# セキュリティ業務への生成AI活用可能性検証レポート
# ～生成AIはセキュリティ業務にどれだけ役立つのか～

ICSCoE9期
「セキュリティ担当者のための生成AIセキュリティ」プロジェクト

# 改訂履歴

| 版数 | 改訂年月日 | 改訂箇所 | 改訂内容 |
|---|---|---|---|
| 第1版 | 2026年7月31日 | - | 初版発行 |

目次


1 はじめに
- 1.1 本レポートの狙いと全体像
- 1.2 免責事項
- 1.3 著作権および知的所有権
- 1.4 本プロジェクト成果物一覧

2 検証概要
- 2.1 活用可能性検証の流れ
- 2.2 検証対象業務・LLMの選定
- 2.3 LLM・業務プロセス評価指標

3 メール解析
- 3.1 既存業務プロセスと生成AI利用時の想定業務プロセス
- 3.2 検証内容
- 3.3 モデル評価結果
- 3.4 生成AI活用業務プロセス移行の効果と課題
- 3.5 業務プロセスの比較とまとめ

4 Threat Intelligence照合
- 4.1 既存業務プロセスと生成AI利用時の想定業務プロセス
- 4.2 検証内容
- 4.3 モデル評価結果
- 4.4 生成AI活用業務プロセス移行の効果と課題
- 4.5 業務プロセスの比較とまとめ

5 セキュリティ戦略修正
- 5.1 既存業務プロセスと生成AI利用時の想定業務プロセス
- 5.2 検証内容
- 5.3 モデル評価結果
- 5.4 生成AI活用業務プロセス移行の効果と課題
- 5.5 業務プロセスの比較とまとめ

6 まとめ・おわりに
- 6.1 まとめ
- 6.2 おわりに

Appendix

参考文献

# 1 はじめに

# 1.1 本レポートの狙いと全体像

## 本レポートの検討に至った経緯

現在、サイバー攻撃は年々高度化・巧妙化しており、企業が対応すべき脅威は増大の一途をたどっている。これに伴い、セキュリティ担当者の業務負荷は急激に高まりつつある。一方で、専門的なスキルを持つ人的リソースの拡充は容易ではなく、業務効率化は喫緊の課題となっている。この状況を打開する手段として、近年ビジネスシーンでの活用が著しい「生成AI」に注目が集まっている。しかし、生成AIをセキュリティ業務へ導入するにあたっては、以下の課題が存在する。

- **定量的評価の不足:** セキュリティという高い専門性が求められる領域において、生成AIの具体的な業務能力を示す客観的な数値データが不足している。
- **モデル情報の偏り:** 現在、クラウド利用に適した最先端モデルの検証は活発に行われている。一方で、機密性・可用性・完全性が最重要視されるセキュリティ業務においては、ローカル環境や日系モデルの活用ニーズも想定されるが、それらの定量的な評価結果は十分とは言えない。

## 本レポートの目的

本プロジェクトでは、セキュリティ業務における生成AIの性能を多様なモデルで技術検証する。その上で、性能と経営資源（コストや実行時間）のバランスを議論し、現実的な活用指針を提案することを目的とする。

## 本レポートがもたらす効果

本レポートを通じて、読者は以下の効果を得られる。

- **性能の客観的把握:** 対象とするセキュリティ業務において、現時点での生成AIの性能を具体的な数値として確認できる。
- **評価手法の習得:** 性能のみならず、経営資源や実装方式を考慮した、生成AIの検証・評価方法・評価観点を理解できる。

## 想定読者

本レポートは、以下を主な想定読者としている。

- **現場の担当者および監督者:** セキュリティ部門や情報システム部門において実務を担う者。
  本レポートでは特に「メール解析」「Threat Intelligence照合」「戦略修正」という3テーマの業務に対する検証を実施している（各業務の詳細は2章2節にて述べる）。
- **導入推進者:** 自社における生成AIの導入・活用を推進している担当者。

## 1.2 免責事項

- 本書は単に情報として提供され、内容は予告なしに変更される場合があります。
- 本書に記載の内容は、独立行政法人情報処理推進機構および産業サイバーセキュリティセンターの意見を代表するものではなく、筆者らの見解に基づいています。
- 本書の利用によるトラブルに対し、筆者らならびに監修者は一切の責任を負いません。
- 本書に誤りがないことの保証や、商品性又は特定目的への適合性の黙示的な保証や条件を含め明示的又は黙示的な保証や条件は一切ありません。
- 本書に登場する組織・企業名・企業情報はすべて架空のものであり、実在の組織・企業とは一切関係ありません。
- 本書の有効期限は、第 1 版の発行日から1年間とします。
- 本書の内容は第 1 版の発行日以前の情報を基に作成されており、最新のモデルや API のアップデートによって以降の挙動と異なる可能性があります。

## 1.3 著作権およびその他すべての知的所有権

- 本書に関する著作権およびその他すべての知的所有権は、独立行政法人情報処理推進機構産業サイバーセキュリティセンター 中核人材育成プログラム9期生「セキュリティ担当者のための生成AI」プロジェクトに帰属します。
- ただし、本書に含まれる第三者の著作物・商標等は各権利者に帰属します。
- 利用に際しては著作権法で認められている範囲でご利用ください。また、引用の際には出典を明記してください。

# 1.4 本プロジェクト成果物一覧

セキュリティ担当者のための生成AIセキュリティプロジェクトでは、本レポートを含めた以下2つの資料を公開している。「生成AIおよびAIエージェントを安全に活用するための手引書」も、セキュリティ担当者だけでなく、AI導入や運用に関わる多くの方にぜひ読んでいただきたい。

## 生成AIおよびAIエージェントを安全に活用するための手引書

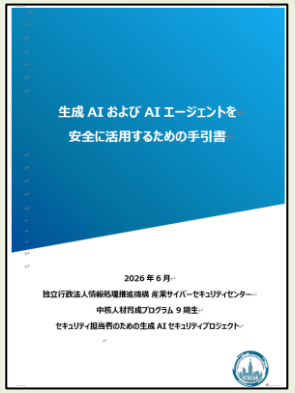
手引書では、企業が生成AIやAIエージェントを安全に導入・運用するために必要なセキュリティ対応事項を、AIシステムの企画から廃棄までのライフサイクル全体を通じて体系的に提供する実践的なガイドである。

同書では、AI導入者・AI管理者・セキュリティ担当者という3つの役割を想定し、統制のために全社レベルで実施すべき事項や個別AIシステムにおけるライフサイクル管理、AIが引き起こす7つの発生被害(リスク)、リスク評価方法、評価結果に基づいた対策検討と残存リスクという構成で整理されている。

各企業が自社のリスクとリソースに応じて、根拠をもって対策を選択・判断できるよう支援することを目指している。

**読んでいただきたい方:**

- 生成AI等を利用する企業の以下担当者
  ・AI導入・管理部門　・セキュリティ部門　・AI活用推進者



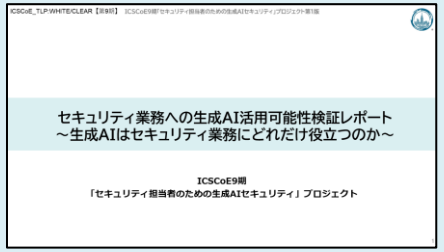

**本書はこちら**

## セキュリティ業務への生成AI活用可能性検証レポート ～生成AIはセキュリティ業務にどれだけ役立つのか～

選定した3つのセキュリティ業務に対する生成AIの性能を多様なモデルで技術検証をした。検証結果をもとに、性能と経営資源のバランスからセキュリティ業務への生成AI活用指針を提案する。

**読んでいただきたい方:**

- セキュリティ部門や情報システム部門において、実務を担う者およびその監督者
- 企業において生成AIの導入・活用を推進している担当者



# 2 検証概要

本章では、セキュリティ業務における生成AIの活用可能性を評価するための、検証の全体概要について説明する。

現状、生成AIの業務適用を検討する際、単に高性能なモデル同士の能力を比較するだけでは、実際の業務環境に即した適切な評価が難しい。業務特性やビジネス要件によってAIの適性が異なる上、導入にあたっては処理時間やコスト、機密情報の管理(クラウドとローカル環境の違いなど)といった実運用上の懸念も存在する。また、モデルの回答精度だけでなく、業務プロセス全体に与える効率化の効果を客観的に測る指標も必要となっている。そこで本検証では、3つのセキュリティ業務と多様なLLMを選定し、性能・コスト・生産性に基づく評価指標を設定することで、生成AIが業務効率化にどの程度寄与するかを多角的に検証した。

## 2.1 活用可能性検証の流れ

本レポートでは、生成AIの業務適用可能性への評価を対象業務・LLM・評価指標の選定から、選定した業務における業務プロセスの整理、AI活用プロセスの設計、技術検証、検証結果評価、業務プロセス評価までを下表の流れで実施した。まず、適用効果を検証しやすい業務と多様なLLMを選定し、性能・コスト・処理時間を含む評価指標を設定した。続いて、現行プロセスの可視化とAI適用後のプロセス設計を行い、実業務データを用いて各LLMの検証を実施した。最後に、検証結果を比較し、生成AIが業務効率化にどの程度寄与するかを評価した。

| | |
|---|---|
| 準備 | **対象業務の選定（2章 2節）**<br>対象業務の選定にあたっては、プロジェクトメンバーで議論し、生成AIの適用可能性や効果を見出すことができる、かつ自社で実施しているセキュリティ業務を抽出した。具体的には、メール解析、Threat Intelligence照合、セキュリティ戦略修正の3業務を選定した。これらは業務特性が異なるため、技術検証結果の方向性に違いが生じることを期待した。 |
| | **対象LLMの選定（2章 2節）**<br>対象LLMの選定にあたっては、単に高性能なモデルを比較するのではなく、様々な企業での導入を想定し、クラウドの高性能モデルと自社環境で動かせるローカルモデル、海外企業モデルと日本企業のモデルなど多様な属性を持つモデルを組み合わせて選定した。性能だけでなく、処理時間、コスト、動作環境や情報管理、日本語適性といった実運用上の観点も総合的に評価できる構成とした。 |
| | **LLM評価指標の決定（2章 3節）**<br>LLM評価指標の決定にあたっては、性能指標であるF1スコアに加え、実業務で重要となる処理時間や消費トークン量などのコスト指標も設定した。これにより、モデル精度だけでなく、生産性や運用コストを含めた総合的な評価が可能となるようにした。また、評価結果を基に業務プロセスでの工数削減効果や想定コストを検討できる構成とした。 |
| 現状把握 | **既存業務プロセスの整理（3～5章 1節）**<br>生成AIの適用範囲を検討するにあたり、まず対象業務の現行プロセスを洗い出し、各工程の役割や作業内容を整理した。特に、脅威情報収集や作業など、人手で実施している判断・分析プロセスを可視化することで、AIが代替・支援し得るポイントを明確化した。 |
| | **生成AI活用業務プロセスの定義（3～5章 1節）**<br>整理した現行プロセスを基に、生成AIの適用によって効率化が見込める工程を特定し、AIが担う処理と人が行う処理を切り分けた。具体的には、脅威情報の収集・照合といった自動化可能な工程をAIに置き換え、担当者はAI出力の妥当性確認や最終判断に注力できるよう、生成AI利用時の業務プロセスを定義した。 |
| 検証・評価 | **検証実施（3～5章 2節）**<br>有効性を検証すべく定義した生成AI利用時の業務プロセスのうち、特にモデルの性能が実現性に直結するタスクについて検証を実施した。検証環境の構築、実業務に即した検証データ作成、プロンプト設計、データ入力といった準備作業を実施した。その後、選定した対象業務ごとに各LLMの技術検証を実施し、処理結果・出力内容・処理時間などのデータを取得した。 |
| | **検証結果の評価（3～5章 3節）**<br>取得した検証結果を基に、対象業務ごとに各LLMの性能・処理時間・API利用時のコストを比較し、モデルの特性と優位性を評価した。特に、F1値による性能評価に加え、平均処理時間や消費トークン量といった生産性・コスト指標を組み合わせることで、実運用を見据えた多角的なモデル評価を行った。 |
| | **定義した生成AI利用業務プロセスの評価（3～5章 4～5節）**<br>技術検証で得られた結果を踏まえ、定義した生成AI利用業務プロセスがどの程度業務効率化に寄与するかを評価した。情報整理、分析・抽出、戦略案作成、承認といった各工程について、生成AIを適用した場合の作業負荷の変化を確認した。作業時間の短縮などの効果を確認するとともに、AIの出力結果に対する人の確認・判断の設計およびAI導入に伴い発生する新たなリスクへの対策など、実業務への適用における具体的な留意事項を整理した。 |

## 2.2 検証対象業務・LLMの選定

表2-1 検証対象業務

| 業務名 | 業務内容 | 参照先 |
|---|---|---|
| メール解析 | 従業員に日々届くメールについて、正規メールと不審メールを識別する | 3章で説明 |
| Threat Intelligence照合 | 日々発信されるセキュリティに関する記事やレポートに対し、自社での対応要否を識別する | 4章で説明 |
| セキュリティ戦略修正 | 外部環境の変化や社内情報の分析結果を受けて、リスクや施策を再評価し、セキュリティ戦略を更新する | 5章で説明 |

表2-2 検証対象LLM

| # | 動作環境 | 開発国 | 公表年 | 本動作環境のコンテキストウインドウ | モデル概要 |
|---|---|---|---|---|---|
| A | クラウド | 米国 | 2026上期 | 1000K（1M）※公表値 | Enterprise版最上位モデル 推論モードあり |
| B | ローカル | 米国 | 2026上期 | 256K | ローカル環境で動作する主要米国モデルAと同系統で推論なし |
| C | ローカル | 米国 | 2025上期 | 256K | ローカル環境で動作する主要米国モデルBと別系統で推論なし |
| D | ローカル | 日本 | 2024下期 | 16K | ローカル環境で動作する主要日系モデル 推論なし |
| E | ローカル | 日本 | 2024下期 | 32K | ローカル環境で動作する主要日系モデルDと別系統で推論なし |

