェルジュは児童が調べたいことを決める補助としての役割を果たしているといえる．さらに，教員へのヒアリング結果からは，AI コンシェルジュの使用によって，普段の調べ学習時よりも「何を調べるか」という部分で躓く児童が減ったことがわかった．加えて，「AI の返事は，考えるヒントになりましたか.」「AI を使って，考えが広がったり変わったりしましたか.」という問 3，問 4 の結果と問 6 の自由記述の内容から，AI コンシェルジュは児童の考え方を広げるきっかけとなることがわかった．これらのことから，AI コンシェルジュの「2)「問」の質を高める機能」は想定通りの働きをしており，探究学習のシーンにおける「テーマが選べない」や「表面的な調べ学習になり問が深まらない」といった課題に対して有効性が示された．

また，教員へのヒアリングでは，「ハルシネーションや教育的に不適切な文章が生成されないため安心して使用できる」といった回答が得られた．ここから，AI コンシェルジュの「1）小学生でも使える AI 機能」を実装することで，生徒児童の情報リテラシー面でインターネットを使用するにはまだ不安があるという課題を解決できると考えられる．

一方で，教員の授業中における見取り部分については，ダッシュボードではなく机間指導がメインであったということがアンケート結果から伺える．このような結果となった原因としては，普段の授業においてダッシュボードのようなものを確認する習慣があまりないことや，机間指導の方が児童の進捗とチャットログを同時に確認できることが考えられ，「4）教員用のログ確認機能」については，課題の残る結果となった．今後は，ダッシュボードにチャットログだけでなく，児童のチャットの傾向から推測できる現在の進捗や指導方針のレコメンドなどの情報を追加することで，ダッシュボード機能がより授業における教員の補助となるように改善する予定である．

6. お わ り に

本稿では，LLM を活用し生徒児童と伴走する AI を目指した「AI コンシェルジュ」開発の取り組みと，その導入効果を評価するための実証実験について説明した．実証実験の結果からは，生徒児童が AI コンシェルジュを使用することで，探究学習をサポートできる可能性があることがわかった．また，AI コンシェルジュの改善案についても把握できた．今回見つかった課題への対応を行い，より探究学習をサポートできるシステムとなるよう改善していく予定である．小中等教育における生成 AI の利用については，安全性の問題や教職員による指導の方法，保護者への理解など課題となる部分は多くあるが，協力や理解を得られるように尽力していく所存である．

最後に，本稿執筆にあたりコンテンツ協力いただいた教育出版様・日本電子図書館サービス様及びメイツユニバーサルコンテンツ様，ご指導いただいた方々と教育 AI のプロジェクトメンバーに感謝申し上げます．

＊ 1　文部科学省が推進する「GIGA スクール構想」に基づき，児童・生徒に 1 人 1 台配備された学習用端末（タブレットやノート PC など）．

＊ 2　Azure が提供している AI アプリケーションの開発サイクル全体を効率化するように設計された開発ツール．
https://learn.microsoft.com/ja-jp/azure/machine-learning/prompt-flow/overview-what-is-prompt-flow?view=azureml-api-2/

＊ 3　Python 製のオープンソース Web アプリフレームワーク．https://streamlit.io/

＊ 4　株式会社 LoiLo（ロイロ）が提供する，全国の小中高や大学など多くの教育機関で導入されている授業支援クラウドサービス．

＊ 5　授業中に教師が教壇から離れ，生徒児童の席の間を巡回しながら，個々の学習状況の把握，個別の助言・指導，評価を行う教育技法．

参考文献　［1］　文部科学省初等中等教育局教育課程課，学習指導要領の趣旨の実現に向けた個別最適な学びと協働的な学びの一体的な充実に関する参考資料，2021 年 3 月，https://www.mext.go.jp/content/210330-mxt_kyoiku01-000013731_09.pdf

［2］　文部科学省・国立教育政策研究所，OECD 生徒の学習到達度調査 PISA2022 のポイント，2023 年 12 月 5 日，https://www.nier.go.jp/kokusai/pisa/pdf/2022/01_point_2.pdf

［3］　文部科学省，総合的な学習・探究の時間に関する現状・課題と検討事項，2025 年 10 月 15 日，https://www.mext.go.jp/content/20251015-mxt_kyoiku02-000045235_4.pdf

［4］　文部科学省，初等中等教育段階における生成 AI の利活用に関するガイドライン，2024 年 12 月 26 日，https://www.mext.go.jp/a_menu/other/mext_02412.html

※　上記の注釈および参考文献に示した URL のリンク先は，2026 年 7 月 1 日時点での存在を確認．

執筆者紹介　高　場　雄　太（Yuta Takaba）

2022 年 BIPROGY（株）入社．プラットフォームサービス本部にて群集マネジメント研究や，AI サービス開発業務に取り組む．2024 年より市場開発本部にて教育分野への生成 AI 活用の企画検討業務に参加．

複数拠点向け太陽光発電量・余剰量予測サービス開発における データ活用手法と実用性評価

Data Utilization Methods and Practicality Evaluation in the Development of a Multi-Site Solar Power Generation and Surplus Forecasting Service

