# 知財部における野良 AI エージェント防止レポート

ChatGPT-5.5

## エグゼクティブサマリー

結論は明確です。知財部で「野良 AI エージェント」は十分に起こり得ます。理由は、Claude Code や Codex が単なる対話 UI ではなく、**ファイル読取・編集、シェル実行、外部接続、hooks、skills、MCP、サブエージェント、リポジトリ同梱の指示ファイル読込** を備えた「実行主体」だからです。RPA で表面化した **台帳なし、責任者不明、共有資格情報、変更管理なし、静かな故障、利用部門単独開発** は、そのまま AI エージェントにも再出現します。[1]

知財部では被害が重くなりやすい点が本質です。個人情報当局は、個人情報を含むプロンプト入力が利用目的達成に必要な範囲内か、また提供事業者が機械学習に利用しないか等を十分確認するよう注意喚起しています。加えて、営業秘密管理の観点では管理体制そのものが重要であり、著作権行政も生成 AI と著作権の整理は現時点の一定の考え方であって法的拘束力を持たないと明示しています。さらに、特許実務は進歩性・記載要件・発明該当性など高度非定型判断を含みます。つまり知財部では、**誤答** よりも **秘密性喪失、権利取得機会の逸失、第三者権利侵害、説明責任不能** が重大リスクです。[2]

実務上の正解は「全面禁止」ではなく、**統制付き市民開発** です。知財部のエージェントは、**ソフトウェア資産 + 設定資産 + 接続資産 + 資格情報 + 実行境界 + ログ資産** として扱い、台帳、承認、固有 ID、Secrets 管理、Sandbox、Network Egress 制御、Prompt/MCP/Hook の版管理、定期再承認を必須化し、**法的結論の確定と外部送信は必ず人が止める** という原則で設計すべきです。国内ガイドラインも、トレーサビリティ、責任者明示、文書化、教育・リスキリングを中核に置いています。[3]

| 最優先で固定すべきこと | 推奨 |
|---|---|
| 認証 | 機密案件は企業管理下アカウントのみ。個人契約・個人課金は禁止 |
| 権限 | `full access` / `danger-full-access` / `approval never` / `bypass` 相当を機密案件で禁止 |
| 接続 | MCP、Hook、Web、外部 DB はデフォルト拒否、案件ごと許可 |
| 変更 | `AGENTS.md`、`CLAUDE.md`、`.claude/`、`.codex/`、`hooks.json`、`.mcp.json` は PR 承認制 |
| 人手介在 | 法的判断、提出物最終版、対外送信、権限昇格は必ず人手承認 |
| 監査 | 台帳、承認ログ、実行ログ、設定変更ログ、接続先ログを相関保存 |

## 前提と評価軸

本報告は、知財部が **公開情報、社内文書、未公開発明、契約、交渉メモ、発明者情報、外部 DB** を扱い得る前提で評価しています。また、ローカル端末、社内リポジトリ、CI、クラウドワークスペースを経由した利用を含みます。未指定の事項は、**一般的な日本企業の知財部で起こりやすい運用** を仮定しています。

本報告でいう **野良 AI エージェント** とは、組織の承認、台帳、責任所在、ログ、権限設計の外で、個人または現場判断で作成・配布・実行されるエージェントを指します。対象は、チャット履歴だけでなく、`AGENTS.md`、`CLAUDE.md`、skills、hooks、rules、MCP 設定、ローカル実行スクリプト、CI 組込ジョブまで含みます。公式仕様上、これらのツールはリポジトリやユーザー領域の設定・指示ファイル、hooks、MCP、subagents で挙動が変わるため、統制対象は「モデル」ではなく「実行面全体」と考える必要があります。[4]

評価軸は、**影響度** と **発生確率** の二軸です。知財部では、金銭被害よりも、秘密性喪失、出願・交渉ポジション喪失、第三者権利侵害、説明責任不能の重みを高く置いています。[5]

