ノート

社内情報を活用する生成 AI の開発

大野　宏*　長谷川　直樹**　土井　康平*　田辺　寛*　小林　豊*
田崎　秀行***　石田　祐介***　武石　紘明***

Research on Generative AI Using Internal Information

OHNO Hiroshi*, HASEGAWA Naoki**, DOI Kouhei*, TANABE Hiroshi* , KOBAYASHI Yutaka*,
TASAKI Hideyuki***, ISHIDA Yuusuke*** and TAKEISHI Hiroaki***

1. 緒　　言

2022 年に公開された OpenAI 社の ChatGPT は，ユーザーが文章で質問すると，AI が自然な文章で回答する手軽さもあり，2 カ月後には登録者数が 1 億人を超えた。その後も AI による自然言語処理技術の進化は著しく，回答精度が向上するとともに画像や音声に対応するものも開発され，用途が広がっている。製造業においても，設計・生産管理・品質保証などの分野で生成 AI の活用が始まっている。工業技術総合研究所は，令和 5 年度の産学官共創ものづくり推進事業「製造業向けエンジニアリング DX に関する調査研究」[1] において，生成 AI の製造業利用に関する調査研究を実施し，令和 6 年度の共同研究事業で当所と（株）WELCON は「社内情報を活用する生成 AI の開発」を行った。開発したシステムは，蓄積された社内データから関連情報を検索し，生成 AI の柔軟な言語処理能力を活用して適切な回答を提供する。本報告では，当該共同研究を実施する中で製造業における生成 AI の有効性を検証し，実業務への適用可能性を評価した結果について述べる。

2. 社内情報を活用する方法

一般的な生成 AI は，公開されている大規模なデータを基に学習しており，幅広い情報に対応し文章などを生成するが，企業の内部情報にはアクセスできないため，社内の特定業務に関する精度の高い回答を出力することは困難である。この課題を解決する方法として，検索拡張生成（RAG：Retrieval-Augmented Generation）やファインチューニングがある。

RAG は，質問に関連するデータを社内のデータベースや文書から取得し，大規模言語モデル（LLM：Large Language Model）がそれを基に回答を生成する仕組みである。例えば，FAQ システムではユーザーの質問に応じて関連する情報を取扱説明書などから取得し，適切な回答を生成する。新しい情報をリアルタイムに反映でき，AI の学習自体をやり直す必要がないことが特徴である。

一方，ファインチューニングは，特定の分野のデータを用いて LLM を追加学習し，専門性の高いモデルを作る手法である。例えば医療分野では，専門論文や診断データを学習することで，より精度の高い診断支援が可能になる。ただし，新しい情報を反映するには追加学習が必要であり，学習データの準備や容量の大きい GPU（Graphics Processing Unit）を搭載した計算機資源の確保などのコストがかかる。

本研究では，業務の効率化と柔軟な情報活用を目的に RAG を活用したシステムを開発した。開発したシステムの概要を図 1 に示す。

3. RAG のしくみ

RAG は，LLM が社内文書やデータベースか

* 研究開発センター
** 下越技術支援センター
*** （株）WELCON

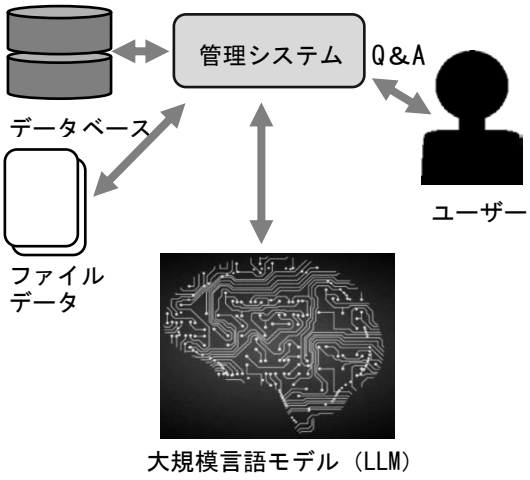

図 1　RAG を活用したシステムの概要

ら関連するデータを検索して参照し，正確な回答を生成する技術である。LLM に一度に大量の文章を読み込ませると回答を生成するまで時間と計算コストがかかるため，RAG では，次の方法で文章を絞り込み，質問と一緒に LLM に渡す。最初に，参照する文章を一定の文字数ごとに分割し，形態素解析を行った上でベクトル化（数値データ化）する。質問も同様にベクトル化し，これらベクトルデータの中から関連度の高い文書を検索（ベクトル検索）し，上位 3～5 件を選定する。次に，これらの文章と質問を LLM に渡して回答を生成させる。生成 AI はまれに間違った答えを生成することもあるため，回答と一緒に参照した文章を表示することで，ユーザーはこれを確認し，回答が正しいか判断できる。

4.　LLM の選定

本研究では文章で質問し回答を得るシステムを開発するため，どの大規模言語モデルを使用するかが重要になる。LLM は ChatGPT に代表されるクラウド型のモデルの他に，自前のパソコンにダウンロードして使用するローカル型のモデルがある。クラウド型のモデルは頻繁にバージョンアップされ性能が向上し続けており，ローカル型のモデルも徐々に性能が向上している。後者は，ネットワークを外部と遮断した状態で使えるため情報漏洩の心配がなく，本研究ではローカル型のモデルを使用した。

