AIを用いた業務改善の実現と効率化について

松本　颯[1]・松本　侑[2]

[1]近畿地方整備局　企画部　情報通信技術課　（〒540-8586大阪府大阪市中央区大手前3-1-41）

[2]近畿地方整備局　企画部　情報通信技術課　（〒540-8586大阪府大阪市中央区大手前3-1-41）.

自営回線網のネットワーク障害対応や資料作成時における複数資料の確認作業は、非常に重要度の高い業務と位置づけられるが、同時に多大な労力が必要になる。これらの業務に対して、生成AIを用いた業務改善の実現の可能性について検証した。検証した結果、ネットワーク障害については従来作業に要する時間が約17時間のものが2分に短縮する結果を確認できた。複数資料の確認作業においても、約70%以上の回答精度を確認できた。ただしトポロジー情報の不足等、さらなる精度向上に向けた課題も判明したため、各種システムの課題と対策案を整理した。

キーワード　業務改善，効率化，AI

1. はじめに

業務改善を検討するにあたって、重要度の高い下記2業務について業務上の課題とその解決案から効率化の実現性を述べる。

1つ目は障害対応業務である。防災官庁である国土交通省にあっては、河川情報や道路情報等災害対応において重要なデータを自営回線網内で通信している。自営回線網にはネットワーク機器と呼ばれるネットワーク内のデータを制御する機器があり、これらの機器に障害が発生した場合、国交省内の通信インフラがストップする恐れがあるため、障害対応は非常に重要度の高い業務に位置づけられる。また1台のネットワーク機器がネットワーク全体を機能不全に陥れる場合もあるため、複数機器のログ解析を行い原因機器を特定する必要がある。この解析作業は保守委託事業者への解析依頼や、特定の専門知識を有する職員個人に依存しており、外部との調整に時間を要することによる通常業務への弊害、初動対応の遅延、解析ノウハウの属人化といった課題を抱えている。

2つ目は資料作成業務である。障害対応は突発的に発生するものであるが、通常業務内では発注資料や事務連絡等資料作成を行う時間が大部分を占める。発注資料を例に挙げると、資料作成時は積算基準、事務連絡、申し合わせ等数多くの内部資料を複合的に確認し作成しなければならず確認作業だけで多大な労力を要する。

上記2つの業務上の課題解決について、生成AIは非常に有用なツールと考える。生成AIはネットワーク機器のログ解析において大量のログから異常や原因を効率的に分析・要約することができる。一方、ログ解析とは一見無関係な複数資料の確認作業においても、文書の意味を理解した検索や要約、自然言語での問い合わせ対応を実現できる。そのため、両者は対象データこそ異なるものの、生成AIを活用することで「必要な情報を迅速かつ分かりやすく取得できる」という共通の効果が得られ、業務効率の向上と利用者負担の軽減が期待できる。

2. 環境説明：導入したAI基盤と主要技術

(1) 近畿地方整備局における生成AI活用の現状

近畿地方整備局では、すでに複数の生成AI基盤が業務利用可能な状態にある。具体的には、商用クラウドサービスであるMicrosoft Copilot Chatと、デジタル庁が推進する政府共通の生成AI利用環境「源内」※1)である。これらは汎用的な文書作成、情報検索、要約等の業務支援において、府省庁横断あるいは庁内全体で広く活用が進められている。

ただ検討当時はMicrosoft Copilot Chatは機密性1情報しか取り扱うことができず、「源内」にいたっては利用不可能であった。上記より、本論文では機密性2情報以上を取り扱うことが可能であるオンプレミス型のローコード開発プラットフォームである「Dify」を選定した。Difyはプロンプトの設定、使用するLLMの選択、RAGで参照するデータベースの指定、入力ファイルの処理、出力形式の指定などを組み合わせ、業務処理の流れを「ワークフロー」として定義しAIアプリケーションを構築できる。表-1にMicrosoft Copilot Chat、「源内」、Difyの各種要件をまとめる。

表-1　業務要件と各基盤の適合性

業務要件	Microsoft Copilot Chat	源内	Dify
(ⅰ)機密性2相当の情報取扱い	△ (検討時)	○	○
(ⅱ)通信途絶時の業務継続	△	△	○
(ⅲ)防災システムとの連携	△	△	○
(ⅳ)拡張性	△	△	○