## 野良 RPA から得る教訓

RPA から持ち込むべき教訓はかなり具体的です。金融向け実務資料では、RPA を **基本構想、企画・計画、開発、運用・保守** の各フェーズで管理し、利用開始前にロボット管理台帳、業務マニュアル、テスト記録を整備し、IT ガバナンス担当が確認・承認し、本番後は運用責任者へ責任移管し、棚卸と変更反映を行うことが示されています。政府の市民開発実務でも、中央コードリポジトリ、RBAC、開発・テスト・本番の分離が推奨されています。[6]

| 失敗モード | RPA で起きたこと | AI エージェントでの再発形 | 効く統制 |
|---|---|---|---|
| 所在不明 | 夜間操作が不正アクセス扱いになったが、実際は未把握の RPA だった | 未登録 CLI エージェント、個人 MCP、野良 Hook が責任不明のまま稼働 | 利用前登録、固有 Agent ID、Owner 明示 |
| 共有資格情報 | 実行主体を後から区別できない | 人のアカウントや PAT をそのままエージェントが利用 | 人と Agent の ID 分離、短命資格情報 |
| 平文秘密情報 | 設定にパスワード・ID を保存 | `.env`、Prompt、Hook、MCP 設定にトークン埋込 | Secrets Manager、平文禁止、検知 |
| 静かな故障 | UI 変更や依存先変更で処理停止・誤処理 | 外部 DB 仕様変更や HTML 変化で誤検索・誤要約が継続 | 依存先一覧、回帰テスト、異常時停止 |
| ブラックボックス化 | 担当者異動後にロボットの意図が不明 | Prompt、Hook、MCP の意図が再現不能 | 台帳、設計書、版管理、引継ぎ |

| 失敗モード | RPA で起きたこと | AI エージェントでの再発形 | 効く統制 |
|---|---|---|---|
| 利用部門単独開発 | 重要業務を現場単独で作り品質低下 | 高機密案件向けエージェントを個人設定で本番利用 | 重要案件は共同承認制 |

RPA で効いたのは、「禁止」よりも **現場活用を認めつつ、中央の標準・レビュー・環境分離・台帳・監査を強制する** 方式でした。AI エージェントでも同じです。禁止だけを強めると、個人設定・個人端末・個人契約へ逃げやすくなります。知財部では、**認可された市民開発** を正面から制度化した方が、結果として野良化を抑えやすいと考えるべきです。[7]

## 技術リスクの整理

コード実行型エージェントでは、統制対象が「処理ロジック」から「処理ロジック + 指示ファイル + 接続設定 + 承認モード + Hooks/Skills/MCP + 監査設定」へ広がります。公式仕様上、Claude Code は managed / user / project / local の設定階層、`CLAUDE.md`、skills、hooks、MCP、subagents、監視機能を持ち、Codex は `config.toml`、`AGENTS.md`、hooks、rules、MCP、sandbox、approvals、managed configuration、Compliance API を持ちます。したがって、**Prompt 資産、設定資産、接続資産、認証資産、実行境界** をまとめて変更管理しなければなりません。[8]

| 面 | 何が危険か | 知財部の固定値 |
|---|---|---|
| 制御面ファイル | `AGENTS.md`、`CLAUDE.md`、skills、hooks、rules、MCP 設定の変更で同じ名前でも別物になる | 制御面ファイルを「コード」と同格に扱い、PR 承認・版番号・差分レビュー必須 |
| 権限モード | 自動承認や full access は誤実行・越権・外部送信の爆発半径を広げる | 機密案件では sandbox + on-request 相当のみ |
| 資格 | ローカル保存資格情報やキャッシ | 企業認証固定、短命資格情報、OS |

| 面 | 何が危険か | 知財部の固定値 |
|---|---|---|
| 情報 | ュが漏えい起点になる | 保護と集中失効 |
| 接続面 | MCP、hooks、Web、第三者サービスが流出・契約逸脱経路になる | Connector/MCP は申請制、案件別 allowlist |
| 自己変更 | エージェントが自分の設定や指示を書き換えると監査不能になる | 制御面ファイルへの write を deny、変更は別 PR |
| 非対話運用 | 人のレビューが入らず、CI で一気に逸脱する | 非対話実行は Green 限定。Amber/Red は専用 review 必須 |
| ログ | 実行・承認・設定変更の相関が取れないと説明不能 | Session, Agent, User, Repo, Config version, Domain, Reviewer decision を必須取得 |