藤　本　容　子

要　約　カーボンニュートラルの実現に向けて再生可能エネルギーの導入拡大が加速する中，太陽光発電を持続的に社会へ定着させるためには，発電事業者を含む関係者にとって経済合理性のある運用を実現することが不可欠である．そのためには，発電販売計画の基礎となる発電量・余剰電力量を高精度に予測し，インバランス料金の抑制につなげることが重要となる．BIPROGY は，2023 年 4 月から提供している太陽光発電量・余剰量予測サービスにおいて，近年拡大しつつある低圧複数拠点モデルへの適用に向け，複数拠点を一括で予測するバルク予測を検討した．本稿では，複数のバルク予測方式について，予測精度・経済効果・開発コストの 3 軸から総合的に評価し，実運用に適した方式を選定した検討過程および検証結果について述べる．

Abstract　As the deployment of renewable energy accelerates toward the realization of carbon neutrality, ensuring economically viable operations for all stakeholders, including power producers, is essential to the sustainable adoption of photovoltaic (PV) power generation in society. To achieve this, it is important to accurately forecast PV generation and surplus power, which serve as the basis for generation and sales planning, thereby reducing imbalance charges. To address this need, BIPROGY investigated within its PV Generation and Surplus Power Forecasting Service, which was launched in April 2023, an approach that performs bulk forecasting for multiple PV sites collectively, with a view to applying it in the emerging multi-site low-voltage PV model. This paper describes the evaluation process and verification results used to identify a practical forecasting method by comprehensively assessing multiple bulk forecasting approaches from three perspectives: forecasting accuracy, economic effectiveness, and development cost.

1. は　じ　め　に

2050 年のカーボンニュートラル実現に向けた取り組みの一環として，再生可能エネルギーの導入拡大が加速する中，太陽光発電はその中心的な役割を担っている．太陽光発電の普及拡大は，脱炭素社会の実現だけでなく，将来的なエネルギー供給の安定化という社会課題の解決にも寄与する．一方で，出力変動に伴う需給調整コストの増加など，普及拡大に伴う新たな課題も顕在化している．太陽光発電を持続可能なエネルギーとして社会に定着させるには，これらの課題を解消し，関係者にとって経済合理性のある運用を実現することが不可欠である．

BIPROGY 株式会社（以降，BIPROGY）は，こうした課題の解決を支援するため，2023 年 4 月から AI 技術を活用した「太陽光発電量・余剰量予測サービス」を提供している．太陽光発電量・余剰量予測サービスは，太陽光発電量および余剰電力量を高精度に予測することで，電力需給の最適化や太陽光発電事業の収益性向上に貢献するものである．

発電事業者は予測結果を基に発電販売計画を策定している．計画と実際の発電量に乖離が生じた場合には，インバランス料金（計画のズレに対するペナルティ）が発生するため，予測精度の向上は，事業の収益性や安定性に直結する重要な要素となる．

太陽光発電量・余剰量予測サービスは従来，大型高圧拠点を対象とした単体拠点向けに展開してきた．しかし，系統混雑や用地確保の課題から大型高圧拠点の新規開発は減少傾向にある．

他方，大型高圧拠点の課題を回避する新たな取り組みとして，低圧（小規模）の発電所を複数拠点まとめて開発する動きがはじまっている．低圧拠点は，地域の電力需要に応じた柔軟な設計ができるだけでなく，政府による分散型電源の導入促進や再生可能エネルギーの主力電源化に向けた制度整備の対象として位置づけられている．また，コーポレートPPA[*1]の普及や企業による脱炭素経営の加速といった市場ニーズにも適合している．

こうした状況を踏まえ，BIPROGYは，低圧複数拠点モデルへの適用を見据え，複数の発電所を一括で予測するバルク予測に着目した．本稿では，複数のバルク予測方式について，予測精度に加え，インバランス料金低減による経済効果や運用コストの観点から評価を行い，実運用に適した方式を選定した検討過程および検証結果について述べる．まず2章でサービス概要とバルク予測について説明した後，3章でバルク予測の課題，4章で課題への対応策および実現パターンを整理し，5章で検証条件および評価方法，6章で検証結果と考察を述べる．

2.　太陽光発電量・余剰量予測サービスの概要と複数拠点化への対応

本章では，太陽光発電量・余剰量予測サービスの概要を説明した後，複数拠点化に向けた対応方法について説明する．

2.1　太陽光発電量・余剰量予測サービスの概要

太陽光発電量・余剰量予測サービス（以降，当該予測サービス）の概要図を図1に示す．当該予測サービスは，AI技術の中でも機械学習を活用した予測モデルに基づき，過去の実績電力データと太陽光発電設備設置地点の気象データを組み合わせて予測を行っている．具体的には，日射量や気温，雲量，風速などの気象情報と発電設備ごとの発電履歴を基に，予測モデルが電力量の変動要因を学習し，将来の発電量や余剰電力量を高精度に予測している．

