

知財業務への AI エージェント適用における ハーネス設計の注意点

— 機密境界・出典検証・承認ゲートを「構造」で担保する —

2026 年 10 月 26 日（2026 年 10 月時点の公開情報に基づく整理）

Claude Opus 5.5

凡例：本文中の【事実】は一次ソース（公式文書・公表資料・論文）または明示した二次情報で確認した事項、【分析】はそれらに基づく推論・提言を示す。上付きの番号は文末の参考文献に対応する。参考文献で「未検証」「二次情報」と付したものは、本調査でリンク先の原文を直接確認できていないことを示す。

要旨

知財業務向けハーネス設計の核心は、「モデルに任せる部分」と「ハーネスが決定論的に強制する部分」を明確に分けることにある。システムプロンプト等の指示文はエージェントの意図を形づくるが、許可範囲そのものは変えない^[1]。したがって次の3点は、プロンプトではなく権限設定・フック・サンドボックス・ワークフローに組み込む必要がある。

- ① **機密境界**：未公開発明を、どのモデル・どの基盤に、どの契約条件で渡すかを案件単位で強制する^[2]。
- ② **出典の機械照合**：特許番号・請求項・引用箇所を一次 DB と照合するまで成果物として出力しない。2026 年 8 月には USPTO が生成 AI による誤引用を理由とする初の懲戒を公表した^[3]。
- ③ **承認ゲート**：出願・応答・期限管理・外部送信など不可逆な操作は、人間の承認なしに実行できない構造にする^[1,4]。

加えて、外部 DB の利用規約（自動取得の禁止）^[5,6]、ログの開示リスク^[7]、自社データによる評価^[8]も、導入前に設計対象として扱うべきである。

1. 要点整理

観点	確認された事実（一次ソース中心）	設計への示唆（分析）
一般原則	最も単純な解から始め、必要な場合にのみ複雑さを加えること、「ワークフロー」と「エージェント」の区別、透明性を推奨 ^[9]	定型業務（期限計算・書誌突合）はワークフロー、探索型業務（先行技術調査・無効資料調査）はエージェントに仕分ける
コンテキスト	トークン量の増加で想起精度が落ちる「context rot」を指摘し、圧縮・構造化ノート・サブエージェントを推奨 ^[10]	公報全文を丸ごと投入せず、文献 ID と引用箇所だけを保持し、必要時に一次 DB から再取得する
長時間タスク	初期化エージェントと進捗ファイルによる引継ぎ ^[11] 、生成者と評価者の分離 ^[12]	調査計画書・検索式ログ・スクリーニング表を引継ぎファイルとし、評価は別エージェントに担わせる
権限	指示文は許可範囲を変えないと明記し、権限とサンドボックスの多層防御を推奨 ^[1,13]	出願系 API・メール送信・外部アップロードは原則 deny、または都度承認（ask）
日本の職能規範	守秘義務（弁理士法 30 条）、新規性喪失、学習の有無を問わないクライアント合意、弁理士によるファクトチェックを要請 ^[2]	案件ごとに「AI 利用可否」「利用可能なモデル・基盤」をハーネス側のメタデータで管理する
米欧の庁	USPTO は AI 利用を禁止しないが署名者の合理的調査義務・誠実義務が及ぶとし ^[14] 、EPO も AI 使用の有無を問わず当事者・代理人が責任を負うと明記 ^[15]	「人間が確認した」証跡（誰が・何を・どの版を）をログに残す
外部データ	J-PlatPat はロボットアクセスを禁止 ^[5] 。特許情報取得 API は機械的アクセスの頻度調整を要求 ^[16] 。Questel Orbit の規約は自動化エージェントによるアクセスを禁止 ^[6]	ブラウザ操作によるスクレイピングをさせず、正規 API・契約上許された経路のみをツール化する

2. ハーネス設計の一般原則と最新動向（2025～2026 年）

2.1 Anthropic の公表文書

【事実】

- **Building effective agents（2024 年 12 月）**：定義済みのコード経路で LLM とツールを組み合わせる「ワークフロー」と、モデル自身が手順を決める「エージェント」を区別する。単純さの維持、計画ステップの明示（透明性）、エージェント向けツールの丁寧な設計を原則とする^[9]。
- **Effective context engineering for AI agents（2025 年 9 月 29 日）**：コンテキストを有限の「注意予算」として扱い、望ましい結果をもたらす最小限の高シグナルなトークン集合を探すことを原