### 検証対象業務の選定

本検証で対象とした業務を表2-1に示す。生成AIの適用可能性を評価するにあたり、セキュリティ担当者の業務負荷が大きく、支援効果が期待でき、かつ適用効果を検証しやすい業務として、メール解析、脅威情報（Threat Intelligence、以下TI）照合、セキュリティ戦略の修正の3業務を選定した。これらは判断業務、情報照合、内容整理といった特性が異なり、生成AIの適性を比較しやすい構成となっている。

メール解析は、メール対策ソフトによる検知メールや従業員からの不審メール報告を担当者が解析し、対応要否を判断する業務である。従業員の報告漏れによる見逃しリスク、高性能な対策ソフトの導入コスト、解析精度の属人化、対応量増加による業務負荷などの課題があり、生成AIによる効率化が期待できることから検証対象として選定した。

TI照合は、日々発信される脅威情報に対し、自社での対応要否を判断する業務である。資産表の更新、脅威情報の収集、自社資産との照合を人手で行うため、担当者の専門性や語学力への依存が大きく、品質と工数にばらつきが生じやすい。また、製品名・バージョン単位での詳細な照合には時間を要し、資産表の記載粒度が不十分な場合は精度低下も起こり得る。限られた人的リソースで継続するには業務負荷が大きいため、生成AIによる支援効果を検証する対象とした。

セキュリティ戦略の修正は、外部情報や社内情報の収集・整理、分析・抽出、戦略案作成、意思決定層の承認といった多段階プロセスで構成される。情報量の多さによる整理負荷、担当者の経験・スキルに依存した分析品質のばらつき、技術的分析を投資判断やロードマップに落とし込む際の高い思考負荷、承認資料の作成工数などの課題がある。生成AIが情報整理や初期案作成を支援することで負荷軽減と品質向上が期待できるため、検証対象として選定した。

### 検証対象LLMの選定

本検証では、モデルの規模（パラメータ数）の大小のみで比較するのではなく、企業における実利用場面を想定し、動作環境・開発国・公表年が異なる多様なLLMを選定した。

対象LLMを表2-2に示す。具体的には、クラウドの高性能モデルに加えて、インターネットと接続せずに自社環境のみで運用可能なローカルモデル、さらに海外モデルと日本企業が開発したモデルを組み合わせることで、各モデルの単体性能だけでなく、動作環境、情報管理、コスト、処理時間、日本語のセキュリティ業務への適用度、導入のしやすさといった実運用上の観点を総合的に評価できる構成とした。また、ローカルで動作させたモデルは同一のハードウェアで動作させたが、ハードウェアスペックの関係からモデルの最大コンテキストウィンドウ未満の環境で検証を実施した。ローカルでの動作環境はAppendix.2に詳細を記す。

# 2.3 LLM・業務プロセス評価指標

表2-3 LLM評価指標

| 観点 | 指標 | 概要・補足 |
|---|---|---|
| 性能 ※1 | F1値 ※2 | • 調和平均と呼ばれ、過検知と見逃しのいずれも評価に含めることが可能な指標<br>• 0:性能低～1:性能高<br>• Recall(見逃しの少なさ)とPrecision(過検知の少なさ)にて計算が可能 |
| | 正解率 ※2 | • 全予測のうち正解した割合を示す指標<br>• 0:性能低～1:性能高<br>• 本レポートでは、「メール解析」業務の解析②にのみ使用する |
| | 人的採点 | • F1値や正解率といった数値指標では評価困難な回答の妥当性等を人間が採点<br>• 本レポートでは、「戦略修正」のみに使用する |
| コスト | 消費トークン数 | • 本来コストは「消費トークン数×トークン単価」によって決定するが、本検証では消費トークン数をコストの代理指標として用いる<br>• トークンの数え方や単価はモデルや環境によって異なる可能性があり、モデル間の単純比較には注意が必要である<br>• ローカル環境の場合はトークン数がコストに直結しないため、機材・設備費用を考慮する必要がある |
| 生産性 | 処理時間 | 各モデルの処理速度を比較する指標 |

※1:検証内容に応じて評価指標を選択
※2:F1値、正解率の算出式およびその詳細は資料末尾のAppendix.3に記載する

## LLM評価指標

本検証では、生成AIをセキュリティ業務に適用できるかを判断するため、モデルの回答精度だけでなく、実運用を見据えた複数の観点から評価を行った。

はじめに、LLM評価の観点として表2-3のとおり「性能」「コスト」「生産性」の3つを設定した。性能については、業務特性に応じてF1値や正解率、人的採点などの指標を用い、見逃しと過検知の両面を考慮した評価を実施した。特にF1値は、PrecisionとRecallの調和平均であり、セキュリティ業務における判断精度を測るうえで重要な指標となる。

次に、コストの評価においては、実際のAI利用料を見積もる上で重要な要素となる消費トークン数を確認した。ただし、トークン数はモデルや環境によって計測方法が異なり、単純比較が困難である。そのため本検証では、ローカルLLMについてもトークン数そのもので比較するのではなく、同等のモデルをAPI利用した場合の推定料金を算定することで、コストをできるだけ近い尺度で比較するようにした。

また、生産性の観点としては処理時間を評価し、モデルが業務フローに組み込まれた際の処理速度を確認した。さらに、これらのモデル評価の結果を踏まえ、生成AIを導入した場合にどの程度の工数削減が見込めるか、またAPI利用料や機材費などのコストがどの程度発生するかについても、一部業務で試算を行った。これにより、単なるモデル性能の比較にとどまらず、業務プロセス全体に与える影響を含めて、生成AIの業務適用可能性を多角的に評価できる構成とした。

## 業務プロセス評価指標

なお、各業務プロセスの効率化については、人手で実施した場合の工数(業務遂行に要する人時)を用いた比較も行った。工数は工程単位で現状とAI利用時を算出し、その増減率から効率化の程度を確認するものである。比較対象は、今回選定した3業務のうちTI照合とセキュリティ戦略の修正の2業務であり、メール解析についてはメール対策ソフトとの費用比較を指標としたため、工数比較の対象外とした。

# 3 メール解析

本章では、セキュリティ業務の一つであるメール解析業務について、生成AIの活用可能性を検証した結果を報告する。

現状のメール解析業務では、メール対策ソフトで検知されたメールや、従業員から「不審である」と報告されたメールを、担当者が解析して判断している。しかし、担当者が業務時間内に解析できる件数は限りがあり、メールが大量に発生した場合にはすべてを網羅して解析することが困難な状況である。また、そもそも従業員が不審メールとして判断できない場合や報告しない場合も考えられる。そこでメール解析を生成AIに代替し、上記の課題が解決可能かを検証した。

メール解析

# 3.1 既存業務プロセスと生成AI利用時の想定業務プロセス

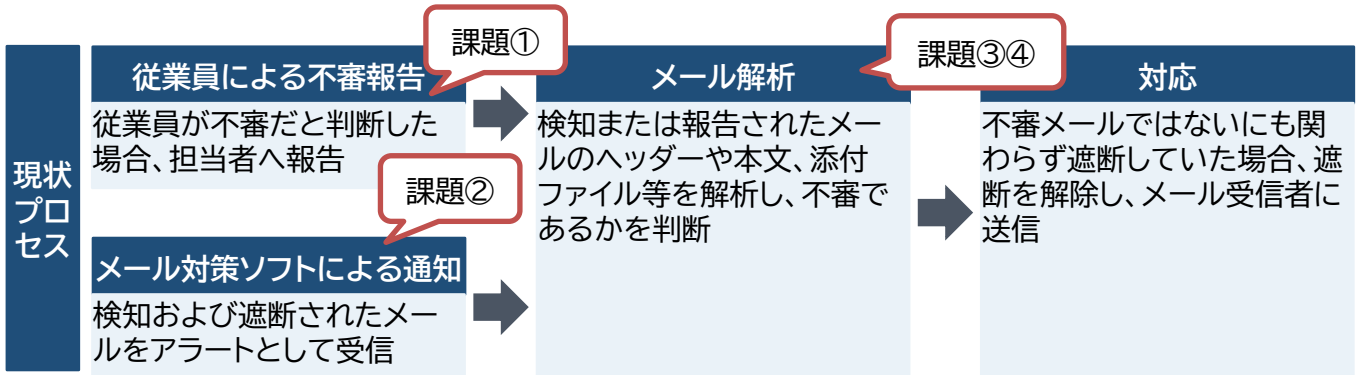



図3-1 現状のメール解析業務プロセス

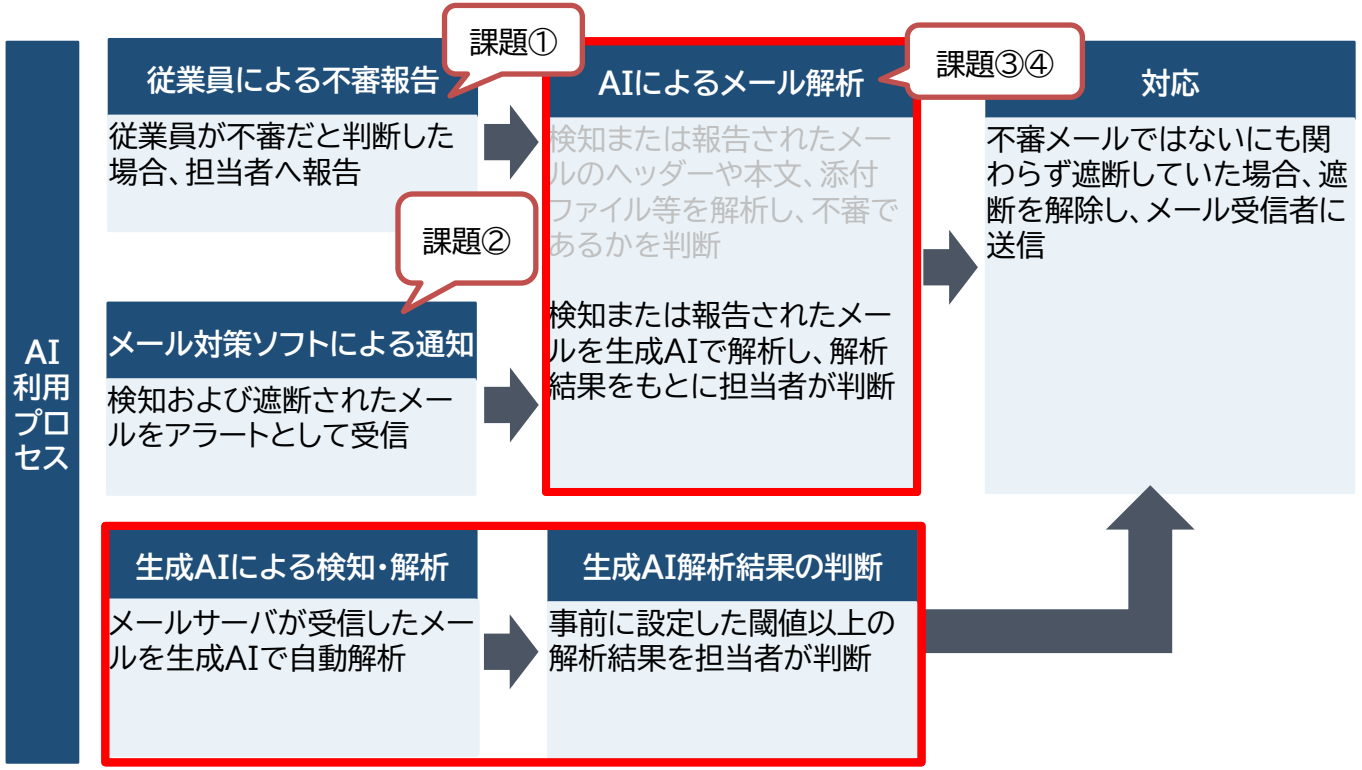



図3-2 生成AI利用時のメール解析業務プロセス

メール解析における業務プロセスは、メール対策ソフトで検知されたメールや、従業員から「不審である」と報告されたメールを、担当者が解析して対応要否を判断し、必要に応じて対応している(図3-1)。

この現状の業務プロセスにおいて、企業は以下4つの課題を抱えている。

- **課題①:従業員によっては不審メールを報告せず、担当者が不審メールを認知できない可能性がある**
  - フィッシングシミュレーションの正しい報告率は18.3%にとどまる[1]
  - 訓練なしの従業員の33.1%がフィッシングに脆弱である[2]
  - 約96%の組織が過去1年に1回以上のフィッシング攻撃を経験している[3]
- **課題②:高性能なメール対策ソフトの費用が増大する**
  - 高度防衛型の場合、月額で数千万程度の費用がかかる[4]
- **課題③:メール解析の分析精度は担当者の能力やリソースに依存する**
  - 大規模SOCでは1日11,000件以上のセキュリティアラートを処理する[5]
  - 中小企業では専任担当者を配置しているのは9.3%、セキュリティ教育未実施は63.6%である[6]
- **課題④:検知または報告件数が大量である場合、検知疲労や解析コストが増大する**
  - SOCアラートの約42%が完全に未調査である[7]
  - アナリストの約70%が精神的に圧倒されると回答しており、米国の年間離職率は約28%である[8][9]

上記の既存業務の課題を解決するため図3-2のとおり、2つのプロセスを生成AIで代替することで従業員や担当者の能力やリソースに依存しない工数削減を実現できるという構成案のもと、検証することとした。

1つ目は、メールサーバと生成AIを連携する構成を想定し、従業員が受信する前に検知させることで従業員の判断能力によらず、高性能なメール対策ソフトに依存しないプロセスが構築できると考える。これを構成案Aとする。

2つ目は、通知後のメールに対して生成AIを利用することで、より不審性の高いメールを識別し、対応に人間のリソースを集中させるプロセスが有効であると考える。これを構成案Bとする。