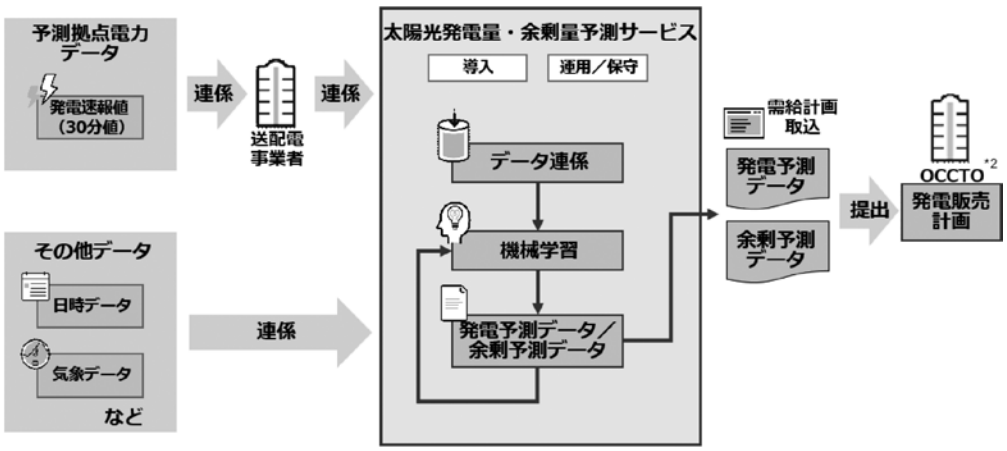

図1　太陽光発電量・余剰量予測サービスの概要図

2.2　単体拠点から複数拠点化に向けた検討

今後，分散型の低圧拠点による太陽光予測のニーズの高まりが予想されることから，BIPROGY では複数拠点を対象とした太陽光予測サービスの開発を検討した．

開発にあたっては，単体拠点向けの既存サービスを活用・拡張する方針とした．この方針で，従来通りに複数拠点を個別に予測した場合，新たな開発費は不要である一方，対象拠点数の増加に伴い，拠点追加時の導入コストや運用管理コストの増大が懸念された．

そこで，拠点ごとに太陽光予測サービスを構築・運用するのではなく，複数拠点をまとめて管理できる予測方式が必要であると考えた．その方式として，複数拠点を統合して一つの予測対象として扱う予測（以降，バルク予測）に着目し，既存サービスの改修を試みた．バルク予測では一部の機能改修を要するものの，複数拠点を一括で管理できるため，拠点追加時の導入コストや運用負荷の大幅な削減が見込まれる．その結果，総合的には費用対効果が最も高いと想定されることから，本方式の適用を検討した．図 2 に予測方法のイメージを示す．

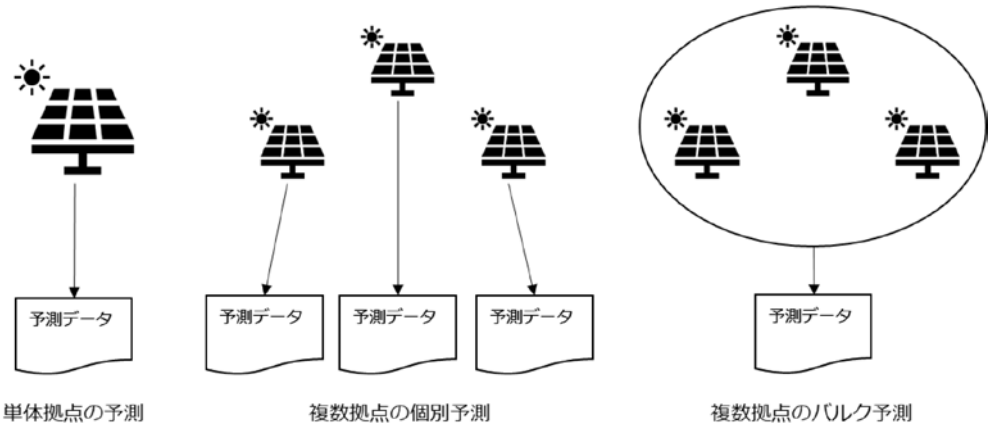

図 2　太陽光予測方法のイメージ

3.　バルク予測の課題

バルク予測の開発にあたっては，「予測精度を確保すること」と，「導入コストを抑制すること」の両立が重要となる．前者は複数拠点をどのようにまとめ，どの気象データを用いるかに大きく依存し，後者は既存サービスからの改修範囲に依存すると考えられる．これらの観点から，バルク予測の開発にあたり，以下三つの課題を整理した．

1）予測対象拠点のグルーピング基準

予測精度を担保するためには，複数拠点の電力データをどのような基準でグルーピングして，一つの予測対象として扱うかを明確にしなければならない．

2）予測に利用する気象地点の選定

バルクとして統合した拠点群に対して，どの気象地点のデータを用いるかが予測精度に影響するため，適切な選定方法を検討しなければならない（単体拠点では，拠点に最も近い気象地点を使用していた）．

