

# 生成 AI のビジネス活用実現に向けた羅針盤

株式会社 野村総合研究所 AI コンサルティング部  
シニアコンサルタント 坂 拓弥

株式会社 野村総合研究所 AI コンサルティング部  
コンサルタント 露木 浩章



## 1 ビジネスにおける生成 AI 利活用の現状

2022年11月のChatGPTリリースを契機として、昨今のビジネス環境では、多くの企業が生成AIのビジネス領域における利活用の検討に注力している。生成AIの利活用検討が注力される理由は、大規模言語モデル（Large Language Model：以降、LLM）を主とした生成AIにより、顧客接点やサービス、既存業務などの多様な局面に対して、大きな付加価値を生み出すことが見込まれるためである。各企業の取り組みとしては、ChatGPTのような生成AIを活用したチャットボットの導入推進が多く見受けられる。一方で、プロダクトや自社固有業務に対する生成AIの適用範囲拡大はまだ道半ばであり、今後各企業が重要なテーマとして対応を図らなければならないと想定される。

これまでの各企業の取り組みは、ChatGPTが広く出始めた初期の段階から、「とりあえず導入してみる」という実験的な導入が多く見られた。これらは、新技術の活用可能性を探るという意味で一定の成果を挙げたが、生成AIに関する知見が蓄積されつつある段階では、計画的な利活用の検討への移行が必要である。しかし、現時点で精度向上や検証スキームを含めたベストプラクティスが整備されているわけではない。一方で、各企業が業務活用を目指して手探りで検討を進めているにもかかわらず、体系的な方法論の整備を待っては、競合に対し

てサービスやプロダクトの品質が大きく劣後しかねない。

本稿では、効果的かつ効率的な生成AIの活用範囲拡大や現場導入の「羅針盤」として、生成AIのあるべき業務活用推進を実践例と併せて提言する。生成AIの業務活用推進は、そもそも先進技術を業務へ活用するための活動であるため、デジタルトランスフォーメーション（DX）やデータサイエンスの活用推進と同様に、適切な業務領域の設定と先進技術による付加価値の実現の両面から検討する活動である。したがって、機械学習の予測モデルを活用したツールなどの検討で、机上での実現性検証を経て現場での導入効果をPoC<sup>※1</sup>で見定めたのと同様に、生成AIの業務活用推進においても同様のスキームに沿って検討していくことが必要である。

## 2 生成 AI のあるべき業務活用推進

本章ではNRIが自社ならびに顧客向けに生成AIの利活用検討を推進するうえで見えてきた「一つのあるべき業務活用推進の考え方」を提示する。なお、本稿で提示する範囲は、生成AIによるプロダクト

※1 PoC（Proof of Concept：概念実証）とは、新たなアイデアやコンセプトの実現可能性やそれによって得られる効果などについて検証すること

図表 1 あるべき業務活用推進の全体像

推進プロセス	目的	検討事項	タスクの具体例
① 利活用構想の設計	対象業務での活用シーンと活用方法の明確化	<ul style="list-style-type: none"> <li>対象業務の選定</li> <li>ソリューションの設計</li> <li>業務KPIの設定</li> </ul>	<ul style="list-style-type: none"> <li>対象業務の棚卸し</li> <li>業務フローの整理</li> <li>アーキテクチャー図の作成</li> <li>データベースの設計</li> <li>NSMの設定</li> <li>KPIの細分化</li> </ul>
② 要求精度の実現	業務要求に対する技術可能性の見極め	<ul style="list-style-type: none"> <li>精度評価方針の設計</li> <li>検証と改善サイクルの実行</li> </ul>	<ul style="list-style-type: none"> <li>検証項目の設定</li> <li>検証用データの準備</li> <li>出力結果の分析</li> <li>改善案の検討</li> </ul>
③ 業務での試験運用	プロダクト設計や導入効果の精緻化	<ul style="list-style-type: none"> <li>業務KPIのモニタリング</li> <li>定量的な活用効果の推計</li> </ul>	<ul style="list-style-type: none"> <li>モニタリング結果の分析</li> <li>改善案の検討</li> <li>ユーザーアンケート</li> <li>活用効果の算出</li> </ul>