則とする。典型的な失敗として、機能が重複しどのツールを使うべきか曖昧になる肥大化したツールセットを挙げ、人間が使うべきツールを断言できないならエージェントにも期待できないとする。長時間タスク向けに、圧縮（compaction）、構造化ノート（エージェント外部のメモリ）、サブエージェント（詳細探索を担い、多くの場合 1,000~2,000 トークン程度の要約だけを返す）を紹介している。事前取得と自律探索を併用するハイブリッド戦略は、法務・金融のように変化の少ない領域に向く可能性があるとしている^[10]。

- **Effective harnesses for long-running agents（2025 年 11 月 26 日）**：初回のみ動く初期化エージェントが機能リスト・git・進捗ファイルを用意し、以降のエージェントは 1 機能ずつ進めて、構造化された成果物で次のセッションへ引き継ぐ^[11]。
- **Harness design for long-running application development（2026 年 3 月 24 日）**：プランナー・生成者・評価者の 3 エージェント構成。同一エージェントに生成と評価をさせると評価が甘くなると指摘し、着手前に生成者と評価者が完了条件を合意する「スプリント契約」を導入している^[12]。
- **Demystifying evals for AI agents（2026 年 1 月）**：タスク・試行・トランスクリプト・結果を区別し、評価対象は「モデル+ハーネス」の組合せであるとする。pass@k（k 回中 1 回成功）と pass[^]k（k 回すべて成功）の使い分け、採点器を検証するためのトランスクリプトの精読を推奨している^[8]。
- **Claude Code のセキュリティ関連文書**：プロンプトや CLAUDE.md はエージェントが「何をしようとするか」を形づくるが、「何が許可されるか」は変えないと明記している。権限（ツール・ファイル・ドメイン）と OS レベルのサンドボックス（ファイル・ネットワーク隔離）を補完的に用いる多層防御を推奨し、プロンプトインジェクションがモデルの判断をすり抜けてもサンドボックスの制約は残ると説明する。ツール実行前に走る PreToolUse フックによる拒否・承認要求の強制、プロキシ経由に限定した通信による任意サーバーへの持ち出し防止も示されている^[1,13,17]。

2.2 OpenAI Agents SDK

【事実】

- 構成要素はガードレール（入力・出力・ツール）、ヒューマン・イン・ザ・ループ、セッション、トレーシング。取消・編集・シェル実行・機微な MCP 操作など副作用の前で人間の承認により実行を止める用途を例示し、承認待ちで一時停止した実行を同じ状態から再開できる^[4,18]。

- Python 版はトレーシングが既定で有効であり、環境変数（`OPENAI_AGENTS_DISABLE_TRACING=1`）、`set_tracing_disabled(True)`、`RunConfig.tracing_disabled`で無効化できる^[19]。JS版は、ゼロデータリテンション（ZDR）ポリシー下でOpenAI APIを使う組織ではトレーシングを利用できないとしている^[20]。

2.3 知財業務への翻訳

【分析】

- ハーネスは「事務所・知財部の業務規程をコード化したもの」と捉えると実務に落とし込みやすい。業務規程のうち、守らなければ事故になる部分（送信先・提出・期限・機密区分）はフックと権限で強制し、品質の勘所（クレーム解釈の観点、調査の網羅基準）はスキルやシステムプロンプトで誘導する、という二層構造が基本となる。
- トレーシング等の周辺機能も送信経路になり得る。トレースに未公開発明の内容が含まれる場合は、送信先・保存期間・無効化設定を事前に確認する。

3. 知財業務特有の注意点とハーネスへの落とし込み

3.1 機密性（未公開発明・出願前情報）

【事実】

- 日本弁理士会ガイドラインは、①入力情報が学習に使われるか否かは新規性喪失や秘密保持契約の遵守に密接に関係すること、②秘密保持規約のない第三者プラットフォームに情報を載せると新規性を喪失する可能性があること、③外部事業者の生成 AI への秘密情報の入力守秘義務（弁理士法 30 条）違反のおそれがあること、④クライアントの秘密情報を入力する場合は学習の有無によらず合意を取るべきことを示している^[2]。
- USPTO ガイダンスは、機密情報の意図しない開示に加え、海外サーバー上の AI ツールの利用が輸出管理・外国出願許可・国家安全保障上の問題を生じ得る点に注意を促している^[14]。