3）改修費用の低減方法

単体拠点向けの既存サービスを活用して，改修費用を最小限に抑える設計が求められる．

これらに対して複数の解決案を比較検討し，予測精度，経済効果および開発コストを評価指標として検証を行い，実運用に適した予測手法を選定した．

4. バルク予測の課題への対応策

前章で述べた課題に対応するため，複数拠点のグルーピングおよび気象地点について検討した．

4.1 課題1），2）に対する拠点のグルーピングと気象地点の選定

バルク予測の開発にあたっての課題1），2）に対して，以下の検討を行った．

拠点のグルーピング単位としては，予測出力の最大単位であるバランシンググループ[*3]（以降，BG）と，より細分化された都道府県単位を候補とした．また，気象地点としては，BGの中心地点および都道府県の中心地点を候補とした．

これらの拠点のグルーピングと気象地点の組み合わせについて，以下の四つのパターン（A～D）を設定し，予測精度およびコストのバランスを評価するため，Bを除く三つのパターンで比較検証を行った．各パターンの詳細は表1および図3に示す．

なお，コストの観点では，拠点を大きなグループでまとめる方が望ましいが，これまでの運用経験から，予測精度の観点では，より小さなグループの方が有利であることが示唆される．そのため，両者のバランスを考慮して最適な組み合わせを選定することが重要となる．

・パターンA：拠点をBG単位でグルーピングし，BGの中心地点の気象データを使用する．
・パターンB：拠点を都道府県単位でグルーピングし，BGの中心地点の気象データを使用する．ただし，気象地点がグルーピング単位外となり，予測対象の気象条件を適切に反映できないことから，精度の担保が困難であると想定されたため，検証対象から除外した．
・パターンC：拠点をBG単位でグルーピングし，各都道府県の中心地点の気象データを使用する．
・パターンD：拠点を都道府県単位でグルーピングし，各都道府県の中心地点の気象データを使用する．

表1 拠点のグルーピングと気象地点の組み合わせ

		気象地点	
		BG中心地点	都道府県中心地点
拠点のグルーピング	BG単位	パターンA	パターンC
	都道府県単位	パターンB （検証対象外）	パターンD

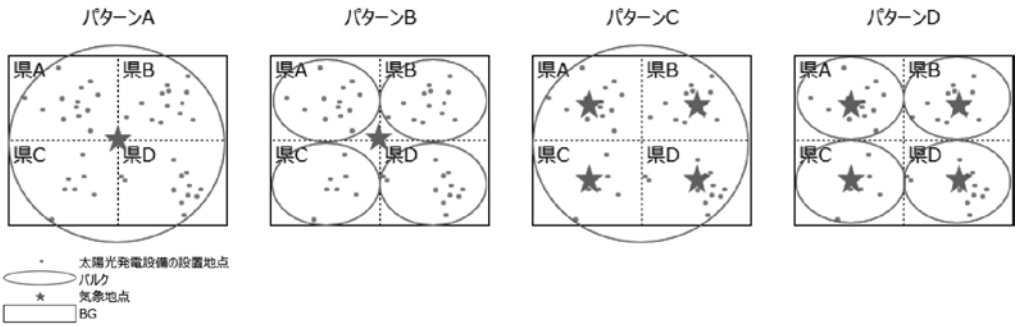

図3　拠点のグルーピングと気象地点の組み合わせイメージ

4.2　課題3）に対する改修費用を抑えた精度向上策

単体拠点向けの既存サービスでは，曜日情報の有無や学習期間の調整が予測精度の改善に寄与した実績がある．これらは既存の予測ロジックのパラメータ設定変更のみで実現でき，データ構造やモデル構造の改修を伴わないことから，課題3）（改修費用の低減方法）に合致する精度向上策と考えた．そこで，バルク予測においても同様の観点を取り入れ，改修費用を抑えつつ精度向上ができるかを検証した．

・曜日情報の有無：特徴量[*4]として曜日情報を含めるか否かが予測精度に与える影響を検証した．

・学習期間の調整：予測モデルの学習に用いる過去データ期間が予測精度に与える影響を検証した．特に複数拠点を統合する場合，学習期間が長すぎると地点数の変動に追随しにくくなるため，適切な学習期間の設定が重要と判断した．

5.　検証条件および評価方法

本章では，検証方法，検証データおよび評価指標について述べる．

5.1　検証方法

本検証では，4章で整理した対応策の有効性を評価するため，予測シミュレーションを採用した．予測シミュレーションは，多様な条件下で複数の予測ロジックを同一条件下で比較でき，導入効果を事前に定量評価できるためである．検証にあたっては，パターンA，C，Dについて，曜日特徴量の有無および学習日数を組み合わせた条件で予測を行い，得られた結果を評価指標に基づき評価する．

5.2　検証データ

検証データには，単一の実運用環境から取得されたデータに基づくデータセットを用いた（表2）．当該データの分析および公開に際しては，企業等を識別できないよう厳格な匿名化および統計的集計処理を施して，プライバシーおよび機密性の保護を確保している．