本章では、生成AIによる不審メールの識別が可能かを性能(F1値)、処理時間、消費トークンで測定し、構成案Aおよび構成案Bの有効性を検証する。

メール解析

# 3.2 検証内容：概要

表3-1 解析①と解析②の概要

| | |
|---|---|
| **解析①** | |
| 入力情報 | 入力-1：通常・不審と判断するための仮想企業情報<br>入力-2：作成した通常メール（250件）<br>入力-3：作成した不審メール（250件） |
| 評価指標 | F1値、Recall、Precision、処理時間、消費トークン数 |
| 構成案Aの判断材料 | 主要 |
| 構成案Bの判断材料 | 主要 |
| **解析②** | |
| 入力情報 | 入力-1：通常・不審と判断するための仮想企業情報（解析①と同じ）<br>入力-4：実際に送付された不審メール（41件） |
| 評価指標 | 正解率、処理時間、消費トークン数 |
| 構成案Aの判断材料 | 補助 |
| 構成案Bの判断材料 | 補助 |

生成AIが構成案Aまたは構成案Bを実務に適用する上で、十分な不審メールの識別能力を備えているかを検証するため、2つの解析を実施した。解析①では仮想企業の業務メールを大量に再現した仮想メール（計500件）で網羅的な精度評価を行い、解析②では実環境で収集された実メール（41件）で実運用相当の挙動を確認する。これら両者を組み合わせることで、データ特性に対するモデルの汎用性もあわせて評価する。

**解析①**

作成した不審メールと通常メールを正しく識別できるかを測定する。本解析では、仮想企業の情報をもとに、検証で使用しないLLMモデルを用いて作成した仮想メール500件を対象とする。入力情報は、表3-1の*入力-1+入力-2*と*入力-1+入力-3*の組み合わせで入力する。解析①の不審メールの識別における評価は、性能（F1値）、Recall（見逃さない割合）、Precision（過検知しない割合）、処理時間、消費トークンをもって実施する。Recallは不審メールを通常メールとして見逃すほど低下し、Precisionは通常メールを不審メールと過検知するほど低下する。また解析①は、構成案Aと構成案Bの主な判断材料として使用する。

**解析②**

実際に送付された不審メールを「不審」と判定できるかを測定する。本解析では、実環境のメールサーバで収集された不審メール41件を対象とする。入力情報は、表3-1の*入力-1+入力-4*の組み合わせで入力する。解析②の不審メールの識別における評価は、性能（正解率）、処理時間、消費トークンをもって実施する。解析②は不審メールのみを対象とした解析であるため、評価指標は正解率としている。なお解析するメール件数が少数であるため、構成案Aおよび構成案Bの補助的な判断材料として使用する。

メール解析

# 3.2 検証内容：データセット

表3-2 解析①と解析②のデータセット

| | 解析① | 解析② |
|---|---|---|
| 共通入力 | **入力-1：通常・不審と判断するための仮想企業情報**<br>自社情報　：ドメイン・会社名・業種・運用サイト　など<br>社員情報　：役職・メールアドレス・メール文体の癖（ですます調など）<br>取引先情報：ドメイン・会社名 | |
| 通常メール | **入力-2：**<br>**作成した通常メール（250件）**<br>社長関連/経理部長関連/取引先関連といった多様なメールを作成した | 使用しない |
| 不審メール | **入力-3：**<br>**作成した不審メール（250件）**<br>仮想企業情報（入力-1）から逸脱したメールを、攻撃手法、ヘッダー情報・検知難易度を掛け合わせて作成した | **入力-4：**<br>**実際に送付された不審メール（41件）**<br>実際に送付されたメールであり、仮想企業情報（入力-1）とは無関係な内容も含まれる |
| 攻撃手法 | BEC（ビジネスメール詐欺）：経営者なりすまし/取引先なりすまし/アカウント侵害　など<br>フィッシング：偽ログインページ誘導/マルウェアの添付/前金詐欺　など | フィッシング：学術論文の投稿勧誘/認証情報窃取/前金詐欺/マルウェアの添付 など |
| ヘッダー情報 | 送信元ドメイン・SPF/DKIM/DMARC認証結果・スパムスコア、受信経路など | |
| 検知難易度 | レベルが高いほど不審としての検知が難しくなると想定<br>レベル1：形式的異常（ヘッダー大幅異常/文体大幅異常）<br>レベル2：軽微な異常（ヘッダー軽微異常/文体異常）<br>レベル3：業務文脈上の矛盾（承認フローの逸脱や組織情報の誤り） | 使用しない |

本検証では、表3-2で示す通り解析①と解析②で異なるデータセットを使用する。

**共通入力**

両解析ともにLLMが「正常な業務メールとして妥当か」を判定するための背景情報として、仮想企業のプロファイル（自社情報、社員情報、取引先情報）を入力する。なお、解析②の入力-4は実環境で収集された不審メールをそのまま使用するため、仮想企業情報と実メール内容が直接関係しないケースも含まれる。

**通常メール**

解析①では社長関連・経理部長関連などのサブカテゴリで構成され、実際の業務で流通しうる正常メールを再現した通常メール250件を使用する（入力-1）。これにより過検知（正常メールを不審と誤判定する）評価が可能となる。解析②では通常メールを使用しない。実環境のメールサーバで収集された実メール41件はすべて不審メールであり、過検知の評価対象データが含まれないためである。

**不審メール**

解析①では仮想企業情報（入力-1）から逸脱したメールを、攻撃手法・ヘッダー情報・検知難易度のレベルを掛け合わせて網羅的に作成した不審メール250件を使用する（入力-3）。解析②では実際に送付された不審メール41件を使用する（入力-4）。実環境で収集された実メールであり、仮想企業情報とは無関係な内容も含まれる。これら両解析を組み合わせることで、「作成メールでの網羅評価」と「実環境での挙動確認」の両面からアプローチの妥当性を検証する。

**攻撃手法**

解析①では経営者なりすまし・取引先なりすましなどの標的型攻撃であるBEC（ビジネスメール詐欺）と偽ログインページ誘導・マルウェアの添付などの非標的型攻撃であるフィッシングの両方を含む幅広い攻撃手法を扱う。解析②はフィッシング系のみを含み、学術論文の投稿勧誘・認証情報窃取・前金詐欺・マルウェアの添付などで構成される。

**ヘッダー情報**

両解析とも、送信元ドメイン・SPF/DKIM/DMARC認証結果・スパムスコア（SCL/BCL/SBRS等）・受信経路など、同一のヘッダー項目を入力する。解析①では 実環境を想定して合成したヘッダー、解析②では実環境で実際に観測されたヘッダーを使用する。これにより、作成メールと実メールの差異がモデル性能に与える影響の比較・分析を可能にしている。

**検知難易度**

解析①は作成メール固有の設計としてレベル1（形式的異常）・レベル2（軽微な異常）・レベル3（業務文脈上の矛盾）の3段階を設定する。レベルが高いほど生成AIによる検知が難しくなる想定であり、モデルが検知可能な難易度を段階的に評価する。解析②は実メールであり、難易度ラベルは付与されない。

メール解析

## 3.2 検証内容：入力設計

| ①入力するメール情報 | |
|---|---|
| minhd（minimal header） | ヘッダー情報のうち、メールアドレス・件名・日時のみ |
| head | ヘッダー情報全て |
| body | 本文のみ |
| full | ヘッダー情報全て・本文 |

✚

| ②出力形式 | |
|---|---|
| zs (zero-shot) | 指定なし（自由形式） |
| st (structured) | 構造化出力 |
| cot (Chain-of-Thought) | 段階的推論 |

✚

| ③企業情報の有無 | |
|---|---|
| wc (with context) | あり |
| nc (no context) | なし |

図3-3　入力プロンプトパターンの組み合わせ

図3-3に示すとおり、検証時の入力プロンプトは「①入力するメール情報」、「②出力形式」、「③企業情報の有無」の3要素を組み合わせた計24パターン（*4×3×2*）を網羅的に評価し、メール識別精度への影響を比較した。各組み合わせは「body-zs-wc」のように略記する（例: body-zs-wc は「①メール情報は本文のみ、②出力形式は指定なし、③企業情報あり」を意味する）。

**①入力するメール情報**

生成AIに入力するメール情報の粒度・範囲を4種類に分けて検証する。メール検出に必要な情報がヘッダーに依存するか、本文に依存するか、両方が必要かを切り分けることで、実運用で必要な情報量を見極められる。minhd（minimal header）は、ヘッダー情報のうち、メールアドレス・件名、日時のみの最小メタ情報である。headは、認証結果・経路情報・スパムスコア等を含むヘッダー情報全てである。bodyは、本文のみの情報である。 fullは、ヘッダー情報全てと本文の全てのメール情報である。通常、メール対策ソフトはヘッダー認証情報を主に用いて機械的に判定を行うため、本検証では生成AIが本文の理解を加えることで識別精度がどのように変化するかを評価する。

**②出力形式**

生成AIへの出力指示の方法を例示や思考手順を含めない自由形式のzs（zero-shot）、JSON形式や指定フォーマットで判定結果と根拠を構造化して出力させるst（structured）、生成AIに「ステップごとに考えてから結論を出す」ように指示するcot（Chain-of-Thought）の3種類に分類し、指示方法の違いが識別精度に与える影響を検証する。これら3種は、標準的なプロンプト戦略であり、モデルや業務によって最適な指示方法が異なると考えられる。

**③企業情報の有無**

仮想企業情報（入力-1）をプロンプトに含めるwc（with context）と含めないnc（no context）の2種類で検証する。企業情報を与えることで業務文脈上の矛盾を含む不審メールの判定材料が増加する。一方で、プロンプトが長くなりコスト・処理時間が増加することが想定される。仮想企業情報を含めることでの検知精度向上と処理コストのトレードオフを評価する。実環境では、企業情報をどこまで詳細にメンテナンスするかが運用負荷に直結するため、精度向上と運用コストのバランスを見極める観点で重要な軸となることが考えられる。

上記3要素の選択肢は、実運用で考えられる主要な変動要素を網羅することを想定している。特定のパターンのみ評価すると、運用条件が変わった際の精度低下リスクを把握できないため、本検証では全ての組み合わせを評価した。

メール解析

# 3.3 モデル評価結果

表3-3 解析①検証結果(★は最高F1値)

| モデル | 最良入力プロンプトパターン | 性能 | | | 処理時間(秒/件) | 消費トークン数(トークン/件) |
|---|---|---|---|---|---|---|
| | | F1値 | Recall | Precision | | |
| クラウドLLM[A](海外) | body-zs-wc | 0.990 | 1.000 | 0.980 | 7.5 | 1377 |
| ローカルLLM[B](海外) | body-zs-wc | 0.994★ | 1.000 | 0.988 | 5.5 | 733 |
| ローカルLLM[C](海外) | full-zs-nc | 0.970 | 0.968 | 0.972 | 5.0 | 654 |
| ローカルLLM[D](日本) | body-cot-wc | 0.818 | 0.708 | 0.967 | 11.5 | 1361 |
| ローカルLLM[E](日本) | body-st-wc | 0.795 | 0.660 | 1.000 | 6.8 | 978 |

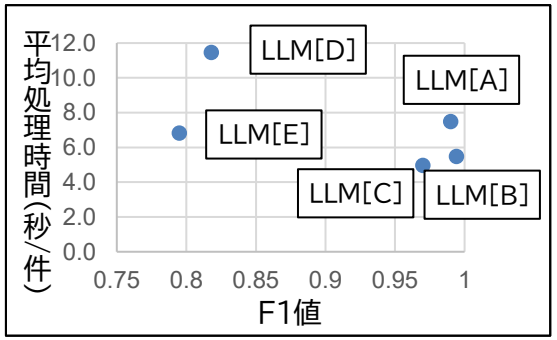



図3-4 解析①・性能評価×生産性評価

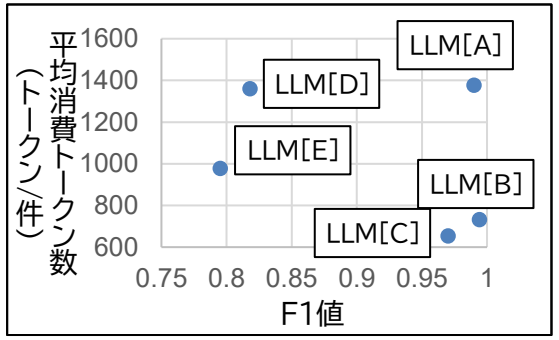



図3-5 解析①・性能評価×コスト評価

表3-4 解析②検証結果(★は最高F1値)

| モデル | 最良入力プロンプトパターン | 正解率 | 処理時間(秒/件) | 消費トークン数(トークン/件) |
|---|---|---|---|---|
| クラウドLLM[A](海外) | head-st-wc | 0.902 | 6.9 | 2026 |
| ローカルLLM[B](海外) | head-st-wc | 0.878 | 6.4 | 1239 |
| ローカルLLM[C](海外) | minhd-zs-wc | 0.927 | 4.1 | 636 |
| ローカルLLM[D](日本) | head-zs-nc | 0.976★ | 4.5 | 733 |
| ローカルLLM[E](日本) | head-st-wc | 0.927 | 5.8 | 1360 |

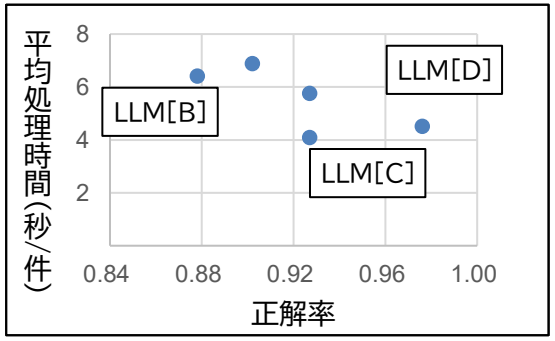



図3-6 解析②・性能評価×生産性評価

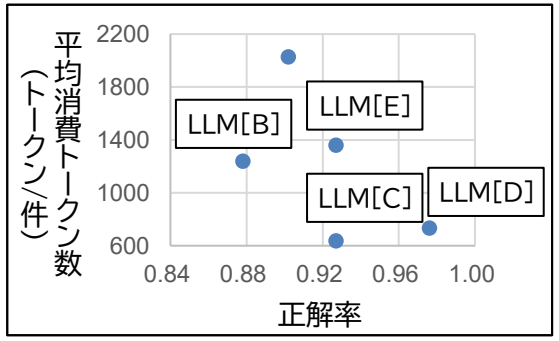



図3-7 解析②・性能評価×コスト評価

解析①②の検証結果は次のとおりである。各モデルの最良入力プロンプトパターンによる検証結果を、表3-3(解析①、評価指標: F1値)と表3-4(解析②、評価指標: 正解率)に示す。また、両結果に基づき、F1値と処理時間の関係を図3-4、図3-6に、F1値と消費トークン数の関係を図3-5、図3-7に示す。なお、クラウドLLM[A]のみ推論モードで動作させた。