出所) NRI 作成

やサービスが確立された後のユーザー等に対する利活用浸透ではなく、生成 AI の利活用先の実現性検証を起点として、最終的に幅広く活用可能となるツールやシステムを構築する段階までの検討プロセスとする。

### 1) あるべき業務活用推進の全体像

生成 AI の業務活用推進において検討すべき重要な点は、図表 1 に示すような①利活用構想の設計、②要求精度の実現、③業務での試験運用である。これらは、順に進めていくものとなっているが、生成 AI の出力結果を確認しながらアウトプット設計を見直すなど、前後のプロセスを並行して検討を行うなど柔軟な対応が必要である。また、業務での活用方法具体化（≒①利活用構想の設計）→技術的な実現性検証（≒②要求精度の実現）→実業務での PoC（≒③業務での試験運用）という検討プロセスは、生成 AI 特有ではなくあくまで先進技術導入における一般的な流れである。したがって、生成 AI 特有の重要論点を明らかにするのではなく、先進技術導入における重要論点を生成 AI の文脈でどう捉えるべきかという観点での深掘りが必要である。

### 2) 推進プロセス①：利活用構想の設計

利活用構想の設計における目的は、業務での活用シーンと生成 AI の活用方法を可能な範囲で具体化することである。この段階では、日々の業務における課題解決に対して、生成 AI が担うべき機能の整合性を確認することが求められる。主な検討事項は、対象業務の選定、ソリューションの設計、業務 KPI (重要業績評価指標) の設定の 3 点である。業務活用に必要なアウトプットと達成目標を、技術的な実現性検証を行う前に明確にすることで、以降の検討の対象範囲と目標を明確化する。

#### (1) 対象業務の選定

対象業務の選定では、生成 AI 利活用テーマを初期インプットとして生成 AI の活用方法を見極める。例えば、営業業務への適用であれば、顧客への提案ストーリーの作成や会議録の要約などの活用シーンは多岐にわたる。さらに会議録の要約という活用シーンにおいても、参加者ごとの発言内容の抽出や商談における決定事項の整理などの用途の違いによって、生成 AI の実装方式が異なり、対象の活用シーンにおいて、生成 AI が担うべき機能を決めな

ければいけない。

具体的な取り組みとして、対象業務の棚卸しと業務フローの整理を挙げることができる。対象業務の棚卸しでは、検討スコープにおける業務を大まかに分類し、生成 AI 利活用検討における有望度を評価する。営業であれば、顧客訪問準備・商談の内容の記録・日々の情報収集といった分類か、さらに細かい形で業務に対して、各業務の内容と課題を整理することが望ましい。それらに対して、想定効果・生成 AI による出力の実現性・データの取得可否を主要な項目として簡易評価を実施し、対象業務を選定する。業務フローの整理では、対象業務の内容をさらに細分化して、現状の業務プロセスに対して、生成 AI の利活用が実現した際の業務プロセスの変更点を明らかにする。このとき、生成 AI による出力内容とそれに基づく人々の行動まで整理することで、生成 AI の活用仮説が明確化される。

## (2) ソリューションの設計

ソリューションの設計では、対象業務における生成 AI の活用仮説を実際の仕組み（実装）に設計する。生成 AI モデルのアーキテクチャー、プロンプト、参照するデータベースの設計、最終的な出力結果を決定する。ただし、業務活用に必要なアウトプットは、全て生成 AI によって行われる必要はない。例えば、特定処理における条件分岐の基準が定義できる場合は、生成 AI を活用せずに、一定な規則で実装するなど、生成 AI が担うべき機能を仕分けることが重要である。また、この段階で設計されたソリューション案は、次項で述べる「推進プロセス②：要求精度の実現」における精度検証の検討対象であり、検証結果に基づきソリューション案を更新していくことが求められる。

具体的には、ソリューションのアーキテクチャー

図の作成を挙げることができる。インプットするプロンプトや参照データベース、生成 AI や規則での処理プロセス、アウトプットされる出力項目を全体の流れとして可視化しておくことで、実装すべき仕組みと検証対象を明示できる形にする。