非対話運用は一段高いリスクとして扱うべきです。Claude Code では non-interactive の -p 実行時に trust verification が無効化される点があり、Codex の auto-review は reviewer を置き換えるだけで sandbox 境界を広げるものではありません。つまり、自動化のときほど「レビューを減らす」のではなく、「境界を狭くする」必要があります。[9]

脅威分類として特に重要なのは、**Prompt Injection**、**Insecure Output Handling**、**Sensitive Information Disclosure**、**Excessive Agency**、**Overreliance**、**Confabulation** です。知財部の運用に引き直すと、優先順位は次のようになります。[10]

| 影響度＼発生確率 | 低 | 中 | 高 |
|---|---|---|---|
| 高 | R4 自己変更・設定ドリフト | R3 過剰権限 MCP / Hook / 秘密参照 | **R1 機密入力の外部送信 / R2 幻覚を含む法的・技術判断の採用 / R6 未登録エージェントの野良運用** |

| 影響度＼発生確率 | 低 | 中 | 高 |
|---|---|---|---|
| 中 | R7 コスト暴騰・無駄な並列実行 | R5 外部 DB や依存先変更による静かな故障 | R8 承認疲れによる漫然承認 |
| 低 | R9 一時ファイル残存 | R10 長期保持の過剰 | — |

## 知財部で高まる組織リスク

知財部の第一の論点は、**一度の入力・一度の送信が不可逆になりやすい** ことです。未公開発明、先行技術検討メモ、共同研究情報、交渉メモ、発明者情報は、個人情報・営業秘密・限定提供データの論点をまたぎやすく、一般部門よりも「入力前の情報分類」が重要です。個人情報当局の注意喚起をそのまま知財部に適用すると、**機密案件では「入れないこと」が第一統制** であり、「入れるなら学習利用なしの確認、目的内利用、最小化、ログ」が必要になります。[11]

第二の論点は、**高度非定型判断は自動化対象ではなく、補助対象** だということです。特許行政は AI 関連技術でも進歩性、記載要件、発明該当性の判断ポイントを事例化しており、著作権行政も生成 AI との関係整理は今後の事例蓄積で見直し得るとしています。したがって、知財部では **検索補助、要約、比較表、論点洗い出し** までは許容しやすい一方、**出願可否、FTO 結論、侵害・有効性意見、著作権結論、提出物最終版、交渉方針の確定** は人間専権に残すべきです。[12]

第三の論点は、**外部 DB 利用は第三者アクセス管理そのもの** だということです。国際的な生成 AI リスク管理では、第三者アクセス主体の棚卸、IP・プライバシー・セキュリティを含むベンダーデューデリジェンス、承認済み提供者リスト整備、第三者変更履歴の保持が推奨されています。MCP や外部接続を使う以上、「どの DB を使うか」は単なる利便性の問題ではなく、**誰が、どの案件で、どの匿名**

**化水準で、どの契約条件の下で接続するか** の統制だと位置付ける必要があります。[13]

| 区分 | 知財部での典型例 | 許容条件 | 推奨判断 |
|---|---|---|---|
| Green | 公開公報の要約、公開情報だけの比較表、ダミーデータでの整形 | 企業認証、台帳、sandbox、外部送信なし | 推奨 |
| Amber | 未公開ドラフトの構造化、請求項比較表、社内文書要約、承認済み外部 DB 照会補助 | Owner 承認、固有 ID、allowlist、ログ、二段階レビュー | 条件付き許可 |
| Red | 出願可否の最終判断、FTO 結論、侵害・有効性意見、対外メール送信、提出物最終版生成、未承認 MCP 追加 | 自律実行は禁止。補助利用でも上席レビュー必須 | 原則禁止または人手限定 |