凡例：○適合、△一定の制限あり

（ⅰ）については、現状機密性2情報以上が全てのAI基盤で取り扱うことができる。(ⅱ)については、近畿地方整備局外のネットワークにMicrosoft Copilot Chat及び「源内」は存在するため、インターネット途絶等の外的なネットワーク障害時に利用できなくなる可能性が高い。一方Difyは近畿地方整備局内に設置されているため、外的要因に左右されにくい利点がある。（ⅲ）についても、Difyについては内部に設置されているため、同様に内部設置されている防災系システムとの連携の容易さを利点として挙げることができる。（ⅳ）について、Difyは後述するRAGと呼ばれる技術により、LLMが学習していない情報を回答として生成することができる。このRAGを利用者自身が管理できること、利用者自身がワークフローを作成できることから他の生成AI基盤より拡張性が高いと整理した。

以上の利点からDifyを用いて、障害対応及び資料作成業務の一部を効率化するワークフローを作成する。障害対応については、ネットワーク機器のログ解析作業、資料作成については、各種資料のチャットボットを作成し、検証を行った。

(2) 主要技術

a) 大規模言語モデル（LLM）

大規模言語モデル（Large Language Model: LLM）は、大量のテキストデータから言語の構造や文脈上の関係を学習し、入力された文章に続く適切な語句を推定しながら回答を生成するAIモデルである。LLMは、単なるキーワード検索とは異なり、利用者の質問意図を踏まえて文章を要約したり、情報を整理したり、手順案を作成したりすることができる。そのため、文書作成支援、問い合わせ応答、情報抽出、分類、データ分析結果の説明など、幅広い業務への適用が期待されている。

b) 検索拡張生成（RAG）

RAG（Retrieval-Augmented Generation）は、LLMが元々学習していない最新の情報や専門的な内部情報を外部データベースから検索し、その内容を基に回答を生成する技術である※2)。業務マニュアル、仕様書、機器の取扱説明書、過去事例などを参照させることで、LLMが根拠の不明確な回答を生成するリスクを抑制し、回答の正確性と信頼性を高めることができる。

3. 検証

(1) ネットワークログ解析ワークフローの評価

a) 検証内容

Difyを用い、アップロードされたログファイルをAIが自動で解析し、要約、時系列イベント、原因の詳細、再発防止・対策案をレポートとして出力する「ネットワークログ解析ワークフロー」を構築した。検証にあたり、過去に実際に発生したネットワーク障害（彦根維持出張所における障害）のログデータを使用し、AIによる解析時間と内容の精度を評価した。この事例では、従来の手法では原因特定までに約17時間を要していた。

b) ワークフローの仕組みと出力内容

本ワークフローは、複数の技術要素を連携させることでログ解析を支援する（図-1参照）。まず、利用者が解析対象のログファイルと質問文を入力する。次に、ワークフローはRAG技術を活用し、あらかじめ外部知識として登録された機器の製品マニュアルや、ログの読み解き方を指示する手順書を参照する。その後、アップロードされた各ログファイルを個別に解析し、最後に複数ログの解析結果を統合して、横断的な観点から原因候補を整理する。最終的に出力されるレポートには、要約、時系列イベント、原因の詳細、再発防止・対策等の提案、まとめが含まれる。

ワークフロー作成にあたっては、解析精度の向上と出力のばらつき抑制のため、単にログ全体をLLMに投入するのではなく、次の工夫を行った。第一に、機器マニュアルに加え、ログの読み方、異常候補の抽出条件、調査順序を記載した手順書を外部知識として登録した。第二に、単一ログファイルごとの一次解析と、複数ログの結果を統合する集約工程を分け、ログ間の関係を段階的に確認できる構成とした。第三に、利用者が入力した事象発生日時を起点とし、その72時間前以降を主な調査対象とすることで確認範囲を具体化した。第四に、ERROR、WARNING、リンクUP／DOWN、電源ON／OFFの短時間反復など、通常とは異なる事象を抽出対象として明示した。第五に、障害発生箇所に近いノードや、接続経路上の関連ノードを優先的に確認するよう指示し、被疑箇所の絞り込みを支援した。これらの工夫により、AIの自由な推測に依存する範囲を抑え、利用者が確認すべきログ上の根