【分析】

- 案件メタデータに機密区分と AI 利用可否（クライアント合意の有無）を持たせる。未公開・出願前案件は、ZDR 契約のある、または自社テナント内（国内リージョン・閉域）のモデルにのみルーティングし、それ以外への送信はフックで遮断する。
- 外部検索のクエリから発明内容を漏らさない。Web 検索・外部 DB 検索のクエリ文字列自体が第三者への開示になり得る。送信前に機密語をマスキングし、外部の公開検索は公開済み情報に限定するのが安全側である。

- サブエージェントや MCP サーバーにも同じ境界を継承させる。Claude Code の文書では、フックやローカル MCP サーバーはサンドボックス外で動作するとされており^[13]、親の制約が自動的に及ぶとは限らない。
- クラウドかオンプレかは二者択一ではない。「データ分類×契約条件（学習不使用・保持期間・所在地）×監査可能性」で決める。弁理士会ガイドラインは秘密保持規約等の有無を基準にしており^[2]、技術的な所在より契約・規約の確認が先決である。

3.2 正確性（ハルシネーション対策と一次ソースへの接地）

【事実】

- USPTO の OED（登録・懲戒局）は 2026 年 8 月 27 日、生成 AI が作成した存在しない引用（明細書・図面・審査経過という内部証拠への引用）を含むクレーム解釈チャートを地裁に提出した特許弁理士を戒告する最終命令を公表した（In re Brian E. Mitchell, D2026-16）。IPWatchdog は、生成 AI 利用を前提とする USPTO 初の懲戒命令と報じている。本人は AI の限界の理解不足と検証不足を認めた^[3]。二次情報によれば、本人は起案用とは別の AI ツールでレビューすれば足りると考えていた^[21]。
- USPTO ガイダンスは、AI で書類の一部を起案した場合でも引用の正確性と主張の法的正当性の確認を求め、署名者には 37 CFR 11.18(b)の合理的調査義務が及ぶとする^[14]。弁理士会ガイドラインも、ファクトチェックは必ず弁理士が行う必要があるとしている^[2]。

【分析】

- 「AI で別の AI をレビューさせた」だけでは不十分というのが、上記懲戒事例の実務的教訓である。LLM による評価は有用だが、番号・引用箇所は決定論的な照合ツールで検証すべきである。推奨する検証パイプラインは次のとおり。

- (1) 出力中の特許番号・出願番号・請求項番号・段落番号・判例名を構造化して抽出する。
- (2) 正規 API または契約済み商用 DB で実在を確認し、引用文の文字列一致まで照合する。
- (3) 不一致・未検証の箇所にはフラグを付け、成果物として出力させない（Stop フック等）。
- (4) 検証結果をログに残す。

- 引用の「存在」だけでなく「内容」も確認する。実在文献の誤要約は番号照合では検出できない。引用箇所の原文を並記し、人間が対照確認できる画面設計とする。

- 評価者エージェントは別コンテキスト・書込み権限なし・可能なら異なるモデルとする（生成者と評価者の分離^[12]を応用した推論）。

3.3 法的・制度的制約

【事実】

- **弁理士法 75 条**：日本弁理士会は 2025 年 4 月、弁理士でない事業者が AI で明細書・特許請求の範囲を作成するサービスや、願書を半自動作成して出願手続を行うサービスを提供する形態について見解を公表した。同時に、非弁行為該当性は個別具体的に判断され、解釈・適用は最終的に裁判所に委ねられると明記している^[22]。
- **弁護士法 72 条**：法務省は 2023 年 8 月、契約書等関連業務支援サービスについて、弁護士が自ら精査し必要に応じ修正する方法で使用する場合は同条に違反しないとの考え方を示した^[23]。
- **USPTO**：AI ツールの利用は禁止されておらず、利用の一般的な開示義務もない。ただし特許性に重要な AI 利用は誠実義務の対象になり得るとし^[14]、AI ツールによる書類の自動提出には慎重な姿勢を示している^[24]。
- **EPO・epi**：EPO 審査ガイドライン 2025 年版（General Part 5）は、AI ツールの支援の有無にかかわらず当事者・代理人が提出書類の内容に責任を負うと明記し^[15]、2026 年版で独立した節に拡充された^[25]。epi のガイドライン（2024 年 11 月 16 日採択）は、法が特に求めない限り EPO・UPC への AI 利用の開示義務はないとしている^[26]。