表 2　検証データの詳細

項目	内容
実績電力量データ	某事業者より提供された過去 12 か月間（2023 年 8 月～ 2024 年 7 月）の実績電力量データで，発電電力量と余剰電力量の両方が含まれている．
予測シミュレーション期間	9 か月間（2023 年 11 月～ 2024 年 7 月） 2023 年 8 月～ 10 月は学習期間のため予測シミュレーション期間の対象外とする．
予測タイミング	前日 7:00 に翌日 24 時間分（00:00 ～ 23:30 の 48 コマ）を予測
地点数	126 地点（複数県に点在．期間中に拠点の増減がある．）

5. 3　評価指標

予測シミュレーション結果を評価するために，以下三つの指標を設定し，これらをバルク予測の開発における総合的な判断材料とした．

1）予測精度

MAPE（平均絶対パーセント誤差）[*5] を用いて，予測精度を評価した．MAPE は機械学習モデルの予測精度を評価する指標の一つで，値が小さいほど予測誤差が小さく，高精度であることを示す．

2）経済効果

発電インバランス料金を算出し，それが発電事業者に与える経済的影響について評価した．発電インバランス料金とは，発電計画値と実際の発電量の差異に応じて発生する調整費用であり，不足インバランス料金と余剰インバランス料金の和（発電インバランス料金＝不足インバランス料金＋余剰インバランス料金）として定義される．

不足インバランス料金は，発電計画よりも実績が少ない場合に発生し，その差分に不足インバランス単価を乗じた金額であり，発電事業者が系統運用者に支払う．一方，余剰インバランス料金は，発電計画よりも実績が多い場合に発生し，その差分に余剰インバランス単価を乗じた金額であり，発電事業者が受け取ることができる．

本検討では，発電インバランス料金がゼロに近いほど経済的に有利であると評価した[*6]．

3）開発コスト

各検証パターンの構築にかかる開発原価を評価した．

6.　検証結果と考察

本章では予測シミュレーションを用いた検証結果をまとめる．

6. 1　検証結果

4. 1 節の拠点のグルーピングと気象地点の組み合わせパターンと，4. 2 節の曜日特徴量，学習日数の影響について検証した結果を記載する．

（1）予測精度

図 4 に各パターンの曜日特徴量の有無（曜日無：点線・曜日有：実線）と学習日数（07, 14, 30, 60, 90）による 9 か月平均 MAPE の比較グラフを示す．

全ての学習日数において，曜日特徴量無しの方が，MAPE が低く，予測精度が高い結果

となった．学習日数が増えるにつれて MAPE は低下し，30 日間の学習が最も高い予測精度を示した．30 日以降は若干の精度低下が見られるものの，ほぼ安定した精度が維持されることが確認された．

曜日特徴量の寄与が低くなった要因としては，検証対象に自家消費のない発電所が多く含まれており，余剰量予測に比べて発電量予測の割合が大きかったことが挙げられる．発電量予測は主に気象条件に依存するため，曜日の影響を受けにくく，結果として曜日無しの方が高い予測精度を示したと推察した．

複数拠点の場合，拠点数および電力量は一度に増えるのではなく，日々増減がある．そのため，学習日数を長く設定すると，発電電力量の変動に対するモデルの追随性が低下する可能性が懸念された．しかしながら，実際には 30 日以降の学習でも予測精度への大きな影響は見られず，一定の安定性が保たれることが分かった．