**■解析①の考察**

海外モデル3種がいずれもF1値0.95以上となり、高い識別性能を示した。クラウド最上位モデル(LLM[A])とローカル上位モデル(ローカルLLM[B])の間で性能差はなく、いずれも高水準(F1≧0.99)で識別可能であった。また、概ね企業情報あり(wc)の方が企業情報なし(nc)よりF1値が高くなる傾向であった。特にLLM[B]は最高性能と比較的高い処理効率を両立し、性能・効率のバランスが最も良好であった。LLM[C]は処理時間・消費トークン数が最少であり、速度やコストを重視する場合に有力である。最良の組み合わせにおける低下の主因は通常メールを不審と誤判定する過検知であり、不審メールを見逃す誤りは少数(LLM[A]およびLLM[B]では0件)であった。

**■解析②の考察**

対象データや評価指標が解析①と異なるため単純比較ができないものの、最高評価となったモデルは解析①と異なる結果となった。5モデル中3モデルにおいて、ヘッダーのみ(head)+企業情報あり(wc)の組み合わせが最高値となった。LLM[D]は正解率が最も高く、平均消費トークン数および処理時間も比較的短かった。LLM[C]はLLM[E]と同等の正解率で、平均消費トークン数および処理時間が最も少なかった。

**■解析①と解析②の比較**

最良モデルがクラウドLLM[A]とローカルLLM[D]で逆転し、最適入力はbody/fullからheadへと変化した。解析②で使用した実際に送付された不審メール(入力-4)は学術・専門的知識に関するメールが多く、仮想企業情報との関連が低いメールもあったため、本文全体よりヘッダー情報のほうが判断根拠として有効となった可能性が考えられる。また、仮想企業情報の質や採用するメール情報によって最適なモデルが変化、またはモデルの性能が十分発揮できない可能性があると考えられる。解析②は不審メール(入力-4)のみを対象としており、通常メールを誤って不審と判定する過検知の評価ができないため、実運用においては通常メールも含めたF1値による総合評価が必要である。

検証結果のまとめとして、下記2点を記す。

- 求める精度、処理量、利用環境に応じて、適切なモデルを選択する必要がある。解析①では海外モデルが高精度であったが、解析②ではローカルLLM[D](日系)が最高値となり、データ特性によってモデルの優劣が逆転する場合がある。
- 性能だけでなく、コスト・処理速度・運用負荷も含めて実環境で総合的に評価した上でモデルを選定することが重要である。導入後の業務フロー・性能維持/向上・データ保護・システム設計も含めた包括的な検討が必要である。

メール解析

# 3.4 生成AI活用業務プロセス移行の効果と課題

表3-5 構成案Aにおける不審メール識別コスト比較

| 製品・モデル | 性能 | 費用 |
|---|---|---|
| メール対策ソフト<br>（参考値・比較用） | 見逃し率 0.1%以下<br>過検知率 0.1%以下<br>※製品仕様より | 利用料<br>160,000円（16円/日・PC1台と仮定し計算） |
| LLM[A]（API） | 見逃し率 0%<br>過検知率 2%<br>※解析①結果より | API利用料<br>約520,000円/日 |
| LLM[B]（API） | 見逃し率 0%<br>過検知率 1.2%<br>※解析①結果より | API利用料<br>約19,000円/日 |
| LLM[B]（ローカル） | | 機材費（12時間で全メールを処理する仮定）<br>約4,000万円（50台）＋予備台数 |

※見逃し率：不審メールを通常メールと判定することであり *1-Recall* で算出。
過検知率：通常メールを不審メールと判定することであり *1-Precision* で算出。

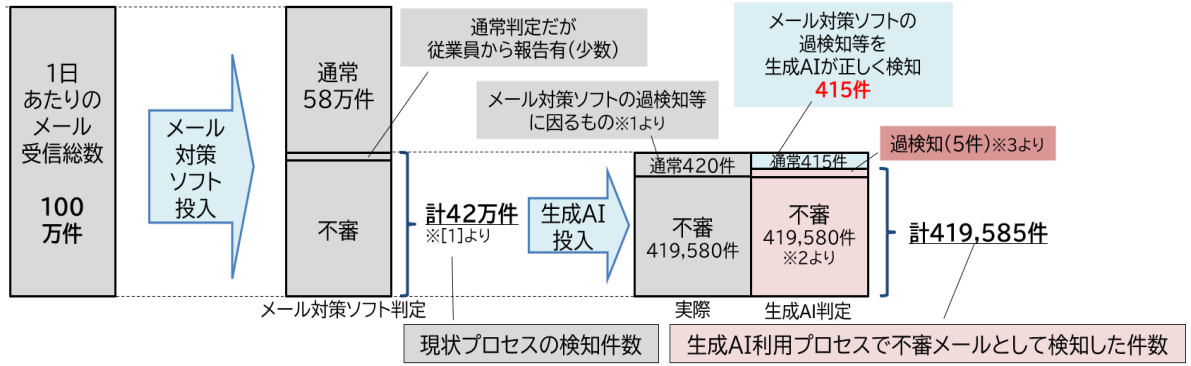



図3-8 構成案Bにおけるメール検知件数の試算

表3-6 構成案Bにおける不審メール識別コスト比較

| 製品・モデル | 性能 | 費用 |
|---|---|---|
| LLM[B]（API） | 見逃し率 0%<br>過検知率 1.2%<br>※解析①結果より | API利用料<br>約8,000円/日 |
| LLM[B]（ローカル） | | 機材費（12時間で全メールを処理する仮定）<br>約1,700万円（21台）＋予備台数 |

※見逃し率：不審メールを通常メールと判定することであり *1-Recall* で算出。
過検知率：通常メールを不審メールと判定することであり *1-Precision* で算出。

本節では、3.3節までに得られた検証結果をもとに、現状プロセスと生成AI利用プロセスにおけるコスト・業務負荷を比較する。比較結果を表3-5（構成案A）、表3-6（構成案B）、構成案Bのメール検知件数試算図を図3-8に示す。あわせて、残存リスクと移行時の課題を整理する。

**■試算条件**

従業員数1万人、1人1台PC貸与、1人・1日あたり受信メール100件を想定する。LLMのAPI利用料は解析①結果の消費トークン数と実行クラウド環境のトークン単価から算出した。

**■構成案Aの効果試算**

表3-5より、LLM[B]のAPI利用料はメール対策ソフトを大幅に下回り、かつ不審メールの認知率が飛躍的に高まる。一方、扱う情報の機密性や処理可用性を加味した動作環境の決定が必要となる。

**■構成案Bの効果試算**

表3-6より、生成AIはメール対策ソフトが過検知した通常メール420件のうち98.8%（=415件）を正しく再判定し、担当者の確認対象から除外できる。これは既存プロセスで不審メールとされた42万件の約0.1%に相当する削減量である。実際に担当者が対応しているメールのうち削減可能な件数が業務効果のポイントとなり、各企業の規模や運用体制に応じた個別検討が必要となる。

**■今後の拡張案**

本検証では不審・非不審の2値判定のみを行ったが、構成案A・Bともに、不審度のスコアリングを導入することで担当者の優先対応をさらに絞り込める可能性があり、より効率的な運用が期待できる。

また、生成AI利用プロセスへの移行に際しては、以下の残存リスクと課題を業務プロセス・システム設計に組み込む必要がある。

**■生成AI利用プロセスにおける残存リスク**

- 過検知により通常のメールまで遮断する構成は、業務への影響度が大きい
- 企業情報の陳腐化や未知の脅威への対応力低下のため、継続的な情報更新・モデル再評価が必要
- 添付ファイルの解析は未検証のため、実運用での精度・トークン数が試算と乖離する可能性あり
- メール本文の長さにより消費トークン数が変動するため、費用・処理時間が試算と乖離する可能性あり

**■生成AI利用プロセスへの移行における課題**

- 最高F1値ではなく過検知・見逃しのバランスを考慮したプロンプト設計とすべき
- 生成AIへの権限付与は最小権限の原則に基づいた設計が必要
- プロンプト入力情報や解析結果の外部流出を防ぐ設計が必要
- 通知量によるアラート疲労対策として脅威の深刻度に応じた通知調整が必要

メール解析

## 3.5 業務プロセスの比較とまとめ

本検証では、メール解析業務を生成AIで代替可能かを評価するため、不審メールの識別精度・処理速度・コストの観点で検証を行った。本検証における結論を表3-7に示す。

表3-7 メール解析の結論

| # | 結論 |
|---|---|
| 1 | 生成AIを利用したプロセスは、不審メールの検知精度の向上や従業員の業務負荷の軽減が期待できる |
| 2 | 生成AIの精度を維持するために、企業情報の整備・定期的な再評価等のメンテナンスが必要となる |

また、生成AI導入前後のメール解析業務を業務品質・スピード・コストの3つの観点で比較した結果を表3-8に示す。生成AI利用プロセスは現状プロセスに対し、いずれの観点においても改善の余地があることが確認された。一方、業務文脈の判断を要する不審メールへの対応には引き続き担当者の関与が必要であり、生成AIによる自動処理と担当者の判断を組み合わせた業務設計が重要となる。

表3-8 メール解析における現状プロセスと生成AI利用プロセスの評価比較

| 評価項目 | 現状のプロセス | 生成AI利用プロセス |
|---|---|---|
| 業務品質(精度) | ・精度が担当者のスキルに依存<br>・業務上の矛盾を含む不審メールは高スキルの担当者でも見逃す可能性あり | ・形式的な異常で検知可能な不審メールは高精度で判定可能<br>・業務文脈の判断が必要な不審メールは担当者の最終判断が必要 |
| スピード | ・1件の判定に数分～数十分かかる | ・大量に処理する場合でも、1件あたり数秒～数十秒で判定可能 |
| コスト | ・人件費が主要コストになる<br>・専門的な知識が必要であり、人材の確保が課題 | ・生成AIのAPI費用、ローカルLLM構築費用が発生するが、人件費の削減効果が期待できる |

### 推奨運用方針

本レポートにおいて推奨する運用方針は次のとおりである。

- 形式的な異常で検知可能な不審メールは生成AIで自動識別し、担当者は業務文脈の判断が必要な不審メールや疑いのあるケースに注力する。このハイブリッド運用を実現するため、プロンプト設計や他システムとの連動を検討する。
- 導入時は**処理速度・機密性・コストを考慮したモデル選定と業務設計**を行うことが重要であり、**運用後も定期的なメンテナンスを行い、精度を維持する必要**がある。
- **不審度のスコアリング導入により**、担当者が優先対応すべきメールを絞り込み、アラート疲労を防ぐ運用とする。
- **運用後は定期的なメンテナンスを継続**し、企業情報の更新・モデル再評価により精度を維持する。

# 4 Threat Intelligence照合

本章では、Threat Intelligence（以下「TI」という。）照合業務について、生成AIの活用可能性を検証した結果を報告する。

TI照合とは、日々発信されるセキュリティに関する記事やレポートに対し、自社での対応要否を判別する業務であり、サイバー脅威に対する組織の防御力を維持するうえで不可欠なプロセスである。近年、確認すべき情報量が増加する一方、サイバー脅威の高度化や変化の加速に伴い、より迅速な判断と対応が求められており、TI照合業務の負担は増大している。

Threat Intelligence照合

# 4.1 既存業務プロセスと生成AI利用時の想定業務プロセス

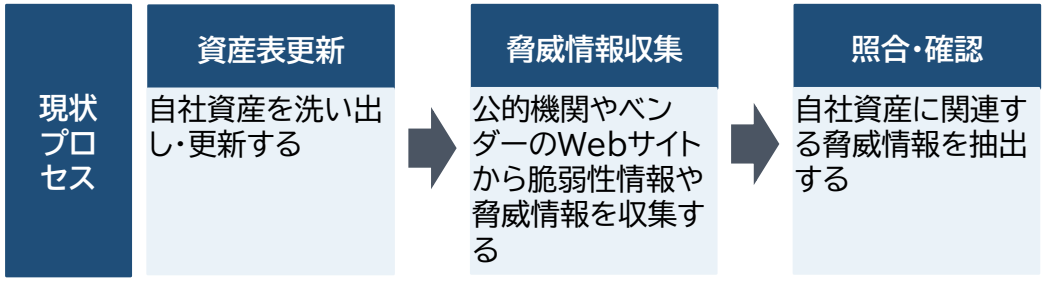



図4-1　現状のTI業務プロセス

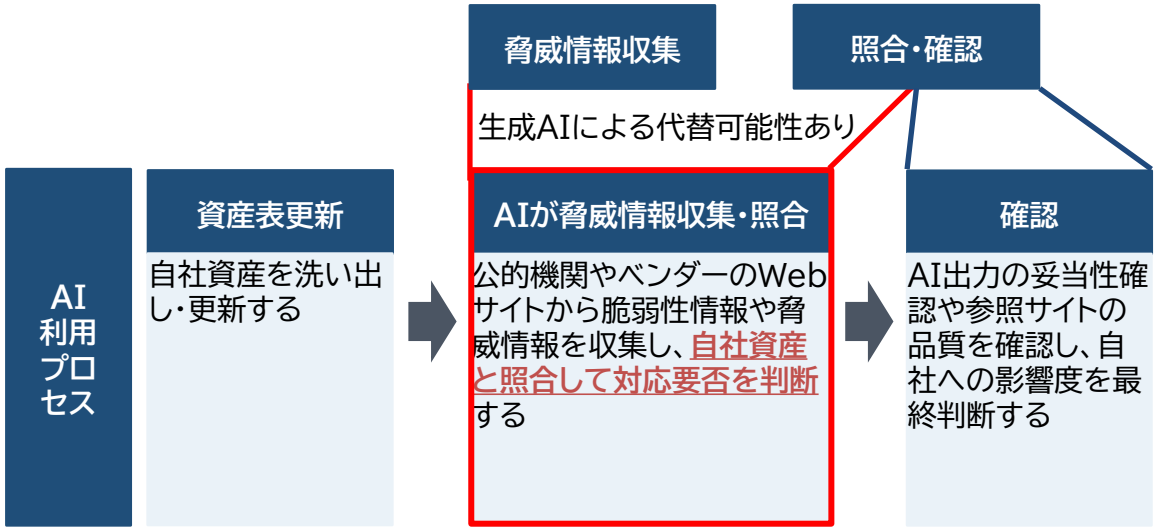



図4-2　生成AI利用時のTI業務プロセス

TIにおける既存の業務プロセスは、資産表の更新、脅威情報の収集、および自社資産との照合を担当者が人手で実施している(図4-1)。照合の結果、該当資産があった場合は後続に対応の検討・対応という業務も発生する。