ローカル型は様々なモデルが提供されており，経済産業省は「競争力ある生成 AI 基盤モデルの開発 GENIAC（Generative AI Accelerator Challenge）」[2] を通じて開発を支援している。本研究ではこのプロジェクトに採択され，2024 年 6 月に公開された（株）ELYZA のモデル Llama-3-ELYZA-JP-8B[3] を使用した。これは Meta 社のモデル Llama3 を基に，日本語の追加学習を行い開発された日本語が使える 80 億パラメータのモデルで，その性能は OpenAI 社の ChatGPT-3.5 Turbo に匹敵する。このモデルは研究および商業目的での利用が可能で，自前のパソコンにダウンロードして利用できる。なお，モデルのサイズが大きいため 24GB 程度の GPU が必要になる。

5.　取り組み内容

5.1　概要

（株）WELCON は早くから DX 推進に取り組み，データの有効活用を検討している。本研究では，①生成 AI を活用し自社の過去の不適合データを活用して新たな不適合を削減する，②新規見積もり依頼に早く正確に対応するため過去の案件から類似品を検索する，③社内規程に関する質問に回答する，④PDF の請求書から必要事項を読み取り別のファイルに転記する，という 4 種類のプログラムを開発した。

5.2　不適合の削減

過去の不適合データを活用し，生成 AI に不適合の発生の傾向と対策を行わせることで，将来の不適合を予測・未然防止できる。また，新たな不良品の削減やコスト削減が実現し，安定した生産が可能となる。特に，不適合事例の分析は経験と知識が必要であるが，システム化することで誰もが利用可能となる。

（株）WELCON では，過去の不適合事例のデータを保存しており，顧客，部署，製品等も記録している。これらを利用し，同じような製品

を新たに製造する場合，過去に不適合が発生していないか検索して分析し，「このような不具合が発生する傾向があるので，こういう点を気を付けて製造してください。」という注意喚起を行い，同じような不具合の発生を減らすことができる。

例えば，主力製品である熱交換器で発生するリークに関し，不適合事例から傾向と対策を質問すると，次のような回答が得られた。

【質問】多くの不適合発生状況から，発生しやすい不適合の傾向と対策方法を教えてください。

【回答】不適合発生状況を分析した結果，発生しやすい不適合の傾向として，以下の点が挙げられます。

傾向 1： 漏れ試験の工程で不適合が発生しており，特に，A 工程で多く発生しています。

傾向 2： A 工程で，測定結果が規格外の場合に不適合が発生しています。以上の傾向から，対策方法を提案します。

対策 1：不適合工程の改善

A 工程の不適合を減らすために，製造方法の〇〇の改善や，前準備の〇〇の徹底が有効です。

対策 2： 測定結果の管理強化

測定結果が規格外の場合に不適合が発生しているため，〇〇測定の結果の管理を強化し，規格外の場合には製品をリジェクトするなどの対策を取ることが有効です。

5.3 過去類似案件の検索

製造業において，生成AIと過去データを活用し，新規見積もりを迅速かつ正確に誰でも作成できるようにすることは，業務の属人化の解消と標準化を促進できる。多くの中小企業では，経験豊富なベテラン社員しか見積もりを作成できないことが多く，人によるばらつきが課題である。

一般的に，新規見積もりの過去類似案件の検索には図面を用い，画像分類AIを使い類似図面を検索する。今回は，見積もりの打ち合わせの段階で図面がないことが多いため，この段階で顧客と打ち合わせしながら作成する依頼書をもとに，過去に作成した依頼書や手配書を比較し類似案件を検索する。具体的には，どのような製品を製作するか，その概要と仕様を決めるため，これを過去案件から抽出し，ベクトル化してコサイン類似度を計算した後，この値が高い順に，例えば10件を表示する。ユーザーがこれらファイルを開いて内容を確認し，参考にする案件を決定する。多くの過去案件から10件に絞って確認するだけでよく，類似案件の検索が容易になる。実際のファイルでは，依頼書・手配書の仕様が画像で張り付けられていることが多く，自動抽出が難しいため製品の大分類と大まかな概要を記載した客先目的で類似案件を検索した。

5.4 社内規程への回答

生成AIを活用し社内規程に回答するシステムを導入することで，社員が迅速かつ正確に規程を確認でき，業務の効率化と法令遵守の強化につながる。問い合わせ対応の自動化により，総務・法務部門の負担を軽減し，属人化を防ぐことができる。

規程を記載した文章ファイルは内容ごとに分かれており数が多いため，最初に質問からどの規程を参照したらいいか生成AIが判断し，次にその規程を対象にRAGを適用し回答するという二段階方式とした。

5.5 PDF 請求書の読み取り

生成 AI を活用し PDF 形式の請求書から必要事項を自動抽出し，予算管理簿のエクセルファイルに自動転記できると，業務効率の向上と人的ミスの削減が可能になる。また，異なるフォーマットの請求書にも柔軟に対応できるため，経理業務の負担軽減やシステム連携の自動化が促進される。