## 実務ガバナンス設計

国内ガイドラインは、トレーサビリティ向上、責任者明示、関係者間の責任分配、文書化、教育・リスキリングを明示しています。したがって、知財部のガバナンスは単なる利用規程では足りず、**責任・記録・変更・棚卸** を持つ制度設計が必要です。[3]

| 統制テーマ | 最低限の実務ルール |
|---|---|
| 利用主体 | 機密案件は企業契約・企業認証のみ。個人アカウント・個人端末は禁止 |
| Agent ID | 人 ID と分離した固有 Agent ID を付与 |
| 権限 | read/write/network/secret を個別申請。最小権限 |
| 実行境界 | Sandbox 必須。機密案件で unsandboxed / full access / no-approval 禁止 |
| 外部接続 | MCP、Hook、Web、外部 DB はデフォルト拒否。案件ごと allowlist |

| 統制テーマ | 最低限の実務ルール |
|---|---|
| 秘密情報 | Prompt、Hook、設定、Repo に平文秘密情報を置かない |
| 変更管理 | AGENTS.md、CLAUDE.md、.claude/、.codex/、hooks.json、.mcp.json、rules は PR 承認制 |
| 人手介在 | 法的結論、外部送信、権限昇格、新規接続追加は必ず人手承認 |
| ログ | 実行、承認、設定変更、接続先、成果物識別子を保存 |
| 再承認 | Amber は 90 日、Green でも 180 日以内。無更新は停止 |
| 廃止 | 資格情報失効、接続削除、ログ保全、台帳クローズを一括実施 |

**Agent 台帳テンプレート**

| Agent ID | 名称・目的 | Owner | 承認者 | 利用区分 | 対象データ区分 | 実行面 | 利用 Repo/Workspace | Prompt/設定資産版 | MCP/Hook/Skill | 認証方式・Agent ID | Secrets 参照元 | Filesystem 権限 | Egress allowlist | 人手テスト | ログ保存先 | 再承認日 | 状態 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AGT-{連番} | {何をするか一} | {知財責任者} | {知財+情シス/セ | Green/Amber/Red | Public/Internal/Confidential/Secret/PII | Local/CI/Cloud | {repo・案件名} | {git tag/hash} | {名称・版・許可先} | {SSO/API/Workspace} | {Vault path 等} | {read-only/workspace-write} | {domain list} | {レビュー者と条 | {SIEM/bucket} | {YYYY-MM-DD | Draft/Approved/Running/Retired |

| Agent ID | 名称・目的 | Owner | 利用区分 | 承認者 | データ区分 | 実行面 | 利用 Repo/Workspace | Prompt/設定資産版 | MCP/Hook/Skill | 認証方式・Agent ID | Secrets 参照元 | Filesystem 権限 | Egress allowlist | 人手ゲート件 | ログ保存先 | 再承認日 | 状態 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 文て | | キュリティ | | | | | | | | } | | | } | | } | |

**認可プロセス**

| 区分 | 申請に必要な証跡 | 承認者 | 許容できる自律度 | 承認期間 |
|---|---|---|---|---|
| Green | 目的、Owner、Repo、権限、ログ、sandbox 設定 | チーム責任者 | 低リスク定型処理のみ | 180 日 |
| Amber | 上記 + データ分類、外部接続先、Secrets 設計、評価結果、人手ゲート | 知財 Owner + 情シス/セキュリティ | 補助作業まで。外部送信・法的結論は不可 | 90 日 |
| Red | 上記 + 法務/個情レビュー、案件責任者レビュー | 自律実行は承認しない | 補助利用のみ。最終成果物は必ず人が起案・決裁 | 個別案件 |

## 運用定着と監査

運用面では、RPA のライフサイクル統制に、AI 特有のトレーサビリティと教育を重ねる構成が最も実務的です。RPA 実務では利用開始前のドキュメント整備・承認、責任移管、棚卸、変更反映が明示されており、セキュア開発フレームワークも安全な実践をライフサイクルに組み込むことを基本としています。[14]

```
flowchart LR
    A[設計<br/>目的・データ分類・権限設計] --> B[承認<br/>知財 Owner・情シス・セキュリティ]
    B --> C[実装<br/>Git 管理・Secrets 参照・Sandbox・Allowlist]
    C --> D[検証<br/>機能/安全/引用/回帰]
    D --> E[本番運用<br/>限定ユーザー・監視・ログ保全]
    E --> F[監査<br/>台帳照合・権限再認証・KPI 確認]
    F --> G[変更<br/>PR 承認・再テスト・再承認]
    G --> E
    E --> H[廃止<br/>資格情報失効・接続削除・ログ保全]
```