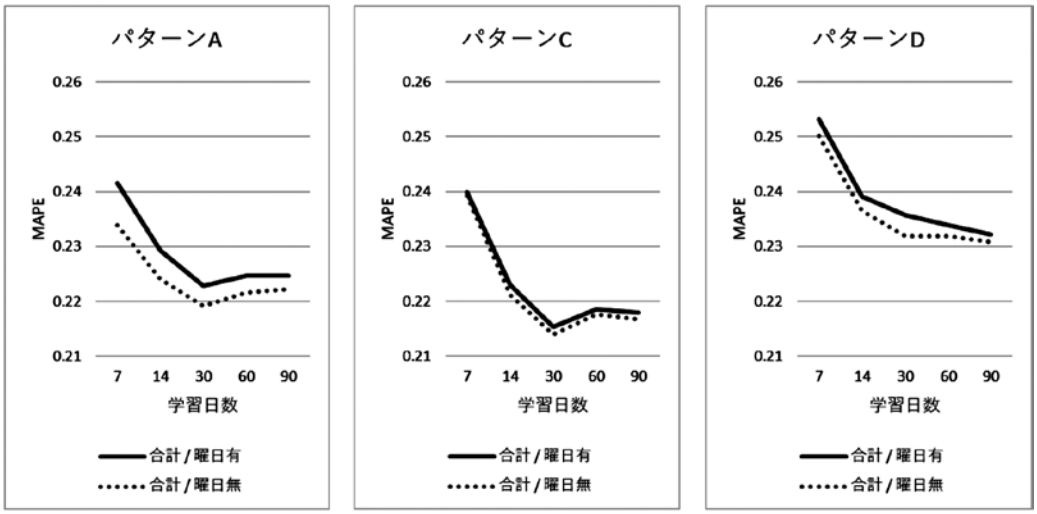

図 4　曜日特徴量，学習日数が予測精度に与える影響

この結果から，各パターンを同条件（曜日特徴量無し，学習日数 30 日）で比較した結果を表 3，図 5 に示す．

パターン C が最も MAPE が低く，高い精度であった．また，標準偏差も最小であり，月ごとのばらつきが小さい安定した予測となった．月別の結果でも全体的に MAPE が低く，特に冬季（1 月）や夏季（7 月）でも他より誤差が抑えられており，季節変動に強いモデルと評価できる．次点はパターン A で，C より劣るが良好なモデルである．パターン D は他と比べて精度・安定性ともに劣る結果となった．

もともと，パターン A は D に比べて広域の気象情報を扱うため，局地的な変動が捉えにくくなり，予測精度が劣ると想定していたが，今回の結果ではその予想に反して，パターン A の方が高い精度を示し，パターン D の方が低いという結果となった．

この要因としては，パターン D では実際の運用を想定し，発電拠点の初期分布ではなく県の中心地点を気象地点として設定している点が挙げられる．そのため，発電拠点の分布と気象地点に乖離が生じた場合，気象地点の局地的な気象現象（例：突発的な雨など）が強く

反映され，予測精度が低下した可能性がある．

一方，パターン A の BG 単位の広域的な気象情報では，局地的なばらつきがならされるため，結果的に予測精度が向上したと推察した．ただし，他のエリアや発電分布でも同様の傾向が得られるかどうかについては，今後の検証が求められる．

表 3　予測精度と安定性

パターン	9 か月 MAPE	標準偏差
A. 曜日無 30	0.219121	0.03823
C. 曜日無 30	0.213996	0.03396
D. 曜日無 30	0.231971	0.03757

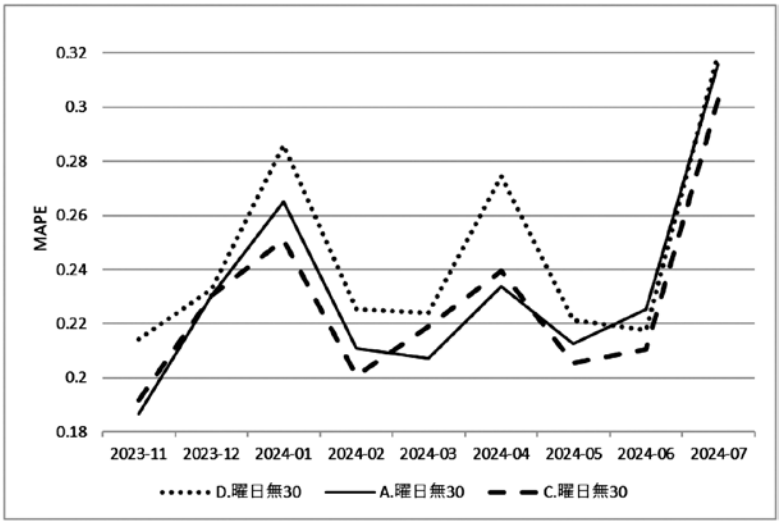

図 5　パターン別 MAPE 推移

（2）経済効果

図 6 に各パターンの曜日特徴量の有無（曜日無：点線・曜日有：実線）と学習日数（07, 14, 30, 60, 90）による 9 か月合計インバランス料金の比較グラフを示す．発電インバランス料金がゼロに近いほど経済的に有利と評価している．

学習日数にもよるが，特徴量として曜日無しの方が経済的に有利な傾向がある．学習日数では，短期学習日数（07 ～ 30 日）は損失が大きく，学習日数 60 日が最も経済的に有利な結果となった．