## (3) 業務 KPI の設定

業務 KPI の設定では、検討の進行と成功を測る指標を設定する。データ活用などの領域では、検証用データに対する精度が目標指標として設定されることが多い。しかし、特に自然言語を出力する生成 AI の精度検証では、正解や到達目標を明確に定義しづらい。また、仮に 9 割程度の精度を達成していたとしても、その数値自体が実際の業務活用への移行可否の根拠となりえない。これらに対しては、階層的かつ業務的に意味のある KPI の設定がポイントである。

具体的には、図表 2 のようなプロダクトのグロースハック<sup>※2</sup>における NSM (North Star Metric : 北極星指標) を提案する。NSM とは、企業がビジネスを成長させるために、事業の本質的な価値を測定するための指標のことである<sup>※3</sup>。NSM は、事業の最終目標である KGI (重要目標達成指標) を達成するための顧客視点での先行指標として設定される。さらに、NSM から KPI を検討し、顧客体験に根差した目標設定を実現できる。各 KPI は、幅 (アクティブ顧客数)、深さ (顧客体験の深さ)、頻度 (アクション頻度)、効率 (価値を実感するまでの時間)

※2 商品やサービスを急成長させるために、データを集めながら改善を繰り返していく手法のこと

※3 Amplitude 社資料より <https://jp.amplitude.com/blog/how-to-find-the-north-star-metric>

図表2 NSMに基づく業務KPI設計



出所) NRI 作成

の4観点で細分化される。また、生成AIの業務活用検討では、ユーザー利用時と机上検証時に分けてKPIを設定し、各指標を基に精度検証の目標やプロダクト設計の制約を検討する。

例えば、生成AIチャットツールの検討をする場合には、NSMが月間の質問送信数、実運用時のKPIがアクティブユーザー数（幅）や問い合わせの解決率（深さ）、机上検証時のKPIが生成された回答への納得度（深さ）や質問に対する回答精度（効率）などになり、単純な精度以外の指標も含めて目標設定することが可能になる。ただし、NSMと各KPIは、初めから完璧な指標を設定することが難しいため、生成AIの利活用検討を進める中での見直しや、場合によっては顧客やユーザーと議論しながらの合意形成が必要である。

### 3) 推進プロセス②：要求精度の実現

要求精度の実現における目的は、利活用構想の設計で具体化された生成AIの活用方針について、精度評価と改善のサイクルを回し、業務上の要求精度に対する実現性を見定めることである。具体的には、ソリューションの設計に基づきモックアップ<sup>※4</sup>構

築やデータ整備を実施する。その後、生成AIからの出力結果を評価し、アーキテクチャー・プロンプトの改善や生成AIの活用方針見直しを検討することによって、業務プロセスにおける生成AIの適切な位置づけを決定する。

#### (1) 精度評価方針の設計

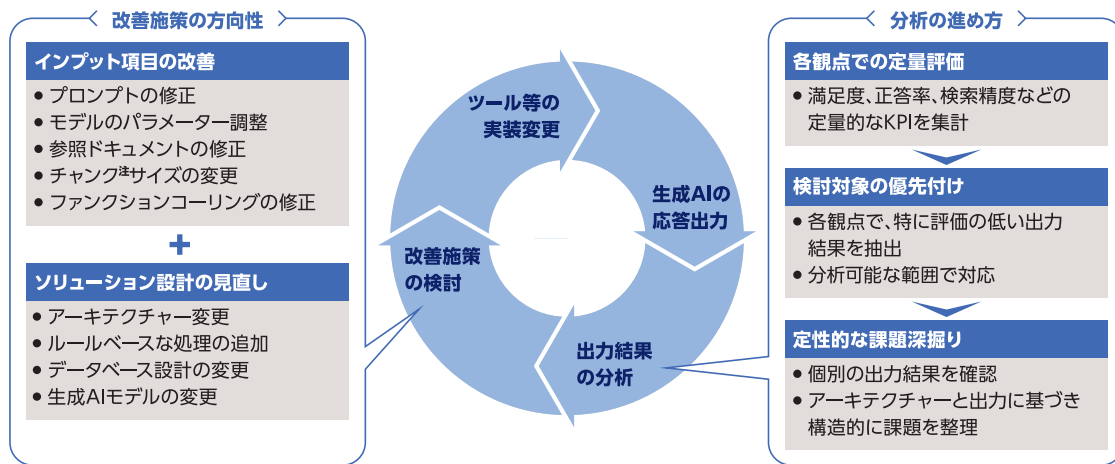
精度評価における指標は、前述のNSMとKPIに関して議論された内容を基に設定される。生成AIの利活用検討においては、入力に対する出力の正しさを示す回答精度とユーザー目線での納得感や満足度の2指標が採用されるケースが多い。このとき、生成AIからの出力が定量的な数値でない場合は、主観的な判断に基づく5段階評価や正誤判定のような形で定量的な数値指標に落とし込む必要がある。