この業務フローにおける課題は、担当者と資産表という2つの観点に分けられる。
担当者の観点では、専門性・資産理解度・言語力など担当者の能力に大きく依存する。以下3点の課題は、限られた人的リソースのもとで高品質かつ継続的なTIを実現するうえで深刻なボトルネックとなっている。

**課題①**

- 脅威情報の読解品質が、担当者の専門性・言語力・脅威動向への理解に依存する。英語の脅威レポートの読解や、高度な攻撃手法の理解には、相応の専門知識と語学力が求められるため、担当者間で品質にばらつきが生じやすい。

**課題②**

- 自社資産とのTI照合には、製品名・バージョン等の細かい粒度での確認が必要となり、多大な時間を要する。

**課題③**

- 継続的な監視を行うには、日次の確認工数と資産表の更新の業務負荷が大きく、担当者が他業務と並行して十分な時間を確保することが困難である。

また、資産表の質も重要な要素となる。資産表の記載粒度が不十分な場合、精度がさらに低下する恐れがあるためである。

上記の課題を踏まえ、担当者の業務負荷軽減と品質向上を目指し、脆弱性情報収集タスクに対して生成AIの活用を検討した。生成AI活用時の想定業務プロセスを図4-2に示す。

まず、事前に定義した対象サイトからスクレイピング等の手段を用いて新規記事を自動収集する。次に、収集した脅威情報と自社の資産表を生成AIに入力し、対応要否を自動判定させる。担当者はAIの出力結果を確認し、対応要否を判断する。このプロセスにより、「情報収集」という従来の業務負荷を「AI結果の確認」に移行させ、大幅な工数削減が期待できる。

この想定業務プロセスの実現可能性を評価するため、本検証では、特に生成AIの能力が要求されるタスクにおける対応要否判断に対する性能(F1値)、処理時間、消費トークンを測定し、有効性を確認した。

Threat Intelligence照合

# 4.2 検証内容

**A. 検証の目的と方法**

本検証の目的は、LLMが自社資産に関する脆弱性情報や脅威情報を適切に抽出・対応要否を判定できるかの「精度」を評価するとともに、業務迅速化をもたらす「処理時間」、および継続的な運用コストを左右する「消費トークン数」の観点を加えて活用可能性を検討することである。

検証方法としては、脅威情報（TIレポート）と仮想企業の資産表を入力情報とし、仮想企業の資産に直接影響がある TIレポートであるか（対応要否判断）をLLMに判定させ、導入価値を測る指標として、対応要否判断の精度（F1値）、処理時間、およびコスト（消費トークン数）の3点を測定した。

**B. 検証データ（入力情報）**

検証にあたり、以下の2種類の入力データを準備した。

- **仮想企業資産表**
  ネットワーク機器、サーバ、ソフトウェア、クラウドサービスなど計15製品について、ベンダー名、型番、バージョン等の情報を記載した資産表を作成した。
- **脅威情報（TIレポート）**
  TIレポートは合計100件とし、仮想企業の資産との関連有無および言語に基づき、以下の構成とした。
  - 仮想企業の資産に直接関係する脆弱性情報や攻撃情報：50件（日本語25件、英語25件）
  - 仮想企業の資産に直接関係しない脆弱性情報や攻撃情報：50件（日本語25件、英語25件）

※TIレポートは公的機関のWebサイトやベンダー公開情報から収集したものであり、TIレポート自体の正確性については本検証の対象外とする。

【（参考）対応要否の判定基準】

- **「対応必要」とするもの**
  - 自社資産と脅威情報の対象製品・サービスが一致し、対象バージョンまたは影響条件に該当するもの
  - 条件付きの脅威（特定の設定が有効・無効である場合に影響を受けるなど）について、資産表から該当が確認できるもの
- **「対応不要」とするもの**
  - 対象となる製品・サービスを保有していないもの
  - 保有製品のバージョン、提供形態または構成が影響条件に該当しないことを確認できるもの
  - ランサムウェアや標的型攻撃などの一般的な脅威情報であり、自社資産との直接的な関連が確認できないもの
    ※ただし、自社資産を対象とする製品・脆弱性等が記載されている場合は「対応必要」とする

Threat Intelligence照合

# 4.3 モデル評価結果

表4-1 TI検証結果(★は最高F1値)

| モデル | 性能 | | | 処理時間(秒/件) | 消費トークン数(トークン/件) |
|---|---|---|---|---|---|
| | F1値 | Recall | Precision | | |
| クラウドLLM[A](海外) | 0.932 ★ | 0.960 | 0.906 | 16.5 | 7166 |
| ローカルLLM[B](海外) | 0.917 | 1.000 | 0.847 | 41.5 | 6069 |
| ローカルLLM[C](海外) | 0.723 | 0.680 | 0.773 | 14.4 | 5652 |
| ローカルLLM[D](日本) | 0.207 | 0.120 | 0.750 | 85.6 | 5516 |
| ローカルLLM[E](日本) | 0.655 | 0.760 | 0.576 | 49.7 | 5793 |

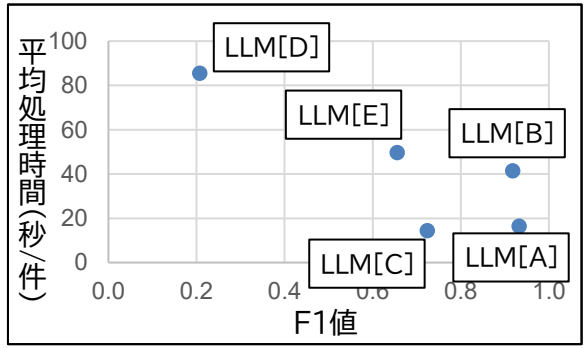



図4-3 性能評価×生産性評価

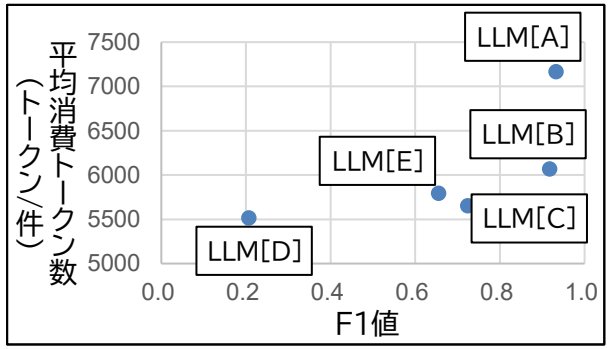



図4-4 性能評価×コスト評価

表4-2 TI記事言語別検証結果

| モデル | 英語記事F1値 | 日本語記事F1値 |
|---|---|---|
| クラウドLLM[A](海外) | 0.962 | 0.902 |
| ローカルLLM[B](海外) | 0.943 | 0.893 |
| ローカルLLM[C](海外) | 0.809 | 0.638 |
| ローカルLLM[D](日本) | 0.077 | 0.313 |
| ローカルLLM[E](日本) | 0.787 | 0.509 |

各生成AIモデルの検証結果を表4-1に示す。また、F1値と処理時間の関係を図4-3に、F1値と消費トークン数の関係を図4-4に示す。なお、クラウドLLM[A]のみ推論モードで動作させた。

**■性能面**

LLM[A]がF1値0.932で最高精度を記録し、次いでローカルLLM[B]がF1値0.917を記録した。この2モデルのF1値は0.910以上であり、他のモデルを大きく引き離す。両モデルともRecallが高く見逃し率が低いことが特徴であるが、LLM[B]はRecall 1.000(見逃しゼロ)を達成した一方、Precisionは0.847にとどまり、過検知が相対的に多い傾向が見られた。その他3モデルのF1値は0.8を下回り、実用的な精度としては難しい結果となった。

**■処理時間**

LLM[C]が14.4秒/件で最速、LLM[A]が16.5秒/件と比較的高速であった。一方、LLM[B]は41.5秒/件、LLM[D]は85.6秒/件と処理時間が長かった。

**■消費トークン数**

LLM[A]が7,166トークン/件で最多、LLM[D]が5,516トークン/件で最少であり、モデル間差異は最大約30%であった。推論モードを使用するLLM[A]はトークン消費が多くなる傾向が確認され、推論過程に伴うトークン消費が影響した可能性が高いと考えられる。

**■記事言語別F1値**

表4-2より、LLM[D]を除く全てのモデルで、日本語記事より英語記事の方がF1値が高い傾向にあった。日系モデルであるLLM[E]でも同様の傾向が認められた。つまり、日系モデルであっても日本語記事のF1値が高いとは限らず、モデル提供元や記事言語だけでは精度を判断できないことが示された。英語との言語相性が良いモデルが多いこと、あるいは英語記事の一般的な構成との相性が良いモデルが多いことが理由として考えられるが、正確な理由の断定には至っていない。

検証結果のまとめとして、下記4点を記す。

- LLM[A]とLLM[B]はF1値が0.9を超え、クラウド・ローカルともに一定程度の精度を持つモデルと確認した。しかし、実際の資産表は本検証で使用した表と比較して資産数が多くなると予想されるため、導入検討時には事前に検証が必要である。
- 処理時間は、1記事あたり少なくとも15秒程度を要し、LLM[A]とLLM[B]の間では約2.5倍の差が生じた。ただし、LLMの実行環境により増減する可能性があることに留意が必要である。
- 消費トークン数のモデル間差異は最大約30%であり、推論系モデルが推論用に確保する領域のためにトークン消費が多くなる傾向が確認された。
- 実運用には**モデルと相性の良い情報源・処理時間を考慮した業務フロー**等の検討が必要である。

Threat Intelligence照合

# 4.4 生成AI活用業務プロセス移行の効果と課題

本節では、前節までに得られたLLMの検証結果をもとに、既存プロセスと生成AI利用プロセスにおける工数を比較する。比較した内容を表4-3に示す。

試算条件として、現状の業務は「担当者1名が毎朝、脅威情報の収集に30分（最大10件程度）、その後の照合作業に30分要する」と設定し、1日あたりの工数を算出した。生成AI利用時の業務プロセスは「脅威情報収集、照合タスクは人手を介さない自動実行で実施」と設定し、工数を算出した。

本検証の直接的な対象である「照合」タスクは、生成AIが出力した結果を担当者が確認する作業のみとなるため、工数が0.5人時から0.25人時に半減する（▲50%）。これにより、脅威情報収集から照合にいたる全体工数は、1日あたりの工数が1人時から0.25人時へ削減され、業務量が約4分の1（▲75%）となる見込みである。

また、本検証の対象外ではある「資産表の更新」「脅威情報収集」タスクについても生成AI利用時のプロセスにおける影響を述べる。資産表の更新タスクについては、生成AI利用プロセスへ移行しても基本的に工数は変わらない想定である。しかし、見直した資産情報をLLMに入力するため、資産情報の選定、加工、マスキング、管理といった追加対応が増加する。脅威情報収集タスクについては、スクレイピング等の手段を用いた夜間など業務時間外での定期実行を想定することで、担当者の工数を0.5人時からゼロに削減できる（▲100%）。

生成AI利用プロセスへ移行することで、人力による情報収集・タスクが不要となり、**担当者は「AIが出力した結果の確認・判断」に注力**でき、工数を**1日1時間から15分へ削減できる検討結果**となった。さらに、**記事数を増加**させることで、**より多くの脅威情報を迅速に把握**し、対策を検討することが可能となる。

表4-3 生成AI利用時のTI業務プロセス工数見積もり ※業務プロセス「照合」を本検証にて実施

| 業務プロセス | 現状の工数（人時） | 生成AI利用時の工数（人時） | 工数増減率（%） | 評価・備考 |
|---|---|---|---|---|
| 資産表の更新 | ー | ー | ー | 検証対象外 |
| 脅威情報収集 | 0.5 | 0（夜間定期実行を想定） | ▲100% | 検証対象外だが、実施を定期実行にすることで工数を削減 |
| 照合・確認 | 0.5 | 0.25（出力結果の確認） | ▲50% | 出力結果の確認のみとなり、15分で実施可能と仮定した場合、工数が半減 |
| 合計 | 1 | 0.25 | ▲75% | 業務量が約4分の1となる |

※表4-3は1日あたりの工数で算出したものである。なお上記の工数に加えて、生成AIが脅威情報を収集する参照先Webサイトの有効性確認や、出力品質維持のためのメンテナンス等の業務が別途発生する点に留意が必要である。