開発したプログラムは，最初に pypdf などの Python ライブラリを使い PDF 請求書ファイルからテキストを抽出する。次に，生成AIがこれを

読んで日付，金額，請求書発行企業名などの必要な情報を抽出し，エクセルファイルに転記する。

まれに請求書の企業名が画像の場合があり，Python ライブラリ pypdf では請求書から企業名を抽出できない。画像からテキストを抽出する Python ライブラリもあるが，有料の AI-OCR に比べて認識率が低い。そのため，あらかじめ企業名の画像を切り出しておき，PDF の請求書を画像化したものと画像処理ライブラリ OpenCV でマッチングを行い，その度合いが高いものを企業名とした。

6. 回答精度の検証と改良

6.1 不適合の削減

本システムの有効性を検証するため，検索条件を「熱交換器に関する不適合の傾向と対策」と設定し，AI の回答を評価した。主な検証結果は以下の通りである。

・一部誤回答は見られるものの，一般的な回答としては十分に有用である。
・厳密な評価では，AI の回答が一般論や発生工程の理解を前提としていることが確認された。
・具体的な数値（耐圧やリークの不適合に関するデータなど）が不足しており，業務活用にはより詳細な情報が求められる。

本用途の精度向上のためには，次の改善策が考えられる。

・RAG の検索対象となる情報量を増やしたり，質問の仕方を変えたりする。
・不適合発生時の具体的な数値データを含めたデータセットの整備を行い，ファインチューニングを行い，業務特化型 AI モデルを開発し，より実践的な回答を得られるようにする。

6.2 過去類似案件の検索

本システムの精度を評価するために，類似度スコアの分析を実施した。その結果，以下の知見が得られた。

・高い類似度を示すスコアは 0.85 を超える一方で，低いスコアは 0.8 未満であった。
・類似度スコアの差が小さい傾向が見られた（例：0.795 ~ 0.819）。
・検索文字列を変更しても類似度スコアに大きな変化が見られなかった。

本システムの精度向上のため，以下の対策を講じることでさらなる改善が見込まれる。

・検証対象となるファイルのサンプル数を増加させる。
・客先目的の記載内容を統一し，表記揺れを抑制する。
・類似度検索の対象範囲を拡張し，客先目的以外の項目も検索対象とする。

過去の依頼書・手配書からは，客先目的しかテキストとして検索できなかったため，スコアの差が小さく，似たような案件が多くなる。そのため，製品の仕様や特徴を加味して情報量を増やし，より類似度に差をつけ似た案件だけを表示する必要がある。

6.3 社内規程への回答

前節に示した方式で簡単な検証を実施した。具体的には，「育児休業の取り方を教えてください。」と質問すると，育児休業規程が選択され，正しい回答が得られた。一方，取締役会について質問すると，取締役会については定款と取締役会規程に書かれているものの，質問と定款の関連度が低いため，定款に書かれている内容について質問しても取締役会規程を参照し誤った回答を出力する。そのため，例えば，「定款には，取締役，取締役の任期について記載されています。」とキーワードを記載したファイルを作るとともに，これを参照しどの規程を見たらいいか判断し，参照して回答するように変更した。なお，規程を読んでキーワードを自動で抽出するプログラムを Python で作成した。

また，RAG は文章を一定文字数ごとに分割するのが一般的であるが，文字数を変えると回答が変わる場合がある。そのため，文章の区切りの良いところで分割するよう変更した。

6.4 PDF 請求書の読み取り

請求書は発行先企業によって書式が異なるが，LLM は日付，金額，請求書発行企業名などの必要な情報を正しく読み取ることができた。

また，社名が画像として張り付けられている請求書に対しては，画像マッチング機能を使い，正しく認識できた。

7. 結　　言

(1) ローカル型の大規模言語モデルを使い，不適合の削減，類似案件の検索，社内規程への回答，PDF 請求書の読み取りの 4 種類のプログラムを開発した。

(2) 試作したシステムの回答精度を検証し，改良方法を検討した。不適合の削減ではより専門的な知識のデータを用意し RAG を適用する，サイズの大きい最新のローカル型のモデルを利用するなどの方法がある。過去類似案件の検索では，製品の仕様や特徴を追加して比較すると，より精度の高い類似品が検索できる。

(3) 社内規程への回答ではキーワードの追加などの改良に取り組み，90%以上の正答率を達成した。一方，正しい規程を選択したものの誤回答することがあり，その場合は参照した文章を表示し，利用者が読んで確認するなど，実利用を目指し改良を継続する必要がある。

参考文献

1) 小林豊ら, 製造業向けエンジニアリング DX に関する調査研究, 工業技術研究報告書, No53（2024）, pp.95-98.

2) GENIAC,https://www.meti.go.jp/policy/mono_info_service/geniac/index.html,（閲覧 2024-10-10）.

3) ELYZA, https://elyza.ai/news/2024/06/26/elyza-llm-for-jp/,（閲覧 2024-10-10）.