さらに、ソリューションの設計に基づき、アーキテクチャー内で検証する観点を事前に整理し、問題点や改善点を詳細に特定することが可能となる。例えば、ドキュメント検索結果に基づいて回答を生成

※4 試作品のこと



図表3 精度検証における改善サイクルの概念図



注) 文書の区切りの単位のこと  
出所) NRI 作成

するチャットボットの場合は、ドキュメント検索部分と結果の出力部分に検証観点を分けて精度評価を行うことで、正確ではないあるいはユーザーからの納得感が低いと判定された応答に対する課題と対応策を明確にする。

また、検証用データに関しては、必要なサンプル数について明確な基準やルールがあるわけではない。そのため、業務での活用範囲や想定される入力の全体像を仮置きして設問を設計することで網羅性を担保し、可能な範囲で検証用データを準備・収集することが望ましい。例えば、事務手続きに関する問い合わせチャットボットであれば、事務手続きの各プロセスや必要書類など具体的な用途に沿って想定問答集を準備する。

## (2) 検証と改善サイクルの実行

精度検証では、図表3のような生成AIからの応答の精査と改善施策検討のサイクルを回す。まず、事前に整理した各検証観点について、特に評価の低い応答を確認し、その要因を分析し、構造的に生成AIの応答評価が低くなる要因を整理する。前述のドキュメント検索に基づくチャットボットの例であれ

ば、参照ドキュメントが違う・参照ドキュメントの内容が応答に反映されていない・出力された文章や構造が望ましくないなどに課題は細分化される。

対して、取りうる改善施策は、技術観点ではプロンプト・ファンクションコーリング<sup>※5</sup>・生成AIのモデルのアーキテクチャー・データベース設計など選択肢がある程度限られてくる。プロンプトやファンクションコーリングの調整は比較的取り組みやすい一方で抜本的な精度向上にはつながらづらい。一方で、アーキテクチャーやデータベース設計の変更は大きな改善につながる可能性があるが、検証するための実装の手間や事前に検討したプロンプトの再検討につながるリスクなどがデメリットとして存在する。理想的にはアーキテクチャーとデータベースの設計を確定し、プロンプト等の検討に移るという流れになるが、検証の中で出力結果が業務要求を満たさない場合には柔軟に改善策を変更するべきである。

※5 プロンプトで与えられたタスク実行に際してLLMが外部的なAPIや関数を呼び出す機構のこと。プロンプトエンジニアリングと同様に、LLMによって適切に関数呼び出しされるよう調整が必要

また、出力結果を精査・改善する過程では、顧客やユーザーといったビジネス・業務サイドを巻き込んで検討を進めることが望ましい。ビジネス・業務目線での出力結果への評価、生成 AI の業務活用方針の見直しまで含めた検討により、業務上の要求水準の変更にも対応しうる技術的な実装が可能となる。

#### 4) 推進プロセス③：業務での試験運用

業務での試験運用における目的は、利用者に対して試験導入することで、プロダクトやサービスの磨き上げと定量的な効果算出を精緻に行うことである。最終的に、本プロセスを経て、利用者への本格導入やその後の運用に関する意思決定が行われる。

##### (1) 業務 KPI のモニタリング

試験導入に至るまでに精度評価と改善によって一定の精度を実現している状態は望ましいが、完全な作り込みを行うことは難しい。したがって、現場での試験導入においても NSM に基づいた KPI をモニタリングする仕組みを構築することで、ユーザー目線でプロダクトやサービスの磨き上げを行う。試験導入における KPI の例としては、アクティブユーザー数や生成 AI 活用によるタスク完了率や正答率などを指標として挙げるができる。