拠を明示しながら、原因候補を整理できるようにした。

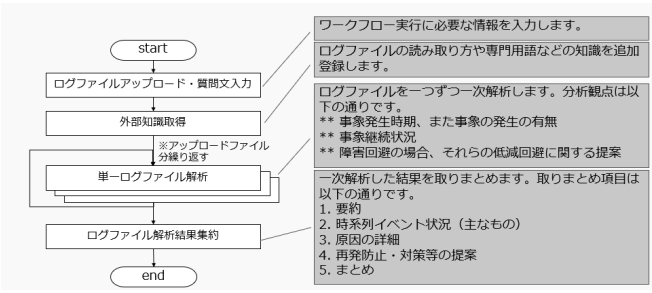

図-1　ログ解析ワークフローの仕組み

c) 結果

本検証では、過去に実際に発生した「彦根維持出張所のネットワーク機器障害」におけるネットワーク障害事例のログデータを対象とし、AIにログファイルと質問文を入力して解析を実施した。本事例において、従来の手法では原因特定までに約17時間を要していたのに対し、本ワークフローは複数のログファイルを横断的に分析し、約2分で解析レポートを生成した。これは従来比で約99%の時間短縮に相当する。

また生成されたレポートは、要約、時系列イベント、原因の詳細、対策の提案を含み、障害状況を把握するうえで必要な情報を備えていた。特に、原因の詳細については、単に結論のみを提示するのではなく、異常が確認されたログの時刻、対象インターフェース、出力値、関連する機器情報を併せて提示しており、利用者がAIの判断根拠を一次情報に基づいて確認できる構成となっていた。実障害の経緯と照合した結果、本ワークフローは原因特定に向けた一次切り分けとして有効であることを確認した。ただし、AIの出力は保守上の最終判断を自動的に代替するものではなく、職員や保守担当者が確認すべき被疑箇所を迅速に絞り込むための判断支援として位置づけることが適切である。

d) 考察及び課題

本ワークフローの導入により、障害発生時に職員自身が迅速に一次切り分けを行えるようになり、初動対応の早期化と復旧時間の短縮が期待できる。また、ログの読み方や異常候補の確認観点をワークフローに組み込むことで、特定の職員の経験に依存していた解析ノウハウを一定程度標準化できる。この点で、専門知識への依存低減と属人化解消にも寄与する。

一方で、本格運用に向けては、トポロジー情報の整備、障害履歴データの整備、監視・検知との連携、生成AI出力の変動への対応といった課題を整理した（表-2）。

表-2　本格運用に向けた課題と対策案1

課題	業務上の影響	対策案
トポロジー情報の不足	ノード間の接続関係が明示されていない場合、複数装置にまたがる障害の真因推定に限界がある。	ノード間の接続関係をトポロジーデータとして整備し、RAGの外部知識に登録する。これにより、接続経路を踏まえた横断的な解析を可能とする。
障害履歴データの不足	過去の類似障害、原因、対処内容がAIから参照できない場合、再発防止策や対応案の具体性が不足する。	過去障害の原因、対処、復旧手順、再発防止策を障害履歴データとして整備し、ワークフローから検索可能にする。
監視・検知との連携不足	障害検知後に人手でログ取得・整形・解析を実行する必要があり、対応の自動化には限界がある。	監視システムのアラートをトリガーとして、被疑装置のログ取得、前処理、ワークフロー解析を連続的に実行する構成を検討する。
生成AI出力の変動	生成AIの回答内容や表現は実行ごとに変動する可能性があり、最終判断をそのまま委ねることは適切ではない。	出力項目を定型化し、根拠ログの提示を必須とする。AI出力は一次切り分け支援として扱い、最終判断は職員または保守担当者が確認する運用ルールを明確化する。

(2) 文書検索ワークフローの評価

a) 検証内容

検証にあたっては、下記の各対象ドキュメントに対して約30問の想定質問および期待回答を基準とし評価を行った。精度評価については学習させた3ファイルごとに行い、それぞれの特性に応じ実務利用に耐えうる指標として「期待精度」の目標値を設定している。

・電気通信設備工事共通仕様書： 目標精度 7割以上
・土木工事共通仕様書： 目標精度 7割以上
・積算担当者会議資料： 目標精度 6割以上

回答の精度は、職員の習熟度や業務への影響度を考慮し、以下の5段階（A～E）で判定した。

評価A： ベテラン職員または有識者と同等の回答レベル
評価B： 回答は正しいが、言い回しが冗長あるいは若干的を射ていない（中堅職員相当）
評価C： 回答は正しいが、重要な情報が一部提示できていない（入社2，3年目相当）
評価D： 回答の方向性は正しいが、若干の誤りやハルシネーション（もっともらしい嘘）を含む
評価E： 明白な誤回答、または業務に支障をきたす重大