**RACI**

| 活動 | 知財 Owner | 利用者 | 情シス /Platform | セキュリティ | 法務・個情 | 内部監査 |
|---|---|---|---|---|---|---|
| 利用企画・案件分類 | A | R | C | C | C | I |
| 技術設計・実装 | C | R | A | C | I | I |
| 承認 | A | C | C | C | C | I |
| 本番運用 | A | R | C | C | I | I |
| ログ監視 | I | I | R | A | I | C |
| 例外承認 | A | I | C | A | C | I |
| 事故対応 | A | R | C | A | C | I |
| 再承認 | A | C | C | C | C | I |

| 活動 | 知財 Owner | 利用者 | 情シス/Platform | セキュリティ | 法務・個情 | 内部監査 |
|---|---|---|---|---|---|---|
| 廃止 | A | R | C | C | I | I |

国内ガイドラインは教育・リテラシーとリスキリングを明確に要求しており、RPA 市民開発の実務でも標準手順の周知と既存 IT 部門との連携が成功要因とされています。知財部では、広い AI 教育よりも、**何を入力してはいけないか、何を自動化してはいけないか** を先に教える方が効果的です。[15]

| 対象 | 内容 | 頻度 | 完了証跡 |
|---|---|---|---|
| 全利用者 | 情報分類、個人情報・営業秘密、禁止行為、承認ルール、幻覚の扱い | 初回 + 年 1 回 | 受講記録・確認テスト |
| Power User | Sandbox、許可された MCP、Hook、Secrets、Prompt 版管理 | 初回 + 半年ごと | 演習合格 |
| 承認者 | 区分判定、例外承認、ログ確認、外部送信判定 | 初回 + 年 1 回 | ケース演習 |

**承認チェックリスト**

| 項目 | 確認内容 | 判定 |
|---|---|---|
| 目的の明確性 | 「何を補助するか」「何をさせないか」が一文で定義されているか | Yes/No |
| データ分類 | Public/Internal/Confidential/Secret/PII が付与されているか | Yes/No |
| 禁止業務 | 法的結論、外部送信、提出物最終化などが禁止されているか | Yes/No |
| 認証 | 企業認証・企業 Workspace に限定されているか | Yes/No |
| 権限 | read/write/network/secret が最小権限か | Yes/No |
| Sandbox | unsandboxed/full access が禁止されているか | Yes/No |
| Egress | allowlist が設定済みか | Yes/No |

| 項目 | 確認内容 | 判定 |
| --- | --- | --- |
| Secrets | 平文秘密情報が repo / prompt / hook にないか | Yes/No |
| 制御面保護 | 制御面ファイルへの write deny があるか | Yes/No |
| ログ | Session/Agent/User/Domain/Decision が記録されるか | Yes/No |
| テスト | 事前評価・回帰テスト・引用確認が完了しているか | Yes/No |
| 再承認 | 有効期限と棚卸日が設定されているか | Yes/No |

**ログ項目一覧**

| 項目 | 必須 | 備考 |
| --- | --- | --- |
| timestamp | 必須 | UTC/JST の両方に対応 |
| session_id / thread_id | 必須 | 一連操作を追跡 |
| agent_id | 必須 | 台帳 ID と一致 |
| user_id | 必須 | 実行起点の人を識別 |
| workspace / repo / branch / commit hash | 必須 | 版と案件を固定 |
| model / runtime | 必須 | 再現性のため |
| prompt asset version | 必須 | AGENTS/CLAUDE/skill/rule/hook 版を含む |
| tool call name | 必須 | edit, bash, mcp, web, file 等 |
| target path / domain / system | 必 | 接触先を示す |

| 項目 | 必須 | 備考 |
| --- | --- | --- |
| approval request / reviewer / decision | 必須 | 手動・自動・却下理由 |
| output artifact hash | 推奨 | 成果物同一性確認 |
| exception / interrupt / timeout | 必須 | 故障解析用 |