これらの指標のモニタリングと併せて、定期的にユーザーからの定性的なフィードバックも収集しながら、精度検証プロセスと同様に課題の抽出とプロダクトの改善を実施する。具体的なアクションとしては、特定のケースにおける正答率が低いのであれば、回答の正確性を向上させるためにプロンプト修正を実施することなどが挙げられる。また、ユーザーの入出力結果がデータとして蓄積できているのであれば、その内容をデータベースに取り込んで精度を向上させることも検討の選択肢としてありうる。例

えば、FAQ のような形でデータを保持し、類似と判断された質問に対してはその FAQ を基にした回答を優先的に出力させることも一案である。

##### (2) 定量的な活用効果の推計

試験導入では、将来的なシステム開発や本格導入も見据えて、定量的な活用効果の測定も実施したい。測定する指標は、結果指標でもあるプロダクトやサービスの KGI を対象とするべきである。KGI を測定することによって、そもそもの生成 AI 利活用検討の達成度を確認することができる。業務効率化がテーマであれば、既存業務に対する業務時間の削減効果を業務量調査やユーザーアンケートによって算出する。昨今の生成 AI 利活用検討では、現場での本格導入を優先して本プロセスがスキップされることも少なくはないが、より効果的な生成 AI の利活用を実現するために効果算出を行うことを推奨する。

---

### 3 ユースケースでの適用例：仮想エキスパートツール

本章では 2 章で述べた生成 AI 業務活用推進の考え方に沿って、NRI で取り組んでいる社内文書検索チャットボット（以降、仮想エキスパートツール）の事例を紹介する。これは、NRI のコンサルタントがリサーチ業務を行うにあたり、散在している資料・文献・データ等の収集・整理にかけている労力を削減するためのツールである。例えば、「顧客のプライバシー保護についてどのような取り組みが必要でしょうか」という質問に対して、「自社のビジネスモデルに合わせた自主的なプライバシー保護の取り組み、すなわち…」といったような応答を返す仕組みを想定している。

図表 4 仮想エキスパートツールの画面イメージ



出所) NRI 作成

## 1) 仮想エキスパートツールにおける推進プロセス

### ①：利活用構想の設計

利活用構想の設計における検討項目は、対象業務の選定、ソリューションの設計、業務 KPI の設定の 3 点である。仮想エキスパートツールの対象業務は NRI のコンサルタントが散在している資料や文献から情報を収集・整理するリサーチ業務である。この業務フローと生成 AI による代替領域を選定した後、生成 AI ソリューションを設計、検証時に用いる業務 KPI を設定した。

#### (1) 対象業務の選定

一般的なコンサルタントのリサーチ業務フローは、調査設計、調査実施、結果の整理の 3 ステップからなる。調査設計では、その調査を通じて得たい示唆と調査結果の仮説、必要な情報源と調査スケジュールを設計する。実際のプロジェクトでは、プロジェクトのゴールからブレークダウンして顧客と協議のうえで設計されるため、生成 AI 活用による効率化の必要性和余地は少ない。一方で、調査実施と結果の整理は異なる。