なハルシネーションが発生

本検証においては、実務において有効な情報提供がなされていると判断できる「評価AからC」までの合計値を期待精度として算出し、妥当性を検証することとした。また、ハルシネーション等の課題が確認された場合、システムプロンプト等の調整により対応可能なものについては再検証を行い、技術的制約により対策が困難な事案については将来的な対策案として整理した。

b) 結果

性能検証の結果、対象ドキュメントの性質によって回答精度に顕著な差異が認められた。共通仕様書（電気通信・土木）については検証に対し、実用レベルとされるA～C評価の合計は78.3%に達した。これは、当初の目標精度である7割以上を上回る結果であり、構造化されたデジタルデータに対する本システムの高い適合性が実証された。

担当者会議資料については検証に対し、A～C評価の合計は33.3%にとどまった。目標精度である6割以上には届かなかったものの、これは検証期間の制約や、回答の繰り返し間違いやすい難解な質問が集中的に行われたことによる偏りも影響している。

ドキュメント別の詳細な評価結果は以下の通りである（表-3）。仕様書に関しては、ベテラン職員と同等以上の回答（A評価）が25%を超え、具体的かつ正確な引用を伴う回答が生成されている。一方で、会議資料については、過去の古い情報の引用や内容の重複により、情報の網羅性に課題を残す結果となった。

表-3　ドキュメント別評価結果

対象資料	A 評価	B 評価	C 評価	D 評価	E 評価
共通仕様書	26.1%	26.1%	26.1%	0.0%	21.7%
積算担当者会議資料	0.0%	0.0%	33.3%	8.3%	58.4%

c) 考察および課題

本検証の結果、本運用に向けては、検索時の情報の漏れと偏り、複数資料にまたがる横断的読解の不足、網羅的な列挙と時系列検索の限界といった課題を整理した（表-4）。

表-4　本格運用に向けた課題と対策案2

課題	業務上の影響	対策案
検索時の情報の漏れと偏り	担当者会議資料は AI が検索時に重要な箇所を特定しづらい傾向にある。また、資料のページをまたぐ際にチャンクが途切れてしまい、本来存在するはずの情報を「持っていない」と回答するケースが確認された。	単一の AI に全てを委ねるのではなく、役割を分散させた複数の AI エージェントを並列稼働させる。
複数資料にまたがる横断的読解の不足	「災害対策業務協定」のように、複数の年次や資料に分散している情報を抽出する場合、AI が一部の資料のみに言及したり、指示語（これ、あれ等）を多用したりすることで、内容が不透明になる課題がある。	資料の種類ごとに一定件数の情報を均等に抽出し、反復的に探索を行う手法を採用する 。情報が不足している場合には自律的に検索を繰り返す仕組みを構築することで、回答の網羅性と正確性を高める。
網羅的な列挙と時系列検索の限界	「条件に合うものをすべて挙げる」ことや、「最新の順に N 個抽出する」といった、データベース（SQL）のような厳密な条件指定に基づく網羅的な検索が不得手であり、利用者が期待する回答を出力しないケースがあった。	AI に SQL（データベース操作言語）を生成させ、文書のメタデータを格納した構造化データベース（RDBMS）を直接検索する Text-to-SQL 機能を併設する。

4. 結論と今後の展望

生成AIを利用したワークフロー形式のネットワーク機器のログ解析及び文書検索の業務改善の可能性について検証した結果、ログ解析及び文書検索の共通仕様書については利用可能な水準を満たしていることが分かった。ただし、生成AIの出力結果はハルシネーションが発生するリスクを常に内包しているため、業務判断の最終決定は職員自身が行う必要があることには留意すべきである。

今後の展望としては、ログ解析は業務内の一部のみをワークフロー形式で構築したが、全工程を自動化することが考えられる。そのためには、障害検知を行う監視システム等の他の防災システムとの連携が必要不可欠である。文書検索は積算担当者会議資料の精度が低い結果から分かるとおり、対象資料によって精度にばらつきがある。このばらつきをなくし、対策案から精度向上を行うことが展望として考えられる。

また本研究で得られた知見を基に、AIの適用範囲を国土交通省が所管する基幹業務へ拡大していくことも考えられる。例えば、道路・河川の巡視点検報告書の作成支援、監視カメラ映像の解析による異常事象の検出、災害発生時における各地からの被害状況報告の集約と分析な

ど、社会インフラの維持管理に直結する業務への応用が見込まれる。また、将来的には、複数の業務システムや知識ベースを連携させ、AIが職員の指示に基づいて情報取得、整理、判断材料の提示までを一貫して支援する形への発展も期待される。

参考文献

1) デジタル庁: ガバメントAI「源内」, https://www.digital.go.jp/policies/genai.

2) Lewis, P. et al.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, arXiv:2005.11401, https://arxiv.org/abs/2005.11401.