監査では、ベンダー側の監視・監査機能だけに依存せず、社内 SIEM で **実行、承認、設定変更、接続先** を突き合わせることが重要です。特に、機密 Prompt 全文の長期保持は避け、案件区分に応じて要約、ハッシュ、部分マスキングを使い分ける方が現実的です。[16]

## KPI・ロードマップ・推奨ソース

**KPI**

| KPI | 定義 | 目標 |
| --- | --- | --- |
| 台帳捕捉率 | 台帳登録済み Agent ÷ 発見済み Agent | 6 か月で 95%以上 |
| 企業認証率 | Amber/Red 案件で企業認証利用の比率 | 100% |
| 固有 ID 率 | 人と分離した Agent ID を持つ比率 | 100% |
| Sandbox 適用率 | Amber/Red で sandbox 有効の比率 | 100% |
| Egress 最小化率 | Amber/Red で allowlist 運用の比率 | 100% |
| 制御面版管理率 | Prompt/MCP/Hook/Rule が Git 管理下の比率 | 100% |
| 無承認接続追加 | 未承認 MCP/Hook/Domain 追加件数 | 0 件 |
| 再承認遵守率 | 期限内再承認完了率 | 95%以上 |
| 事故検知時間 | 初回兆候から停止判断まで | 4 時間以内 |

## 監査チェック

| 監査観点 | チェック内容 |
|---|---|
| 発見統制 | 端末・Repo 内の `.claude/`、`.codex/`、`AGENTS.md`、`CLAUDE.md`、`hooks.json`、`.mcp.json` と台帳を照合 |
| 認証統制 | 個人契約・個人認証が Amber/Red で使われていないか |
| 権限統制 | write/network/secret が申請どおり最小権限か |
| 変更統制 | 制御面ファイル変更が PR 承認と一致するか |
| 接続統制 | 新規 Domain/MCP/Hook 追加が承認済みか |
| 秘密管理 | Repo / Prompt / Hook / Settings に平文秘密情報がないか |
| ログ統制 | Session, Agent, User, Domain, Decision の相関が取れるか |
| 廃止統制 | 異動・退職・案件終了後の ID と接続が閉じているか |

## 導入ロードマップ

| 期間 | 主要施策 | 努力度 |
|---|---|---|
| 短期 | 暫定ポリシー発行、Amber/Red で個人契約禁止、台帳開始、三区分導入、危険モード禁止、MCP 申請制 | 低〜中 |
| 短期 | 実行ログ最小セット収集、案件 ID 付与、Owner 明示、再承認期限運用開始 | 中 |

| 期間 | 主要施策 | 努力度 |
|---|---|---|
| 中期 | 企業認証強制、Agent 固有 ID、Secrets Manager 連携、allowlist、Sandbox 標準化、SIEM 連携 | 中〜高 |
| 中期 | 回帰テスト、引用検証、四半期監査、教育制度化 | 中 |
| 長期 | Ephemeral runtime 標準化、ポリシーコード化、接続先自動棚卸、退役自動化、監査証跡半自動生成 | 中〜高 |

**推奨ソース**

| 優先 | ソース | 用途 |
|---|---|---|
| A | □url□Claude Code Security□turn8search9□/ □url□Permissions□turn8search1□/ □url□Settings□turn5search0□/ □url□Monitoring□turn5search1□ | 権限、Sandbox、Managed Settings、監視 |
| A | □url□Codex Agent approvals & security□turn10search0□/ □url□Sandbox□turn10search7□/ □url□Authentication□turn12search6□/ □url□Managed configuration□turn12search3□/ □url□Governance□turn10search5□ | 承認、Sandbox、認証、企業統制、監査 |
| A | **経済産業省[17]** — ☒ url ☒ AI 事業者ガイドライン☒ turn24view3 ☒ / ☒ url ☒営業秘密管理指針☒ turn24view5 ☒ | 国内ガバナンス、営業秘密 |
| A | **個人情報保護委員会[18]** — ☒ url ☒生成 AI サービス利用の注意喚起☒ turn27view0 ☒ | 個人情報入力と学習利用確認 |
| A | **文化庁[19]** — ☒ url ☒ AI と著作権☒ turn33search0 ☒ / ☒ url ☒ AI と著作権に関する考え方☒ turn33search1 ☒ | 著作権リスク、法的確実性の限界 |