調査実施では、インターネットや公開されているデータベース、書籍など既に整理された情報を調査

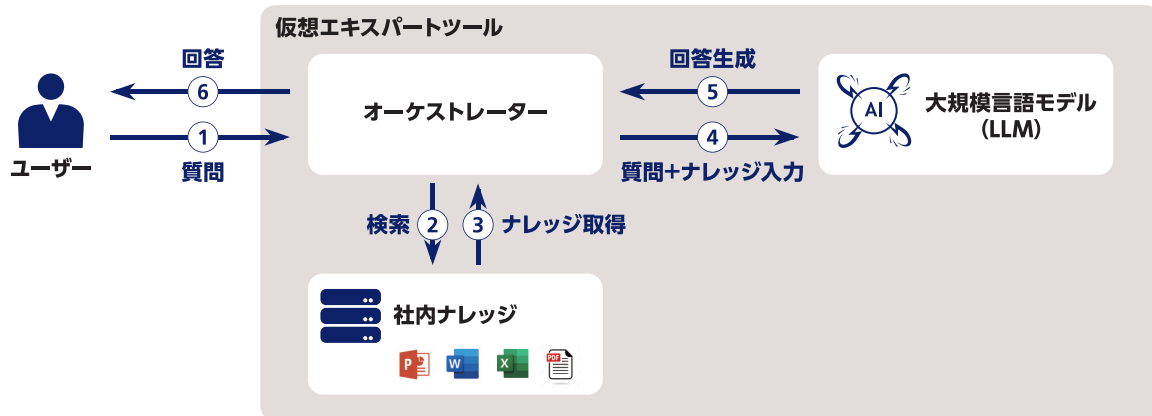
するデスクトップリサーチや、直接的な観察と体験、有識者インタビュー等に基づくフィールドリサーチを行う。広範なデスクトップリサーチを通じて仮説を絞り込み、絞り込まれた仮説について深いフィールドリサーチを行うケースが多い。デスクトップリサーチはコンピューターからアクセスできる情報が大半のため生成 AI 活用が有望視される。

結果の整理は、収集した情報を示唆と合わせて資料化する工程である。リサーチ業務における成果物資料は異なるプロジェクト間でも同一のフォーマットになるケースが多く、こちらもまた、生成 AI や DX 化余地の高い業務である。結果の整理は既定のフォーマットさえあれば人の手による結果の整理も簡潔になることから、より効率化余地の大きい調査実施（特にデスクトップリサーチ）の業務代替を生成 AI 活用仮説とした。

#### (2) ソリューションの設計

ソリューションとして、生成 AI を活用したチャットベースのナレッジ検索ツールを設計した。過去の社内ナレッジ資料を基にした質問回答機能を構築、必要に応じて資料化する機能も付加することによって、リサーチ業務の自動化が実現される。こ

図表 5 仮想エキスパートツールのアーキテクチャー図



出所) NRI 作成

のような資料検索+回答生成の仕組みを Retrieval Augmented Generation (RAG) と呼ぶ。RAG 以外にも、生成 AI を追加学習して内部パラメーターを更新する仕組み (Fine-tuning) も有効だが、コスト効率の観点から本検討のように RAG を採用するケースが多い。

本検討で採用した仮想エキスパートツールのアーキテクチャーを図表 5 に示す。システム全体のフローを管理するオーケストレーター<sup>※6</sup>を通じて、ユーザーからの質問に適したナレッジ文書を検索、質問と合わせて LLM に入力することでユーザーは社内ナレッジに基づく回答を得られる。ナレッジが複数の文書にまたがって存在するケースを想定し、質問と類似度の高い上位五つの文書をプロンプトに加える構成とした。

### (3) ツール導入における業務 KPI の設定

業務 KPI はデータを起点に決定されることが望ましいが、検討の初期フェーズでは仮説として仮置きされることが多い。仮想エキスパートツールの検討においても、初期の業務 KPI ツリーは議論に基づき設定、リリース後にデータを収集・分析、業務 KPI ツリーを更新する運用とした。

仮想エキスパートツールの初期の業務 KPI ツリーを図表 6 のように設定した。対象業務の選定に基づき KGI を月間の文献調査業務の削減時間、NSM を月間質問送信数とした。

NSM に合わせて運用時の KPI と机上検証時の KPI を設定した。特に初期の試験運用で一定のユーザーに繰り返し使ってもらうツールとするため、机上検証の段階から文書のジャンルの広さ（公共や金融系の文書等）や検索できる文書数を逆算して定義した。後述する机上検証ではこれらの条件下で、検索・回答精度と回答へのユーザー満足度が要求水準まで達するかどうかを検証する。

簡易のユーザーアンケートを通じて、机上検証時の KPI に対する要求水準を試算した。専門とするコンサルティング分野の異なる回答者全員に共通した回答は、「検索は 9 割近い一致率の精度が望ましいが回答は低精度でも構わない」であった。最低限、知りたい情報が載った資料が分かれば十分であり多少の回答間違いがあっても資料を直接見るため問題ないという考えである。