【分析】

- 企業知財部が自社案件に使う場合と、特許事務所や無資格事業者が他人のために報酬を得て使う場合とで、75 条リスクの所在は異なる。後者で「書類作成から提出まで」をエージェントが一気通貫で行う設計は制度上最も議論を呼ぶ形態であり、弁理士の精査・修正・承認の工程を技術的に飛ばせない形で組み込むべきである。
- 法令解釈には未確定の部分が残る。社外向けサービスとして提供する場合は、グレーゾーン解消制度の活用や弁理士会見解の更新を継続的に確認する前提で設計する。

3.4 外部データ・ツール利用時のリスク

【事実】

- **J-PlatPat**：一般の利用を妨げるおそれのある大量ダウンロードやロボットアクセス（プログラムによる定期的な自動データ収集）を禁止し、発見時は予告なくアクセスを制限するとしている。プロフェッショナル用途には民間サービスの利用を推奨している^[5]。

- **特許庁の特許情報取得 API**：利用申請制で、機械的アクセスを国内特許情報取得 API は 1 分 10 回以下、OPD（ワン・ポータル・ドシエ）関連 API は 1 分 5 回以下に調整するよう求めている^[16]。
- **商用 DB 等の規約**：Questel Orbit（2020 年 7 月版）は、ロボットや自動化されたインテリジェント・エージェント等によるアクセス・情報取得、データベースの全部または一部のダウンロード、情報を用いた派生物の構築を禁止している^[6]。PatSnap Eureka（2026 年 6 月 15 日更新、個人・セルフサービス向け）は、プラン等で明示的に許された場合を除き自動的手段によるスクレイピング・データ抽出を禁止し、プロンプトや出力を第三者のモデル・MCP サーバーへ送ることは利用者の自己責任としている^[27]。Patentfield もアクセス回数制限の意図的な回避やロボット等による不正アクセスを禁止事項に挙げている^[28]。Espacenet は、異常に多いアクセスで他の利用者を妨げる者に対し予告なくアクセスを拒否し得るとしている^[29]（自動取得には OPS を使うべきとするフェアユース規定は二次情報でのみ確認^[30]）。
- **プロンプトインジェクション**：Anthropic は、エージェントの挙動が処理対象のコンテンツ（ファイル・Web ページ等）に影響され得るリスクに対し、権限・サンドボックス・プロキシによる多層防御を推奨している^[13,17]。

【分析】

- **ブラウザ操作エージェントに J-PlatPat を巡回させる設計は避ける**。規約上のリスクに加え、アクセス制限を受ければ業務が止まる。正規 API・契約済み API・MCP のみをツールとして登録し、レート制限はツール側で強制する（モデルの自制に頼らない）。
- **商用 DB は、契約上「AI エージェント経由のアクセス」と「LLM への投入」が許されるかを個別に確認する**。許諾範囲をツール定義に反映し、ベンダー公式の API・MCP を優先する。
- **特許文献・拒絶理由通知書・Web ページは「データ」であり「命令」ではない**。外部取得テキストは、書込み・送信権限を持たない読取り専用サブエージェントで要約してから親エージェントに渡す。公開公報や第三者ページに埋め込まれた指示文によって送信・削除・外部アクセスが実行される経路を、物理的に断つことが目的である。

3.5 人間の関与（承認ゲートと責任分界）

【事実】

- OpenAI Agents SDK は副作用の前で承認により実行を止める仕組みを公式機能として提供し^[4]、Claude Code もツールごとの allow/ask/deny 設定とフックによる承認要求の強制が可能である

[1]。USPTO は AI による自動提出に慎重な姿勢を示し^[24]、弁理士会ガイドラインは最終責任が弁理士にあるとしている^[2]。

【分析】

操作のリスクに応じて、次の 4 階層で承認ゲートを設計することを提案する。

階層	操作例	ハーンネス設定
L0 読取り	公開公報の検索・要約、社内ナレッジ参照	自動（ログのみ）
L1 下書き	明細書ドラフト、応答案、クレームチャート、調査報告ドラフト	自動生成可。「未検証」透かしと検証フラグを必須とする
L2 外部影響	クライアント・発明者・代理人へのメール送信、外部 DB への機密クエリ	都度承認（ask）。送信先はホワイトリスト管理
L3 不可逆	出願・応答・補正・年金納付・期限管理 DB の更新・放棄	原則 deny。担当弁理士・責任者の二重承認と承認者 ID の記録を条件に、エージェント外の正規システムで実行