| 優先 | ソース | 用途 |
|---|---|---|
| A | **NIST**[20] — □url□AI RMF 1.0□turn16search7□/ □url□Generative AI Profile□turn0search10□/ □url□SSDF□turn31search0□ | リスク管理、生成 AI 特有リスク、セキュア開発 |
| B | **OWASP**[21] — □url□□LM Top 10□turn24view0□/ □url□Agentic Applications 2026□turn24view1□/ □url□Agentic Skills Top 10□turn30view0□ | 脅威分類、過剰自律、スキル/プラグインリスク |
| B | **情報処理推進機構**[22] — □url□AI セキュリティ□turn24view2□/ □url□AISI 既知の攻撃と影響☒ turn32search1 ☒ | 日本語の AI セキュリティ実務資料 |
| B | **特許庁**[23] — ☒ url ☒ AI 関連技術に関する特許審査の事例☒ turn24view6 ☒ | 非定型判断を人手に残す根拠 |
| B | **金融情報システムセンター**[24] — ☒ url ☒ RPA 導入にあたっての解説書☒ turn18search4 ☒ | 台帳、承認、運用、棚卸の RPA 教訓 |
| B | □url□Digital.gov Citizen Development in RPA□turn22search18□ | 統制付き市民開発の実務 |

**Open questions / limitations**

本報告は公開情報を前提にしたため、社内で既に使っている ID 基盤、プロキシ、DLP、SIEM、外部 DB 契約条件、代理人との連携形態までは個別最適化していません。実装時には、**既存社内標準に寄せること、案件類型ごとに Green/Amber/Red を細分化すること、外部 DB と代理人メールの扱いを別ルールに切り出すこと** が必要です。

[1] [22] Claude Code overview - Claude Code Docs

https://docs.anthropic.com/en/docs/claude-code/overview?utm_source=chatgpt.com

[2] [11] 生成 AI サービスの利用に関する注意喚起等について -個人情報保護委員会-

https://www.ppc.go.jp/files/pdf/230602_kouhou_houdou.pdf?utm_source=chatgpt.com

[3] [15] https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf?utm_source=chatgpt.com

[4] [8] [18] [19] [23] Claude Code settings - Claude Code Docs

https://docs.anthropic.com/en/docs/claude-code/settings?utm_source=chatgpt.com

[5] [17] https://www.meti.go.jp/policy/economy/chizai/chiteki/guideline/r7ts.pdf

https://www.meti.go.jp/policy/economy/chizai/chiteki/guideline/r7ts.pdf?utm_source=chatgpt.com

[6] [14] https://www.fisc.or.jp/document/fintech/file/rpa20200120_07.pdf

https://www.fisc.or.jp/document/fintech/file/rpa20200120_07.pdf?utm_source=chatgpt.com

[7] [24] 5 Tips for Implementing Citizen Development in Your RPA Program | Digital.gov

https://digital.gov/2021/08/16/5-tips-for-implementing-citizen-development-in-your-rpa-program?utm_source=chatgpt.com

[9] Security - Claude Code Docs

https://code.claude.com/docs/en/security?utm_source=chatgpt.com

[10] OWASP Top 10 for Large Language Model Applications | OWASP Foundation

https://owasp.org/www-project-top-10-for-large-language-model-applications/?utm_source=chatgpt.com

[12] [20] AI 関連技術に関する特許審査の事例について | 経済産業省 特許庁

https://www.jpo.go.jp/system/laws/rule/guideline/patent/ai_jirei.html?utm_source=chatgpt.com

[13] [21] Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile

https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf?utm_source=chatgpt.com

[16] Monitoring - Claude Code Docs

https://docs.anthropic.com/en/docs/claude-code/monitoring-usage?utm_source=chatgpt.com