※6 複数のコンポーネントやサービス、プロセス間のタスク実行順序やデータフローを管理する機構のこと



図表 6 仮想エキスパートツールの業務 KPI ツリー

KGI	NRI社員のリサーチ業務にかかる時間の削減/月			
NSM	月間の質問送信数			
細分化観点	幅 アクティブ顧客数	深さ 顧客体験の深さ	頻度 アクション頻度	効率 価値実感までの時間
運用時の KPI	<ul style="list-style-type: none"> <li>アクティブユーザー数</li> <li>新規ユーザー数</li> </ul>	<ul style="list-style-type: none"> <li>ツール利用時間/週</li> <li>解決率</li> </ul>	<ul style="list-style-type: none"> <li>アクセス数/週</li> </ul>	<ul style="list-style-type: none"> <li>質問解決時間</li> <li>質問応答の往復回数</li> <li>正答率</li> </ul>
机上検証時の KPI	<ul style="list-style-type: none"> <li>検索できる文書のジャンルの広さ</li> </ul>	<ul style="list-style-type: none"> <li>回答への納得度 (ユーザーアンケート)</li> </ul>	<ul style="list-style-type: none"> <li>検索できる文書数</li> </ul>	<ul style="list-style-type: none"> <li>文書検索精度</li> <li>質問回答精度</li> </ul>

出所) NRI 作成

## 2) 仮想エキスパートツールにおける推進プロセス

### ②：要求精度の実現

要求精度の実現では、仮想エキスパートツールの試験運用を開始する前に、業務 KPI をどの程度実現可能か机上検証を行った。この時点での業務 KPI 達成は必須ではないが、一定の回答精度が達成されていることが望ましい。本節では、仮想エキスパートツールの検討における精度評価の考え方、検証と改善サイクルについて述べる。

#### (1) 精度評価方針の設計

仮想エキスパートツールで採用する RAG の精度評価をするために評価観点と評価方法を決定した。

評価観点は、検索精度と生成精度の二つに大別される。検索精度は、ユーザーの質問に関連する文書を適切に検索できているかどうか（検索文書の関連性）で評価される。この精度が低い場合、適切な情報を LLM のプロンプトに組み込むことができず期待される回答を得ることが難しくなる。生成精度は、生成された回答が質問と関連しているかどうか（回答の関連性）と、生成された回答は文書の情報に基づいているかどうか（回答の忠実さ）とで評価でき

る。これらの精度が低い場合、質問とは無関係の回答や文書とは無関係のうその回答になる懸念が高まる。

評価方法は、事前に人の手で正解データを用意する方法と、その都度人の手で出力を評価する方法、LLM で評価する方法の三つがある。検索精度用の正解データは文書選定によって用意できるが、回答精度用は回答文の作成が必要なため負荷が高い。したがって、本検討においては、回答精度をその都度人の手で評価した。評価には回答の関連性と回答の忠実さについて、5段階の評点をつけた。LLM による評価は人間の評価と同程度の精度で実施できるとする報告もあるが、本検討のデータにも適用できる保証はないため不採用とした。

#### (2) 検証と改善サイクルの実行

事前のユースケースヒアリングで得られた約 30 の質問についてナレッジ検索と回答生成実験を行った。検索精度は質問との類似度が高い上位五つの文書に正解文書が入っているかどうかで評価した。検証結果の一例を図表 7 に示す。

実験の結果、検索精度の改善余地が大きいことが明らかとなった。事前のユーザーアンケートにおい

図表7 検証時の課題と対応方針

	精度	課題	対応方針
検索文書の関連性	73%	質問とは別業界の類似文書が多く検索上位に現れてしまう	【データ整備】 事前に文書に業界カテゴリーを付与。 業界カテゴリーに属する文書群の中から検索
回答の関連性	4.2	文書中の表現をそのまま引用してしまい、質問に対する直接的な回答になっていない	【プロンプトエンジニアリング】 初めにひとことで回答するよう、 回答形式を指示
回答の忠実さ	3.9	文書中に記載のない情報も文書中に書いてあったかのように回答してしまう	【プロンプトエンジニアリング】 文書に記載のある事実と、LLMが独自に生成する事実を分けて回答するように指示