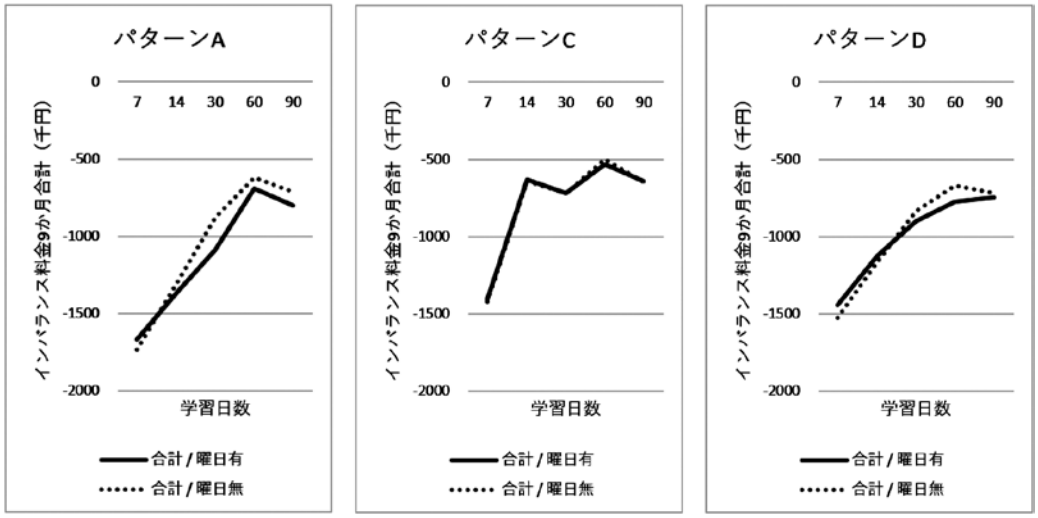

図 6　曜日特徴量，学習日数がインバランス料金に与える影響

以上の結果から，各パターンを同条件（曜日特徴量無し，学習日数 60 日）で比較した結果を表 4，図 7 に示す．

表 4　経済効果

パターン	インバランス料金月平均（千円）
A. 曜日無 60	-69
C. 曜日無 60	-56
D. 曜日無 60	-74

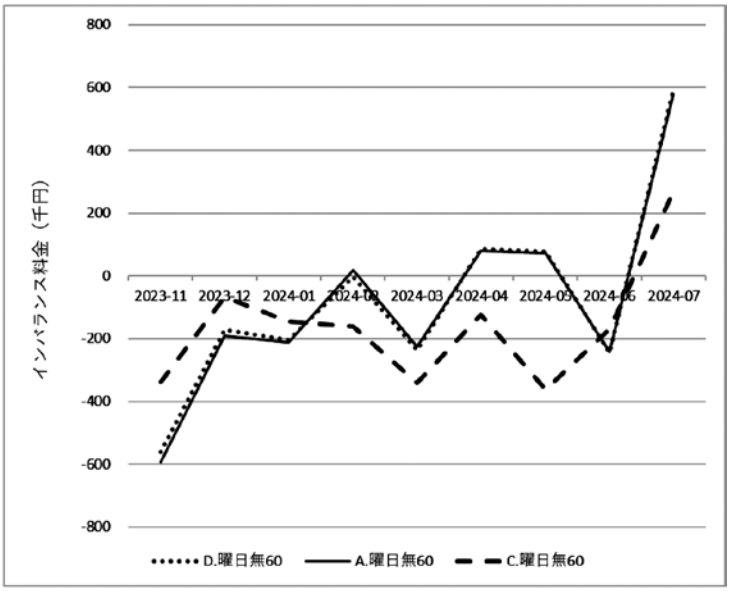

図 7　パターン別インバランス料金推移

パターン A，C，D のインバランス料金を月別に分析した結果，最も経済効果が高いと評価できるのはパターン C であり，次点はパターン A であった．パターン C はインバランス料金の平均値が最もゼロに近く，全期間を通じて安定した経済効果を示している．一方，パターン A と D はほぼ同様の傾向が見られた．2 月から 5 月にかけては，パターン C よりも良好な結果を示している月もあったが，全体を通して評価すると，パターン C が最も優れた経済効果を示したと結論づけられる．

さらに，上位パターンである A，C について，某事業者にサービスを提供した場合の月平均予測料金と実質予測料金を算出した結果を表 5 に示す．パターン A，C 間には大きな差は見られなかった．

予測料金とは，サービス提供時に発生するサービス利用料を指し，実質予測料金とは，BIPROGY がインバランス料金を負担するサービスプランの場合に発生するサービス利用料金である．それぞれ以下の式で表される．

予測料金＝予測電力量（kWh）×ランニング費用単価（円/kWh）

実質予測料金＝予測料金－インバランス料金

表 5　月平均予測料金および実質予測料金

パターン	実績電力量（kWh）	予測電力量（kWh）	インバランス料金（比較指数）	予測料金（比較指数）	実質予測料金（比較指数）
A. 曜日無 60	354,590	340,375	1.000	1.000	1.000
C. 曜日無 60		339,867	0.807	0.999	0.964

（3）開発コスト

パターン A，C，D の開発原価について評価した．

開発は，既存の単体拠点向け予測を活用・拡張する方針を前提とし，主にデータ構造の違いに着目した．なお，特徴量や学習期間は設定値であり，開発作業に直接的な影響はない．

既存の単体拠点向け予測では，予測対象地点に対して 1 ヵ所の気象地点データを保持するデータ構造となっている（予測対象地点：気象地点＝1：1）．バルク予測においては，パターン A および D では，一つのバルクに対して使用する気象地点が 1 ヵ所である（予測対象地点：気象地点＝1：1）ため，既存データ構造を流用できる部分が多く，影響範囲は限定的である．

一方，パターン C では，一つのバルクに対して複数の気象地点を使用する（予測対象地点：気象地点＝1：N）ため，データ構造が既存と異なる．この構造の変更により，テーブル設計やバッチ処理など，広範な修正を要するため影響範囲が大きくなる．パターン C の開発原価はパターン A および D の約 3.25 倍と試算される．