- 期限管理はエージェントに「判断」させず、「検知・提案」に限定する。期限計算は決定論的なルールエンジンで行い、LLM は差分の説明や対応案の起案にとどめる。

3.6 監査・トレーサビリティ

【事実】

- United States v. Heppner (S.D.N.Y., Rakoff 判事、2026 年 2 月 17 日付意見書) は、刑事被告人が弁護士の指示なく一般消費者向けの Claude とやり取りした文書について、秘匿特権もワークプロダクトの保護も認めなかった。理由として、AI は弁護士ではないこと、プライバシーポリシーに照らすと機密性の合理的期待がないこと、弁護士の指示によるものではないことを挙げている^[7]。法律事務所の解説は、機密性は判断理由の一つにすぎず、エンタープライズ版でも法的要請に応じた開示の権利を留保している場合があると指摘する^[31]。一方、同判決には批判的な論考もある^[32,33]。
- USPTO の ASAP!パイロット（2025 年 10 月開始）では、AI が生成した検索結果通知（ASRN）が最大 10 件の文献と検索式を示し、公開時に包袋の一部となる^[34,35]。

【分析】

- FTO・鑑定業務では、「弁護士・弁理士の指示の下で AI を使った」ことが分かる形でログを残す。米国訴訟の可能性がある案件では、エンタープライズ契約、指示記録、アクセス権限の限定が防御線となる。ただし同判決は米国刑事事件の地裁判断であり、評価も分かれている。
- ログは「立証資産」であると同時に「開示リスク」でもある。検索式・母集団・スクリーニング判断・採否理由の保存は FTO 調査の網羅性の立証に役立つが、ディスカバリー対象となり得ることも踏まえ、保存範囲・期間・アクセス権限を法務と事前に決める。
- 庁側の AI 出力も記録対象になる。庁の AI 検索結果が包袋に残る以上、出願人側の AI 調査結果との差分管理（何を把握していたか）が、IDS や誠実義務の観点で重要になる（米国案件についての推論）。

3.7 評価 (eval)

【事実】

- PatRe (arXiv:2605.03571、2026 年 5 月) は、Office Action 生成と出願人の反論を含む特許審査の全過程をモデル化した最初のベンチマークと称し、480 件の実案件と 2 つの評価設定 (oracle / retrieval-simulated) を備える。LLM は審査官役より出願人役 (反論生成) で良い成績を示し、LLM 採点と専門家評価の相関は $r=0.7285$ と報告されている^[36]。なお、OA 生成の判断正解率が上位モデルでも約 50~52.8%にとどまるとの数値は、要約サイトによる二次情報である (URL 未特定)。
- 特許応答生成システム PRO は、GPT-4 比で 6 種類の誤りにわたり忠実性が 39%高いと報告されている (他論文の先行研究紹介による記述)^[37]。特許起案エージェントを評価する LLM ジャッジ自体の評価を扱う研究も現れている^[38]。

【分析】

- 公開ベンチマークは「足切り」用途にとどめ、自社のゴールデンセットを作る。推奨は次の 3 点である。

(1) 過去の自社・事務所案件から、正解の引例 (審査官引用 + 社内で後日判明した文献) 付きの調査課題を作り、再現率 (recall) を主指標とする。調査では見落としが致命的なため、適合率より再現率を重視する。