出所) NRI 作成

て「検索された文書を閲覧できれば回答の質が悪くても問題ない」という回答が多かったことから、改善の優先度は検索の方が高い。

検索ミスエラー分析したところ、別業界の類似文書が検索上位に現れるケースが大半であった。そこで、業界別に文書データを整理し、ユーザーには質問時に業界を選択してもらうことで別業界の資料が検索されないよう対応した。実際にプロジェクトで行う調査業務も特定の業界やソリューションに絞った調査が大半である。前述の対応を含む再実験の結果、検索精度は89%まで上昇し要求精度を達成することができた。

### 3) 仮想エキスパートツールにおける推進プロセス

#### ③：業務での試験運用

現在、仮想エキスパートツールはNRIコンサルティング事業本部にて試験運用中であり、利用者から高く評価されている。現場ではリサーチ業務以外にも、新入社員のオンボーディングや事務手続きに関する問い合わせの自動化等にも活用されており、その活用範囲を広げている。改善のサイクルとしては、精度改善と業務KPIの見直しを実施してい

る。精度改善について、図表6に示す運用時KPIをモニタリングしながらユーザーインターフェース(UI) /ユーザーエクスペリエンス(UX)と机上検証時KPI、つまり検索対象文書の拡大と解決率の低い質問への検索・回答精度を高めている。また、業務KPIについてはNSMの変更を実施した。仮想エキスパートツールには、質問の回答結果に加えて検索結果の文書のダウンロード機能が備わっている。ユーザー利用データ分析を通じて、このダウンロード回数が業務の削減時間と相関があることが明らかとなったため、NSMをダウンロード回数に変更した。

今後の展望は三つある。一つ目は、定量的な活用効果の推計である。試験運用を通じて得られたデータを基に、KGIである業務削減効果を精緻化し、本格導入等の意思決定の材料とする。二つ目は、既存業務のさらなる効率化に向けた機能拡張である。本章の1節で挙げたようなリサーチ業務の結果の整理の効率化等が挙げられる。三つ目は、データの拡充による社内ナレッジのさらなる蓄積である。仮想エキスパートツールでは、生成AIを活用した自動資料匿名化と資料共有に対するインセンティブ設計を

検討している。NRI では、これらの取り組みによって、コンサルティング業務の Copilot 構想の実現を志向している。

#### 4 おわりに

本稿では、生成 AI の利活用を進める際の考え方とその具体事例を、仮想エキスパートツールを題材に紹介した。

生成 AI のビジネス活用が進んでいる企業では、「使って試してみる」フェーズから、「本格導入してその効果を確認する」フェーズへと進んでいる。したがって、生成 AI の応答精度だけでなくビジネス効果まで見据えた取り組みを行うために、業務サイドと技術サイドの両面から検証／改善のサイクルを回していくことが今後の生成 AI 活用の焦点になるだろう。生成 AI の利活用が既に進んでいる企業であれば、作成したツールの検証プロセスがビジネスの成功 (KGI) とユーザー体験 (NSM) に結びついているかどうかを確認し、付加価値を向上させるための運用と改善のサイクルを回しているかどうか再考することが望ましい。今後検討を深めていく企業であれば、まずは小規模なユースケースでスモールスタートし、生成 AI や活用推進に関するナレッジを蓄積することが取り組みの第一歩となる。

(監修：松崎 智彦)

●…… 筆者  
坂 拓弥 (ばん たくや)  
株式会社 野村総合研究所  
AI コンサルティング部  
シニアコンサルタント  
専門は、DX、アナリティクス、事業戦略  
立案・実行支援、人材戦略立案・実行支  
援など。直近では、生成 AI の利活用支  
援などに関わる  
E-mail: t2-ban@nri.co.jp

●…… 筆者  
露木 浩章 (つゆき ひろあき)  
株式会社 野村総合研究所  
AI コンサルティング部  
コンサルタント  
専門は、DX、アナリティクス、自然言語  
処理、アプリケーションの開発・導入支  
援など。直近では、CRM データや生成  
AI の利活用支援などに関わる  
E-mail: h-tsuyuki@nri.co.jp