一方、生成AI利用プロセスへの移行に際しては、以下の残存リスクと課題が存在する。実際のTI業務に生成AIを導入する場合は、以下の点を十分に考慮する必要がある。

**■ 生成AI利用プロセスにおける残存リスク**

- 生成AI利用プロセスに移行しても、資産情報の正確かつ定期的な更新は必要不可欠である。資産情報は可能な限り自動更新が望ましいが、手動で更新する場合は、記載粒度にばらつきが生じると分析品質に直結する恐れがある。
- Web検索を伴うプロセスであるため、API費用の高騰や実行時間の増大が懸念される。ただし、生成AIが情報収集するサイトを事前に定義しておくことで、これらの問題を一定程度抑制することが可能である。

**■ 生成AI利用プロセスへの移行における課題**

- 特にローカルLLMで運用する場合、ハードウェア構成等の制約によりモデルが同時実行できるタスク数は1～3タスク程度に限られることがある。そのため、大量の記事を処理するにはバッチ処理などのスケジュール設計が必要となる。
- 自社の資産情報といった機密情報をLLMに入力する必要があるため、情報漏洩リスクに留意し、適切なセキュリティ対策を講じる必要がある。

Threat Intelligence照合

## 4.5 業務プロセスの比較とまとめ

本検証では、TI照合業務の生成AIへの代替可能性を評価し、分析品質を維持したまま工数を削減できる可能性を確認した。結論を表4-4に示す。

表4-4 Threat Intelligence照合検証結論

| # | 結論 |
|---|---|
| 1 | 正確な資産情報を入力することで、脅威情報収集および照合を効率化できる。試算では業務量が約4分の1に削減される見込みである。 |
| 2 | 生成AI利用プロセスに移行しても、人間による出力結果・品質の確認や機密情報の管理、情報源の有効性確認は引き続き必要である。<br>すなわち、人間の役割は消滅しない。 |

また、生成AI導入前後のTI業務を業務品質・スピード・コストの3つの観点で比較した結果を表4-5に示す。

「業務品質（精度）」については、現状プロセスでは担当者のスキルや資産表の粒度に依存するのに対し、生成AI利用プロセスではクラウド・ローカルともにF1値0.9以上の高精度な判定が可能であり、特に英語記事において精度が高い傾向が確認された。

「スピード」については、現状プロセスでは1記事の理解に約3分を要し、英語記事の理解にはさらに時間がかかるのに対し、生成AI利用プロセスでは1件あたり数十秒～数分で判定可能である（ただしWeb検索の時間は別途必要）。

「コスト」については、現状プロセスでは人件費が主要コストとなるのに対し、生成AI利用プロセスでは生成AIのAPI費用やローカルLLM構築費用が発生するが、人件費の削減効果が期待できる。

表4-5 3つの観点（業務品質・スピード・コスト）での比較

| 評価項目 | 現状のプロセス | 生成AI利用プロセス |
|---|---|---|
| 業務品質（精度） | ・担当者のスキル、資産表の質に依存する | ・クラウド/ローカルともにF1値が0.9以上と高精度なモデルが確認された<br>・記事とモデルの相性により精度が変化する可能性が示唆された<br>・引き続き資産表の質に依存する |
| スピード | ・1記事を理解するのに約3分かかる想定とした<br>・判定や英語記事の理解にはさらに時間を要する<br>・他業務もあるため、多くの時間を割くのが困難である | ・自社資産の量にも依存するが、1件あたり数十秒～数分で判定可能である<br>※本検証の対象外であるWeb検索の時間が必要となる |
| コスト | ・人件費が主要コストとなる | ・生成AIのAPI費用、ローカルLLM構築費用が発生するが、人件費の削減効果が期待できる |

### 推奨運用方針

本レポートにおいて推奨する運用方針は次のとおりである。

- 脅威情報収集は、事前に定義した対象サイトから業務時間外（夜間等）に新規記事を自動収集・照合し、その結果を担当者が確認するプロセスを構築する。これにより、処理対象とできる記事数が現状プロセスよりも増加し、業務の質が向上する。
- 高い品質で業務を実施するために、資産情報を正確かつ定期的に更新する体制を整備する。

# 5 セキュリティ戦略修正

現状のセキュリティ戦略（以下「戦略」という）の策定では、対外脅威、社内環境、既存対策、予算、人員、経営方針等、多くの情報を人手で収集・分析し、施策やロードマップに落とし込む必要がある。人手による戦略策定は、企業の実態を反映しやすい一方で、担当者の経験や過去資料の品質に依存しやすく、作成に多くの時間を必要とする。

本章では、この課題に対して生成AIを活用した場合に、業務負荷の軽減、論点整理の効率化、品質の安定化にどの程度寄与するかを検証する。

セキュリティ戦略修正

# 5.1 既存業務プロセスと生成AI利用時の想定業務プロセス

| | 情報整理 | 分析・抽出 | 戦略案の作成 | 承認 |
|---|---|---|---|---|
| 現状プロセス | 対外脅威、規制、業界動向、社内環境、既存対策等を収集・更新 | 事業影響、既存戦略とのギャップ、主要リスク、対応課題等を整理 | 重点施策、KPI、ロードマップ、体制・投資案等を作成 | 意思決定層によるレビューと承認 |
| 課題 | 資料が膨大で、収集・整理に時間がかかる | 担当者の経験・スキルによる分析品質のばらつき、過去資料の品質に依存しやすい | 技術的分析を投資判断・ロードマップに落とし込む負荷が大きい | 承認に向けた資料作成、代替案の提示、修正対応などに工数を要する |

図5-1 現状の業務プロセス

**課題整理**

①戦略の品質は担当者の練度に依存する
②担当者の経験、業界知識、過去の資料への依存が大きく、品質の再現性に課題がある
③戦略の策定に多くの時間を要している

| | 情報整理 | 分析・抽出 | 戦略案の作成 | 承認 |
|---|---|---|---|---|
| 生成AI | 企業情報、既存戦略、脅威情報等を要約・構造化 | 脅威の影響、ギャップ、主要リスクを整理 | 短期施策、KPI、ロードマップ案を作成 | ・代替案、追加の論点、説明材料を提示<br>・承認資料の作成フォロー |
| 人間 | 入力情報の選定、機微情報の除外・マスキング | 出力内容の確認・補正 | | ・承認資料の作成<br>・最終案として修正・承認し、説明責任を負う |

図5-2 生成AI利用時の業務プロセス

現状の戦略策定における業務プロセスを図5-1に示す。

情報整理においては、対外脅威、規制、業界動向等の外部情報と、社内環境、既存対策等の社内情報の収集が必要となる。これらの情報量は膨大であり、収集や整理に時間を要することが課題となる。

情報整理後は、集めた情報を分析し、事業影響、既存戦略とのギャップ、主要リスク、対応課題を整理する。この分析・抽出作業では、担当者の経験やスキル、参照する過去資料の品質によって、分析結果にばらつきが生じやすい。

分析・抽出が完了すると、戦略案の作成を行う。ここでは、重点施策、KPIを整理し、それを実現するためのロードマップ、体制、予算配分などの戦略案を作成する。この作業では、技術的な分析結果等を投資判断に結び付け、実現可能な期間や優先順位をロードマップに落とし込む必要があり、担当者の負荷が大きい。

最後に、経営層等の意思決定層によるレビューと承認が行われる。担当者は承認に向けて、代替案や補助資料を作成し、指摘事項に対応する必要があるため、資料作成・修正にも工数を要する。

生成AIの利用を想定した業務プロセスを図5-2に示す。

本検証では現状の課題に対し、人間が一からすべてを作成するのではなく、生成AIが初期案を作成し、それを人間が確認・補正する業務プロセスを想定した。生成AIの活用は、人間の判断をすべて置き換えるものではなく、初期案の作成や情報整理を支援する位置づけである。そのため、入力情報の選定、実行可能性の判断、経営判断等との整合確認、最終承認は人間が担う必要がある。

**本検証では、上記の「生成AI利用時の業務プロセス」を前提に、生成AIが短時間で作成した初期案が、人間が構築する戦略の質にどの程度近づくかを検証した。業務プロセスを通じて、出力品質、処理時間、消費トークンを測定し、業務における有効性を確認する。**

セキュリティ戦略修正

# 5.2 検証内容：検証の目的と方法・検証データ


図5-3　戦略見直しの流れ

## A. 検証の目的と方法

想定業務プロセスを踏まえ、図5-3のとおり、実際に生成AIが既存の戦略をもとに、突発的な環境変化に対応した実用的な戦略案を考案できるかを検証した。検証方法は次のとおり。

- 仮想企業において、3か年戦略の2年目終了時点で新たな対外脅威が発生し、残り1年の戦略を見直す必要があるという状況を設定した。
- そのうえで、脅威分析・ギャップ分析・主要リスク分析・短期戦略作成を生成AIに実施させ、出力結果を評価した。

## B. 検証データ（入力情報）

検証にあたり、生成AIに与える入力情報として表5-1を設定した。対外脅威（サイバー脅威、地政学リスク、供給制約等）については、業界ごとに1件ずつ戦略見直しが必要となり得る状況をそれぞれ設定し、生成AIがその影響を戦略に反映すべきと判断するか、また、どのように反映するかを確認した。

検証対象とする企業は、表5-2のとおり、5つの業界について仮想企業を設定した。

複数業界を対象とした理由は、業界ごとに事業特性や想定される脅威が異なり、単一業界のみでは生成AIの有効性が特定業界に限定される可能性を十分に確認できないためである。複数業界で評価を行うことで、戦略見直しにおける生成AIの汎用性と、業界固有の特性やリスクにどの程度対応できるかを確認した。

表5-1　入力情報

| 区分 | 入力内容 |
|---|---|
| 企業情報 | 企業概要、事業領域、IT・OT環境、制約条件、BCP、過去インシデント |
| 既存戦略 | 重点施策、ロードマップ、対策状況、KPI、運用体制、投資制約 |
| 対外脅威 | 業界ごとに、戦略見直しを要する可能性が高い新たな対外脅威（サイバー脅威、地政学リスク、供給制約等）を1件入力 |

表5-2　検証企業と想定する対外脅威

| 業界 | 企業概要 | 想定する対外脅威 |
|---|---|---|
| 鉄鋼業 | 世界的に事業展開する鉄鋼事業者 | 海峡封鎖に伴うエネルギー価格高騰、調達の不安定化 |
| 鉄道・交通インフラ | 地方大都市を中心に隣県にも展開する私鉄 | 国家支援型攻撃の公開アラート（注意喚起） |
| エネルギー・電力インフラ | 一地方に展開する一般送配電事業者 | マルウェア潜伏が疑われる不審通信の検知報告を受領 |
| ICTサービス・システム | グローバルにDX・セキュリティなどを提供する企業 | セキュリティ製品の正規更新プログラム経由の攻撃情報 |
| 製造業（機械・電機） | 国内外に製造装置や工場自動化機器の製造、販売を行う事業者 | 周辺地域の情勢悪化による部品供給遅延および便乗したサイバー攻撃情報 |

セキュリティ戦略修正

## 5.2 検証内容：検証手順・評価基準・前提条件

表5-3 指示内容と入力順序

| 入力順 | 指示内容 |
|---|---|
| 1 | 各仮想企業の企業情報・既存戦略・対外脅威を入力、脅威分析の指示 |
| 2 | ギャップ分析の指示 |
| 3 | 主要リスク分析の指示 |
| 4 | 短期戦略の作成指示 |

表5-4 検証条件

| 検証条件（比較の前提） |
|---|
| 5つの仮想企業に対し、同一の指示文を使用 |
| 企業ごとに独立したチャットにおいて、同一順序で入力 |
| 各タスクの出力は、2名の評価者が5段階で評価 |
| 評価者2名の平均値を各タスクの評価値として採用 |

**C. 検証手順**

検証の進め方は、企業ごとにチャット形式で生成AIに指示を入力する方式とした。また、指示のばらつきによる影響を避けるため、各仮想企業に対する指示内容は共通とした。なお、一連の指示は各仮想企業ごとに同一のチャット内で行い、表5-3のとおり段階的に入力した。

**D. 評価基準**

検証条件を表5-4に示す。生成AIの出力は、文章として整っているだけでは十分ではなく、戦略見直しにどの程度活用できるかが重要である。そのため、本検証では、対象業界に知見を持つ評価者1名と、他業界の評価者1名の計2名が評価を行った。

対象業界に知見を持つ評価者が業界固有の事情や前提を踏まえた妥当性の確認、他業界の評価者が前提知識が限定的な読者にも論理構成や説明内容が理解できるかの確認を実施した。これら2つの観点から評価することで、生成AIの出力が業界固有の特性に適合しているか、また実務上の検討資料として分かりやすく利用できるかを確認した。

生成AIに実施させた内容と評価の観点を表5-5に示す。これをもとに、脅威分析・ギャップ分析・主要リスク分析・短期戦略作成の各タスクを5段階で採点した。

**E. 前提条件**

本検証では、ローカルLLMは閉域環境での利用を想定しWeb検索機能を使用せず、クラウドLLMについてはWeb検索機能の利用を制限していない。このため、Web検索の利用可否が出力品質に影響した可能性はあるが、Web検索機能をクラウドLLMの利用環境を構成する要素の一つとして扱い、その影響の切り分けおよび個別評価は対象外とした。

表5-5 実施タスクと評価の観点

| タスク | 生成AIに実施させた内容 | 評価の観点 |
|---|---|---|
| ① 脅威分析 | 新たな対外脅威が対象企業に与える影響を整理 | 妥当性、網羅性、論理性、表現品質、利用効果 |
| ② ギャップ分析 | 既存戦略と新たな脅威に対する不足・不整合を整理 | 既存施策との接続、対応漏れ、論理性 |
| ③ 主要リスク分析 | 優先すべきリスクとリスク許容度を整理 | 事業影響、リスク優先度、許容度の妥当性 |
| ④ 短期戦略の作成 | 残り1年の施策、KPI、半年区切りロードマップを作成 | 優先順位、実行可能性、法的観点、利用効果 |