(2) 応答は、補正の根拠記載 (新規事項追加の有無)、引用の正確性、禁反言リスクの観点別ループバックで採点する。

(3) 同一課題の複数回試行で pass^k を測り、ばらつきを把握する。業務品質として求められるのは「k 回すべて合格」である^[8]。

- 評価対象は「モデル＋ハーネス」であるため^[8]、モデル更新・プロンプト変更・ツール追加のたびに回帰評価を回す。

4. 実例・公表事例

【事実】

- **USPTO**：2025 年 10 月 20 日に ASAP!を開始した。内部 AI ツールが CPC 分類・明細書・請求の範囲・要約から先行技術を検索し、関連度順に最大 10 件を通知する^[34]。2026 年 4 月 16 日に延長と手数料免除が発表された^[39]。
- **特許庁（日本）**：「AI 技術の活用に向けたアクション・プラン」令和 7 年度改定版で「生成 AI の特許審査業務への適用」を新設し、令和 7 年度に技術実証を行うとした^[40]。
- **EPO**：2025 年 5 月から、テレビ会議による口頭審理の議事録作成を AI で支援するパイロットを開始した（録音は議事録発行後に削除）^[41]。
- **NTT ドコモビジネス**：2026 年 7 月 30 日に「アイデアシート作成支援 AI エージェント」の提供を開始した。キャノンが同年 7 月から一部の開発部門でパイロット利用しており、明細書案作成・先行技術調査・拒絶理由対応向けのエージェントへの展開も予定されている^[42]。
- **オムロン**：研究開発向け生成 AI 基盤「RD Buddy」上に知財 AI エージェントを実装した事例を公表している^[43]（詳細は本稿では未精査）。
- **ストックマーク**：「Aconnect」の特許調査エージェントに、関連性判断の条件を言語化する支援機能を追加した^[44]。
- **失敗例**：前述の USPTO 懲戒事例^[3]。米国の裁判所では、Mata v. Avianca 事件で架空判例 6 件を引用した弁護士 2 名と所属事務所に、連帯で 5,000 ドルの Rule 11 制裁が科された^[45]。生成 AI のハルシネーションが問題となった事例を集積するデータベースは、2026 年 10 月 25 日更新時点で 2,217 件を収録している^[46]。

【分析】

- 公表事例の多くは、発明整理・調査の初期スクリーニング・議事録といった、可逆的で人間のレビューを前提とする工程である。出願・応答の提出まで自律実行する公表事例は確認できなかった。L0～L1 から始め、L3 は人間に残す段階的導入が、現時点の実務的な均衡点である。

5. 実務向けチェックリスト

A. 設計方針

- ☐ 業務ごとに「ワークフロー（固定手順）」と「エージェント（探索）」を仕分けたか（期限計算・書誌突合はワークフロー）
- ☐ 「プロンプトで頼むこと」と「権限・フックで強制すること」を一覧で分けたか

B. 機密・データ

- ☐ 案件ごとに機密区分と AI 利用可否（クライアント合意の有無）をメタデータで持ち、ルーティングを強制しているか
- ☐ 利用モデル・基盤の学習不使用、保持期間（ZDR）、所在地、法的開示時の扱いを契約で確認したか
- ☐ 外部検索クエリへの発明内容の流出を、マスキングまたは遮断しているか
- ☐ トレーシングや評価ツールなど周辺 SaaS へのデータ送信を棚卸したか

C. 正確性

- ☐ 特許番号・請求項・段落・判例の自動抽出と一次 DB 照合を実装し、未検証のものは出力をブロックしているか
- ☐ 引用原文を並記し、人間が対照確認できる画面になっているか
- ☐ 評価エージェントは別コンテキストで、書込み権限を持たない構成か

D. 外部データ・セキュリティ

- ☐ J-PlatPat 等のスクレイピングを禁止し、正規 API（レート制限をツール側で強制）だけを登録したか
- ☐ 商用 DB 契約で、AI エージェント経由のアクセスと LLM への投入の可否を確認したか
- ☐ 外部取得テキストは読取り専用サブエージェントで処理し、送信・書込み権限から隔離しているか
- ☐ MCP サーバーやプラグインの出所を管理し、フックを管理者配布に限定しているか

E. 人間の関与

- ☐ L2（外部送信）は都度承認、L3（出願・応答・期限・放棄）はエージェントから実行できない構造にしたか
- ☐ 承認者 ID・承認対象の版・承認日時を記録しているか

- ☐ 弁理士（弁護士）が精査・修正する工程を技術的に飛ばせない構造か（弁理士法 75 条・弁理士法 72 条への対応）

F. 監査・ログ

- ☐ 検索式・母集団・スクリーニング判断・採否理由を再現可能な形で保存しているか
- ☐ ログの保存範囲・期間・アクセス権限を法務と決めたか（米国案件はディスカバリー・秘匿特権も考慮）

G. 評価・運用

- ☐ 自社のゴールデンセット（調査の再現率、応答のループバック）を作ったか
- ☐ モデル・プロンプト・ツールの変更時に回帰評価を回す運用になっているか
- ☐ インシデント（誤引用・誤送信）の報告とハーネス修正のループがあるか