6. 2　考察

前節の検証結果を基に，予測精度・経済効果・開発コストの三つの観点から各パターンを総合的に評価し，実運用への適合性を判断した（表 6）．表 6 の評価記号は，各評価観点におけるパターン間の相対的な優劣を示す．予測精度については MAPE および月別の精度の安定性，経済効果についてはインバランス料金の月平均額および月別推移，開発コストについては既存

システムへの影響範囲および開発原価を評価指標として判定した．

予測精度および経済効果の両面において最も優れた結果を示したのはパターンＣであり，特に季節変動に対する強さやインバランス料金の安定性が高く評価された．しかしながら，パターンＣは複数の気象地点を扱うため，既存システムとの互換性が低く，データ構造の大幅な変更を要することから，開発コストが他パターンと比較して著しく高くなる．

予測精度や経済効果ではパターンＣが最も優れていた．しかし，次点であるパターンＡと比較すると，今回の規模の電力量における開発コストの回収期間は，パターンＡが1.7年に対し，パターンＣは5.8年と長期に及ぶ．

また，拠点数の増減が発生するタイミングで導入コストが発生する点も考慮した．単体拠点向け予測と比較すると，バルク予測では導入作業が軽微であり，導入コストは相対的に小さい．さらに，バルク予測のパターンＡ～Ｄの比較においては，バルク規模が大きくなるほど運用コストは低減すると想定している．

経済効果の差に比べて，開発コストの差が非常に大きかったため，総合的に勘案すると，パターンＡが有力な選択肢と判断した．

なお，本事例は開発コストと予測精度のバランスを重視した一例であり，精度・経済効果・開発コストのいずれを重視するかは，プロジェクトの目的や状況に応じて異なる点に留意が必要である．

表6　総合評価

	パターンＡ	パターンＣ	パターンＤ
予測精度	○	◎	△
経済効果	○	◎	○
開発コスト	◎	△	◎

7. おわりに

本稿では，AIを用いた予測機能の開発において，仮説検証のプロセスや評価指標を整理した．これにより，開発効率の観点から有用な知見を得られた．

仮説検証のプロセスや評価指標の整理は，関連技術の開発を進める上で応用可能な有用性を持っており，太陽光予測に限らず，他領域への応用も期待される．特に，予測精度そのものではなく，インバランス料金など事業収益への影響を経済効果として定量化し，開発コストと合わせて総合的に方式を選定するプロセスは，AI予測モデルを実サービスとして展開する際に共通する評価の枠組みであり，電力領域におけるインバランス単価予測や市場価格予測，需要予測などに応用できる可能性がある．

今後は，本検証で得られた知見を基に，カーボンニュートラルの実現と電力需給の安定化という社会課題の解決に向けて，これらの予測領域への応用を進めていく予定である．最後に，本プロジェクトの遂行にあたりご協力いただいた関係者の皆様，ならびに本稿の作成にご尽力いただいた皆様に，心より感謝申し上げる．

＊1　企業が再エネ発電事業者と直接契約し，長期的に電力を購入する仕組み．
＊2　電力広域的運営推進機関は，電源の広域的な活用に必要な送配電網の整備を進めるとともに，全国大で平常時・緊急時の需給調整機能を強化するため，専門的知見と強い事業者間調整機能を有する組織として2015年4月に設立．電気事業法に基づく認可法人として，中立で公平な業務運営を行っている．OCCTOまたは広域機関と呼称される．
https://www.occto.or.jp/occto/
＊3　電力の需給バランスを調整するために，複数の電力事業者や需要家が集まって形成するグループのこと．特に計画値同時同量制度の下で，インバランス（計画値と実績値の差）の精算単位となる事業者群を指す．
＊4　機械学習モデルが予測を行うために判断材料として利用する入力データの要素．
＊5　絶対誤差値（予測値－実績値の絶対値）を実績値で割り，その総和をデータ数で割って求めた平均値のこと．
＊6　計画より発電量が少ない場合，発電事業者は不足インバランス料金を支払う必要がある．一方，多い場合は余剰インバランス料金を受け取れるが，その単価は通常，電力市場価格よりも低いため，計画通りに売電した場合と比べて，収益性が劣る可能性がある．

参考文献

［1］　環境省，太陽光発電の導入支援サイト，2026年4月，
https://www.env.go.jp/earth/post_93.html
［2］　太陽光発電量・余剰量予測サービス | AI予測サービス，BIPROGY，
https://www.biprogy.com/solution/service/ems_power_prediction.html
［3］　BIPROGYニュースリリース「独自のAI技術で高精度のバルク予測を実現し，小売電気事業者と発電事業者の費用負担リスクを軽減」，BIPROGY，2025年6月，
https://www.biprogy.com/pdf/news/nr_250630.pdf

※　上記注釈および参考文献に示したURLのリンク先は，2026年7月3日時点での存在を確認．

執筆者紹介

藤　本　容　子（Yoko Fujimoto）
2023年BIPROGY（株）入社．エネルギー領域における「太陽光発電量・余剰量予測サービス」の開発・保守・導入を担当．カスタマーサクセスエンジニアとして，課題分析と対応を通じてサービス価値の向上に取り組んでいる．