【5段階評価】 5：ほぼそのまま利用可能／4：軽微な修正で利用可能／3：一定の修正を前提に利用可能／2：大幅修正が必要／1：再作成が必要

セキュリティ戦略修正

# 5.3 モデル評価結果

表5-6 タスクごとの評価結果

| モデル | 脅威 | ギャップ | リスク | 戦略 | 総合平均 | 平均処理時間(秒) | 平均消費トークン数 |
|---|---|---|---|---|---|---|---|
| クラウドLLM[A]（海外） | 4.4 | 4.3 | 4.3 | 4.2 | 4.30 | 101 | 12,378 |
| ローカルLLM[B]（海外） | 4.0 | 3.8 | 3.7 | 3.8 | 3.83 | 490 | 11,232 |
| ローカルLLM[C]（海外） | 3.3 | 3.0 | 3.2 | 3.0 | 3.13 | 149 | 10,295 |
| ローカルLLM[D]（日本） | 3.5 | 3.6 | 3.2 | 3.2 | 3.38 | 984 | 14,298 |
| ローカルLLM[E]（日本） | 2.9 | 2.6 | 2.0 | 2.9 | 2.60 | 229 | 10,267 |

【評価】5:ほぼそのまま利用可能／4:軽微な修正で利用可能／
3:一定の修正を前提に利用可能／2:大幅修正が必要／1:再作成が必要

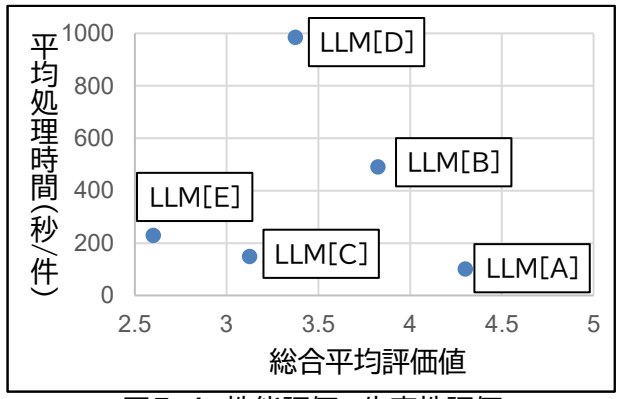



図5-4 性能評価×生産性評価

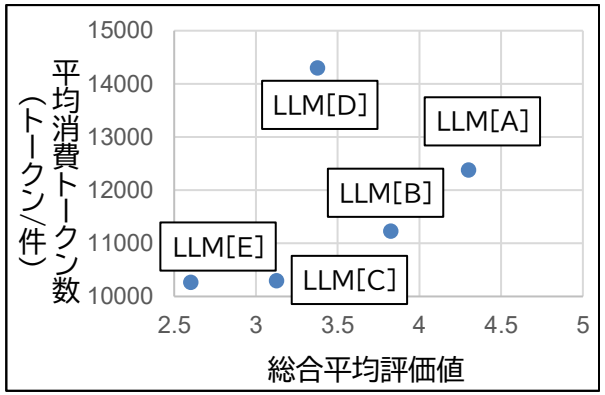



図5-5 性能評価×コスト評価

表5-7 生成AI活用時の推奨役割分担

| 生成AIが貢献できる領域 | 人間が注力するべき領域 |
|---|---|
| ・対外脅威の影響を整理<br>・既存戦略との差分を抽出<br>・リスク候補の洗い出し<br>・施策案、KPI、ロードマップの初稿を作成<br>・経営説明に向けた論点の整理 | ・企業固有の事情の理解<br>・事業影響判断<br>・リスク許容度の決定<br>・予算、人員、期間を踏まえた実現性の評価<br>・法令・社内規程との整合の確認 |

検証の結果は次のとおり。なお、クラウドLLM[A]のみ推論モードで動作させた。

**■タスク別評価**

タスク別の評価結果を表5-6に示す。クラウド推論モデルであるLLM[A]が最も高い評価となり、全タスクで4点以上を獲得した。これは、軽微な修正を前提に初期案として活用できる水準であった。一方、ローカルLLMでは、LLM[B]が一定の分析品質を示しており、LLM[A]には及ばないものの、機微情報を外部送信できない業務では、クラウドLLMの代替候補となり得る。なお、日系モデルと海外モデルの区分だけで評価結果の差を説明できる明確な傾向は確認できなかった。評価結果には、モデル性能等が複合的に影響した可能性がある。

タスク別に見ると、脅威分析やギャップ分析は比較的高い評価となり、入力情報をもとに影響や不足点を整理するタスクでは、一定の有効性が確認された。一方、リスク分析や戦略作成も一定の評価を得たが、事業影響、優先順位など企業固有の事情を踏まえた判断が必要となるため人による妥当性確認の重要性は高い。なお、推論モデルであるLLM[A]ではこのギャップが少ないように見受けられた。

また、点数には表れていないが、LLM[D]では他モデルで示されなかった項目も含まれ、比較的高い網羅性が見られ、複数のLLMを並行使用することで網羅性を担保できる可能性が示唆された。しかし、いずれのモデルでも、平均評価が4.5以上(5点でほぼそのまま利用可能のため、それに迫る点数)となったタスクはなく、人間による確認や修正を前提とすべきであると示された。

**■処理時間・トークン消費**

図5-4では、平均処理時間と総合平均評価値を、図5-5では、平均消費トークン数と総合平均評価値をマッピングした。今回の検証条件では、LLM[A]が最も高い総合平均評価を得たうえ、処理時間も最短であり、品質と生産性の両面で優位性を示した。

ここまでの評価結果から、生成AIは戦略の完成版を自動作成するものではなく、論点整理や初期案作成を支援するツールとして有効であると考えられる。生成AIが戦略見直しで貢献できる領域と人間が注力するべき領域を表5-7に示す。生成AIは、情報整理等の領域で効果を発揮しやすい。一方、事業影響、リスク許容度等は企業固有の判断を伴うため、人間が注力するべき領域である。

検証結果のまとめとして、次の2点を記す。

- 一部のモデルでは初期案作成に活用できる水準であったが、いずれのモデルも人間による根拠確認と専門家レビューが前提となる
- 機微情報、処理時間、トークン消費等を踏まえ、業務の特性やコスト等に応じてクラウドLLMとローカルLLMを使い分けるべきである

セキュリティ戦略修正

# 5.4 生成AI活用業務プロセス移行と課題

生成AIを利用することで、戦略見直しにおいて一定の効果を確認できた。そこで、生成AIを活用した戦略見直しにおいて、どの程度の工数削減が見込まれるかを表5-8の前提条件に基づいて試算した。試算結果は、現状の工数合計を200人時と設定した場合、生成AI利用時では116人時となり、生成AI利用時の追加プロセスを含めても、全体で約42%の工数削減が見込まれる。ただし、この数値は試算であり、実際の削減効果は企業の業務プロセスや、利用するモデルの性能によって変動することに留意が必要である。

表5-8 生成AI活用による工数削減の試算

| 業務プロセス | 現状の工数（人時） | 生成AI利用時の工数（人時） | 工数増減率 | 評価 | 前提条件 |
|---|---|---|---|---|---|
| 情報整理 | 40 | 24 | ▲40% | 既にある情報の整理等は生成AI利用が効率的<br>生成AIに入力する情報の選定や、出力結果の判断等では工数が増加 | 【現状総工数】<br>200人時と設定<br>【生成AI利用時の追加プロセス】<br>・入力情報の選定<br>・出力根拠の確認<br>・機微情報の管理<br>・レビュー記録 |
| 分析・抽出 | 60 | 30 | ▲50% | 評価観点を定めた上での分析で時間削減、網羅性で効果を発揮 | |
| 戦略案の作成 | 60 | 30 | ▲50% | 生成AIで初期案を作成し、人が補正することで工数を削減 | |
| 意思決定層による承認 | 40 | 32 | ▲20% | 承認に向けた資料作成で貢献 | |
| 合計 | 200 | 116 | ▲42% | 全体では約42%の削減 | |

生成AIの利用は工数削減効果が見込まれる一方で、生成AI利用プロセスにおける残存リスクがある。また、生成AIプロセスへの移行においては、新たな統制プロセスの整備等、移行における課題に対応する必要がある。

**■ 生成AI利用プロセスにおける残存リスク**

- 生成AIは一般的なベストプラクティスに基づく妥当な戦略案を提示できるが、自社の経営方針、予算、人員、システム構成、リスク許容度を十分に反映できない可能性がある。
- モデルの性能や入出力制約により、入力した情報が網羅されずに戦略が立案される可能性がある。
- 入力情報に含まれていない情報（例：経営層や事業部門が重視しているが明文化されていない優先事項等）は、生成AIによる戦略案に十分反映されない可能性がある。
- 生成AIは対策の重要度を整理できるが、実際の予算、人員、期限、既存プロジェクト、運用負荷を踏まえた実行可能性の判断には限界がある。

**■ 生成AI利用プロセスへの移行における課題**

- 自社の経営方針、予算、人員、システム構成などの機密情報をLLMに入力する可能性があるため、情報漏えいリスクに留意し、入力情報の管理、アクセス制御、ログ管理などの対策を講じる必要がある。
- 生成AIが作成した案について、短い期間に多くの生成物やバージョンが乱立し、管理が追いつかなくなる可能性がある。
- 生成AIによる作成物は、生成AIを用いて作成したこと、承認状況、責任者・承認者を明示するラベル管理を行い、その管理手順を正式な業務プロセスとして定める必要がある。
- 生成AIの動向は日々変化するため、期間をおいて発生する戦略立案業務では、実施時に都度入力情報を整理し、新モデル検討・モデル比較を実施することで、出力品質を確保（その時点の生成AIの能力を最大限享受）することができる。

セキュリティ戦略修正

## 5.5 業務プロセスの比較とまとめ

本検証では、生成AIが対外脅威の変化を踏まえた戦略見直しにおいて、情報整理・初期分析・施策案作成を支援できることが確認できた。一方で、企業固有の事情、実行可能性、説明責任に関する判断は生成AI単独では完結しないことも確認できた。

そのため、生成AIの出力は人による確認が不可欠であり、統制プロセスと組み合わせて活用する必要がある。本検証における結論を表5-9に示す。

表5-9 結論

| # | 結論 |
|---|---|
| 1 | 生成AIは、戦略見直しにおける入力情報を適切な粒度で入力することで、情報整理・初期分析・施策案作成を効率化できる |
| 2 | 生成AIの出力は検討材料として活用し、最終的な判断・確認・説明責任は企業側が担う必要がある |
| 3 | 生成AI活用は、工数削減だけでなく、入力情報の管理、出力根拠の確認、専門家レビュー等の統制整備とあわせて評価する必要がある |

これまでの検証結果を踏まえ、既存プロセスと生成AI利用プロセスの比較を表5-10に示す。現状プロセスと比べて、初稿作成や比較検討において負荷軽減や工数削減が可能である一方で、説明責任を果たすため、出力根拠の確認や生成AIが作成したものに対するレビュー記録が必要となる。

また、生成AI利用プロセスにおける残存リスクの考慮も必要となるため、生成AIによる効率化だけでなく、生成AI利用に関する統制を業務プロセスに組み込む必要がある。

表5-10 現状プロセスと生成AI利用プロセスの評価比較

| 評価項目 | 現状プロセス | 生成AI利用プロセス |
|---|---|---|
| 業務品質 | 熟練者が関与すれば高いが、担当者差が大きい | 高性能な生成AIと専門家レビューにより、一定の品質安定化が期待できる |
| 網羅性 | 担当者の知見に依存し、論点が偏る可能性がある | 対外脅威、ギャップ、施策候補の抜け漏れの補完に有効 |
| スピード | 情報収集、分析、資料化に時間がかかる | 初稿作成や比較検討を短時間で実施できる |
| 実行可能性 | 社内事情を反映しやすい | 生成AI単独では不十分であり、人間による補正が必要 |
| 説明責任 | 担当者が検討の根拠を説明しやすい | 出力根拠の確認とレビュー記録が必要 |

### 推奨運用方針

本レポートにて推奨する運用方針は次のとおりである。

- 生成AIを、対外脅威の変化を短期間で戦略に反映するための**支援ツールとして活用**する。
- 生成AI活用による作業時間の削減効果を生かし、統制プロセスを組み込んだうえで、従来よりも短い間隔(四半期・半年・年1回など)で戦略を見直す。
  - 巧妙化・高速化するサイバー脅威のトレンドに迅速に対応し、**セキュリティ対策の実効性向上**を図る。

# 6 まとめ・おわりに

## 6.1 まとめ

本レポートでは3種類のセキュリティ業務に対して実施した技術検証や、ビジネスへの活用可能性を検討した結果を示してきた。まとめは下記のとおりである。

**【メール解析】**

- F1値99%以上の精度を出すことも可能で、生成AI利用プロセスにおいて**不審メールを発見する精度向上や従業員の業務負荷軽減を期待できる結果**となり、メール対策ソフトと比較してコスト面での優位性が見られる場合もあった
- 一方、見逃し・過検知のバランス設定や、精度維持のためのメンテナンスが必要となる
- **低難度のメール識別を生成AIで自動化**し、企業情報と照合する必要のあるケースの判断に担当者を注力させることが望ましい

**【Threat Intelligence照合】**

- 本検証の資産数の条件下では、F1値90%以上の精度を出すことも可能であり、現状プロセスと比較すると**幅広い情報を短時間で自社の資産情報と照合することが可能**となる
- 自社への導入検討時には、資産数に応じた処理時間増加や、モデル性能・情報源との相性に留意されたい
- 脅威情報は、事前に定義した対象サイトから**深夜に新規記事を収集・解析し、朝担当者が確認する**運用が推奨される

**【戦略修正】**

- 対外脅威の変化を踏まえた論点整理、既存戦略との差分抽出、初期案の作成に**時間削減効果**を発揮することを確認した
- 業務の性質から**固有リスクは高くなる**ため、動作環境や生成物の確認は慎重な検討が必須である
- 生成AIの作業時間削減効果を生かし、**環境の変化に応じて1年以内の短い周期で戦略を見直す**ことが可能となる