6. 留意事項

- **法令解釈は未確定である。** 弁理士法 75 条と AI サービスの関係は弁理士会の見解にとどまり、最終判断は裁判所に委ねられている^[22]。Heppner 判決は米国刑事事件の地裁判断であり、日本法上の秘密保持・秘匿特権の扱いに直接は当てはまらない。
- **二次情報に依拠した点がある。** USPTO 懲戒事例における「別の AI でレビュー」という事情、PatRe の判断正解率の数値、EPO のフェアユース規定・OPS 無料枠は二次情報による。オムロン事例の詳細は未精査である。Clarivate Derwent、LexisNexis、Shareresearch、CyberPatent Desk の規約原文は確認できていない。
- **規約・製品仕様は頻繁に変わる。** 弁理士会ガイドラインも規約の定期的な確認を求めている^[2]。本稿の規約引用（Questel は 2020 年 7 月版、PatSnap は 2026 年 6 月更新版）は、導入時に必ず最新版を確認されたい。

参考文献

-
- [1] Anthropic 「Claude Code Docs: Permissions」 <https://code.claude.com/docs/en/permissions>
 - [2] 日本弁理士会「弁理士業務 AI 利活用ガイドライン」2025 年 4 月 <https://www.jpaa.or.jp/cms/wp-content/uploads/2025/04/AI-services-guideline.pdf>
 - [3] IPWatchdog 「USPTO Issues Its First AI-Predicated Discipline Order Involving Hallucinated Cites to Intrinsic Record」2026 年 8 月 28 日 <https://ipwatchdog.com/2026/08/28/uspto-issues-its-first-ai-predicated-discipline-order-involving-hallucinated-cites-to-intrinsic-record/>
 - [4] OpenAI 「Guardrails and human review (Agents)」OpenAI API Docs <https://developers.openai.com/api/docs/guides/agents/guardrails-approvals>

- [5] INPIT「J-PlatPat ご利用上のご案内」 <https://www.j-platpat.inpit.go.jp/help/ja/c09/c0900.html>
- [6] Questel「Orbit / IPBI General Terms of Sale」2020 年 7 月版 <https://www.questel.com/wp-content/uploads/2020/07/orbit-ipbi-gts-july2020.pdf> ※より新しい版の有無は未確認
- [7] United States v. Heppner (S.D.N.Y., Rakoff 判事、2026 年 2 月 17 日付意見書) の解説 : Debevoise Data Blog (2026 年 2 月 17 日)、O'Melveny「SDNY First-of-Its-Kind Ruling: AI-Generated Documents Are Not Privileged」 <https://www.debevoisedatablog.com/2026/02/17/update-judge-rakoff-issues-written-opinion-that-ai-generated-documents-are-not-protected-by-privilege/> <https://www.omm.com/insights/alerts-publications/sdny-first-of-its-kind-ruling-ai-generated-documents-are-not-privileged/> ※意見書原文は未参照
- [8] Anthropic「Demystifying evals for AI agents」Anthropic Engineering, 2026 年 1 月 ※原文 URL は未検証。内容は要約記事 (<https://www.beri.net/learning/anthropic-demystifying-evals-ai-agents>) で確認
- [9] Anthropic「Building effective agents」Anthropic Engineering, 2024 年 12 月 <https://www.anthropic.com/engineering/building-effective-agents>
- [10] Anthropic「Effective context engineering for AI agents」Anthropic Engineering, 2025 年 9 月 29 日 <https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents>
- [11] Anthropic「Effective harnesses for long-running agents」Anthropic Engineering, 2025 年 11 月 26 日 <https://www.anthropic.com/engineering/effective-harnesses-for-long-running-agents> ※原文 URL は未検証。内容は要約記事 (<https://www.beri.net/learning/anthropic-effective-harnesses-long-running-agents>) で確認
- [12] Anthropic「Harness design for long-running application development」Anthropic Engineering, 2026 年 3 月 24 日 ※原文 URL は未検証。内容は要約記事 (<https://agentsdesign.dev/article/anthropic-harness-design-long-running-apps/>) で確認
- [13] Anthropic「Claude Code Docs: Sandboxing」 <https://code.claude.com/docs/en/sandboxing>
- [14] USPTO「Guidance on Use of Artificial Intelligence-Based Tools in Practice Before the United States Patent and Trademark Office」Federal Register, 2024 年 4 月 11 日 <https://www.federalregister.gov/documents/2024/04/11/2024-07629/guidance-on-use-of-artificial-intelligence-based-tools-in-practice-before-the-united-states-patent>
- [15] EPO 審査ガイドライン 2025 年版の AI 関連改訂に関する解説 : Solve Intelligence「EPO Guidelines 2025」、Boult「Updates to EPO Guidelines relating to Artificial Intelligence」 <https://www.solveintelligence.com/blog/post/epo-guidelines-2025> <https://www.boult.com/bulletin/updates-to-epo-guidelines-relating-to-artificial-intelligence/> ※いずれも二次情報 (ガイドライン原文は未参照)
- [16] 特許庁「特許情報取得 API の利用の手引き」第 1.3 版 <https://ip-data.jpo.go.jp/files/%E7%89%B9%E8%A8%B1%E6%83%85%E5%A0%B1%E5%8F%96%E5%BE%97API%E3%81%AE%E5%88%A9%E7%94%A8%E3%81%AE%E6%89%8B%E5%BC%95%E3%81%8Dv1.3.pdf>
- [17] Anthropic「Claude Agent SDK Docs: Secure deployment」 <https://code.claude.com/docs/en/agent-sdk/secure-deployment>
- [18] OpenAI「Human-in-the-loop」OpenAI Agents SDK (JavaScript) Docs <https://openai.github.io/openai-agents-js/guides/human-in-the-loop/>
- [19] OpenAI「openai-agents-python」GitHub リポジトリ (Tracing 設定の記載) <https://github.com/openai/openai-agents-python>
- [20] OpenAI「openai-agents-js」GitHub リポジトリ (ZDR 下での Tracing 利用制限の記載) <https://github.com/openai/openai-agents-js>
- [21] PAICE (Substack)、生成 AI による誤引用と裁判所に関する記事 <https://paice.substack.com/p/ai-hallucinated-citations-courts> ※二次情報