**検証全体から得られたこと**

本書では3種類のセキュリティ業務において、**業務やタスクによって最適なモデルや構成が異なることを実証**した。また、業務の特性に応じたモデル選定やシステム、業務プロセス設計が必要であり、**セキュリティ担当者の業務を十分にフォローできる有用性を見出した。**一方、与えるタスクや実行条件によって最適なモデルや構成が変化するため、厳密な最高性能の追求にはコスト、時間を要すると考えられる。性能を追求している間に新たなモデルが利用可能となる可能性もあるため、あらかじめ性能、コスト、処理時間、動作環境などの必要要件を定義しておくことが推奨される。

本検証ではモデル評価として、各モデルの性能、消費トークン数および処理時間を比較した。その結果、単純な性能比較では**クラウド型の高性能モデルが全体的に優れていた**が、**タスクによってはローカルLLMも近い性能を示した。**特に、より少ないトークン数で高性能な結果を示したローカルLLMは、同等クラスのモデルをクラウドAPIで利用する場合に従量課金を抑えられる可能性があり、性能を維持しつつ、コスト面で優位となり得る。また、各LLMで推論や出力に必要なトークン数は異なるため、モデル選定に当たっては性能だけでなく、消費トークン数や処理時間を含めて評価することが重要である。さらに、本検証では業務プロセス評価として、生成AIを適用した場合と既存の業務プロセスにおける作業時間を比較し、**対象とした業務の一部で工数削減が見込めることを確認**した。なお、ローカルLLMでは、トークン数が直接的な従量課金に結びつかないため、処理時間や計算資源が許容される範囲で、段階的な推論や回答の再評価に計算量を配分し、さらなる性能向上を図るアプローチとしても期待される。

本検証から、すべてのタスクを最新かつ最高性能のモデルで構築する必要はなく、コストや処理時間を含め、**ビジネス上の要件とセキュリティ上の要件を両立できるモデル**を選択することが重要であることが示された。また、本書では5つのモデルを用いて検証したが、公開年が直近のモデルで良い結果が得られるケースが大半であった。生成AIの技術進化やモデルの更新は今後も継続するため、導入時には**モデルを容易に変更できる構成**とし、将来の入れ替えに伴うシステム全体のコストも考慮しておく必要がある。

## 6.2 おわりに

本検証を円滑に遂行し、実効性のある成果を得るためには、**「AI」と「セキュリティ」の双方の知見を有する人材の確保、あるいは両専門家の緊密な協働（コラボレーション）**が不可欠であることを認識した。具体的な理由は以下の2点である。

- 生成AIの知識が不足している場合、モデルの特性を踏まえたプロンプト設計や、AIの能力を最大限に発揮する業務プロセスを検討することが難しい
- セキュリティの知識が不足している場合、検証対象や評価基準を適切に設計できず、性能以外の観点から業務への活用可能性やリスクを評価することが難しい

生成AIをセキュリティ業務へ適用するためには、モデルの評価を行うだけでなく、対象業務の特性やリスクを理解した上で、業務プロセスや評価方法を設計する必要がある。

本検証の「検証設計」から「モデル評価」のフェーズにおいて**生成AI・セキュリティの双方に精通した人材を投入し、要した工数は以下のとおりである。**これらの工数には、対象業務の整理、検証方針の設計、データの準備、プロンプトの作成、モデルの実行、結果の評価および考察が含まれる。なお、記載した工数は本検証の条件に基づく概算であり、対象範囲、データ量、利用環境および担当者の習熟度によって変動する。
※ローカルLLM環境の構築は除く

- メール解析　：0.6人月
- TI　　：2人月
- 戦略修正　：1人月

今後、AIエージェントの利用が一般化するにつれて、生成AIを既存の業務プロセスに組み込む動きが加速すると考えられ、課題感のある業務では**継続的に検証を実施**することで先進的な活用が可能となる。

各企業では、AI活用検証を一過性の取り組みで終わらせず、**生成AIとセキュリティの双方を理解する人材の育成・確保および両分野の担当者が連携できる体制の構築**を進めることが望ましい。まずは、定型的な作業や人手による確認を前提とした支援業務から段階的に適用し、効果とリスクを確認しながら活用範囲を拡大していくアプローチが重要である。

# 謝辞

本レポートの作成にあたりまして、ご協力いただきました企業の皆様には多大なるご支援・ご尽力を賜りました。この場を借りて心より御礼申し上げます。

また、産業サイバーセキュリティセンター 中核人材育成プログラムの講師である、門林 雄基先生、小林 和真先生には、本書の元となるプロジェクトのメンターとして、ご指導・ご助言を賜りました。改めて御礼申し上げます。

そして、本レポートの作成や本プロジェクトをともに実施した、下記メンバーの皆様にも感謝を伝えたいと思います。

**セキュリティ担当者のための生成AIセキュリティプロジェクト**
**技術検証チーム メンバー**

【プロジェクトリーダー】
星 凜太郎　　山田 敏大

【チームリーダー】
榛葉 光　　和田 浩樹

【メンバー】
亀井 祐樹　　坂脇 孝弘　　佐藤 舞　　島袋 洋一
下出 有実　　中島 しおり　　向井 耕司

# Appendix

# Appendix.1：技術検証・モデル評価概要

本検証では、クラウドLLM・ローカルLLMの回答結果と、予め用意した正解データや評価基準を比較してLLMの性能等を評価した（図6-1）。

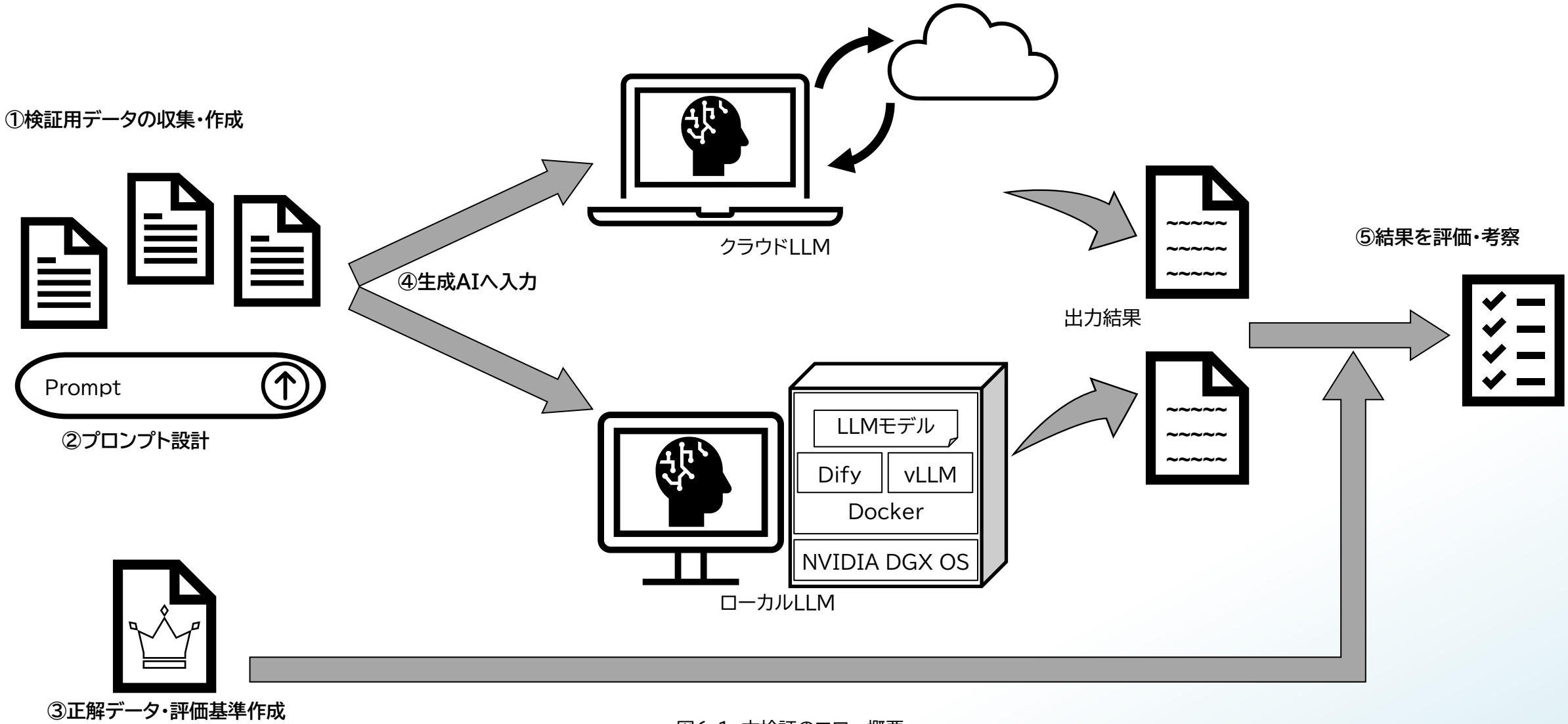



図6-1 本検証のフロー概要

# Appendix.2：ローカルLLM動作環境

本検証で評価した5モデルのうち、クラウドLLM[A]を除く4モデル(ローカルLLM[B]～[E])は、機密性の高い情報を扱う運用環境を想定し、ローカル環境で動作させた。本Appendixでは、その動作環境の主な仕様を表6-1に示す。

【主な仕様】

表6-1　本検証で使用したローカルLLMの動作環境

| 項目 | 仕様 |
|---|---|
| 機器名 | Lenovo ThinkStation PGX |
| OS | NVIDIA DGX OS（Ubuntu Linux Proベース） |
| プロセッサー | 20コア ARM（Cortex-X925×10 + Cortex-A725×10） |
| メモリ | 128GB LPDDR5x（統合型、256bit、273GB/s帯域） |
| ストレージ | 1TB / 4TB NVMe SSD（自己暗号化機能対応） |
| グラフィックス | NVIDIA Blackwell Architecture（第5世代Tensor / 第4世代RTコア搭載） |
| AI処理能力 | 最大1000TOPS、1 PetaFLOPS（FP4） |
| ディスプレイ出力 | HDMI 2.1a（最大8K出力、マルチチャネル音声対応） |
| ネットワーク | 10GbE対応RJ-45、2台結合用NVIDIA ConnectX-7 Smart NIC（最大200Gbps） |
| 本体寸法 | 約150（W）×150（D）×50.5（H）mm |

【動作環境】

モデルをコマンドレベルまたはチャットインターフェースにて操作した。

# Appendix.3：モデルの評価指標

分類モデルの性能は、予測結果と実際の正解の対応関係から4つに分類されるTP/TN/FP/FNを基本要素として算出する。本検証では、3章メール解析の「不審メール」と判定すること、4章TI照合の「対応必要」と判定することをPositive(陽性)と定義する。

- TP(True Positive) ：モデルが正解と予想して実際に正解だった数
- TN(True Negative) : モデルが間違いと予想して実際に間違いだった数
- FP(False Positive) :モデルが正解と予想したが実際は間違いだった数 (過検知に相当)
- FN(False Negative): モデルが間違いと予想したが実際は正解だった数 (見逃しに相当)

【F1値】

F1値は、分類モデルの性能を総合的に評価する指標 であり、PrecisionとRecallの調和平均で計算される(0～1の範囲、数値が高いほど高性能)。両者のバランスを反映するため、Precisionのみ・Recallのみで評価する場合よりも、過検知と見逃しの両面を考慮した評価が可能となる。

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

また、内訳であるPrecisionとRecallを見ることで過検知・見逃しのどちらが多いかを確認可能である。

- Recall(再現率)：本当に正解であるもののうち、モデルが正しく正解と予想できた割合(0～1)を表す。**数値が高いほど見逃し率が低い**

$$Recall = \frac{TP}{TP + FN}$$

- Precision(適合率)：モデルが正解と予想したもののうち、本当に正解であった割合(0～1)を表す。**数値が高いほど過検知率が低い**

$$Precision = \frac{TP}{TP + FP}$$

【正解率(Accuracy) 】

正解率(Accuracy)は、全評価対象のうち、モデルが正しく予想できた割合(0~1、高いほど高性能)である。本検証では、Negativeクラスを含まないデータセット(例:3章解析②)で過検知の評価ができない場合に用いる。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

# 参考文献

# 参考文献

[1] Proofpoint , “Proofpoint 2024 State of the Phish Report”,2025, https://www.proofpoint.com/sites/default/files/threat-reports/pfpt-us-tr-state-of-the-phish-2024.pdf .
[2] KnowBe4 , “2025 Phishing By Industry Benchmark Report”,2026, https://www.knowbe4.com/resources/reports/phishing-by-industry-benchmarking-report .
[3] AAG IT Support , “The Latest 2025 Phishing Statistics”, 2026, https://aag-it.com/the-latest-phishing-statistics .
[4] Innovation & Co. , “ITトレンド「【2026年】メールセキュリティソフト13選比較」”, 2026, https://it-trend.jp/mail_security/article/56-0014 .
[5] Palo Alto Networks , “Forrester Study: The 2020 State of Security OperationsI”, 2020, https://www.paloaltonetworks.com/blog/2020/09/state-of-security-operations/ .
[6] IPA , “2024年度 中小企業における情報セキュリティ対策に関する実態調査”, 2025, https://www.ipa.go.jp/security/reports/sme/nl10bi000000fbvc-att/sme-chousa-report2024r1.pdf .
[7] Microsoft , “Unify now or pay later: New research exposes the operational cost of a fragmented SOC”, 2026, https://www.microsoft.com/en-us/security/blog/2026/02/17/unify-now-or-pay-later-new-research-exposes-the-operational-cost-of-a-fragmented-soc .
[8] CyberSierra , “What Is Alert Fatigue and How to Combat It in Your SOC”, 2025 https://cybersierra.co/blog/alert-fatigue-in-soc/ .
[9] UnderDefense , ”Alert Fatigue in Cybersecurity: the SOC Playbook to Eliminate It”, 2025, https://underdefense.com/blog/alert-fatigue-cybersecurity/ .
[10] デジタルアーツ株式会社, “Digital Arts Security Reports”, 2026, https://www.daj.jp/security_reports/45/ .