- [22] 日本弁理士会「AI 等を用いた業務支援サービスの提供と弁理士法第 75 条との関係について」2025 年 4 月
<https://www.jpaa.or.jp/cms/wp-content/uploads/2025/04/AIservices-article75.pdf>
- [23] 法務省「AI 等を用いた契約書等関連業務支援サービスの提供と弁理士法第 72 条との関係について」2023 年 8 月
<https://www.moj.go.jp/content/001400674.pdf>
- [24] Greenberg Traurig「USPTO Issues Guidance on Use of AI-Based Tools」2024 年 4 月（解説）
<https://www.gtlaw.com/en/insights/2024/4/uspto-issues-guidance-on-use-of-ai-based-tools>
- [25] EPO 審査ガイドライン 2026 年版に関する解説：Solve Intelligence「EPO Guidelines 2026: key changes」、Mathys & Squire「European Patent Office previews the 2026 Guidelines for Examination」
<https://www.solveintelligence.com/blog/post/epo-guidelines-2026-key-changes> <https://www.mathys-squire.com/insights-and-events/news/european-patent-office-previews-the-2026-guidelines-for-examination/> ※いずれも二次情報
- [26] epi（欧州特許代理人協会）、生成 AI の利用に関するガイドライン（2024 年 11 月 16 日採択） ※原文 URL 未確認
- [27] PatSnap「Eureka Terms of Service」2026 年 6 月 15 日更新 <https://eureka.patsnap.com/terms-of-service/>
- [28] Patentfield「利用規約」第 5 条 https://en.patentfield.com/terms_of_use ※検索スニペットでの確認。全文・改定日は未検証
- [29] EPO「Espacenet: Disclaimer（旧利用条件）」 https://ep.espacenet.com/?locale=en_EP&view=disclaimer ※検索スニペットでの確認
- [30] prudai「EPO OPS」（EPO フェアユース規定・OPS 無料枠の解説） <https://prudai.com/en/sources/epo-ops> ※二次情報
- [31] Herbert Smith Freehills Kramer「New York court finds that client chats with Claude are not privileged」2026 年 2 月 <https://www.hsfkramer.com/insights/2026-02/new-york-court-finds-that-client-chats-with-claude-are-not-privileged>
- [32] Harvard Law Review Blog「United States v. Heppner」2026 年 3 月
<https://harvardlawreview.org/blog/2026/03/united-states-v-heppner/>
- [33] Lawfare「AI and Privilege After United States v. Heppner」 <https://www.lawfaremedia.org/article/ai-and-privilege-after-united-states-v.-heppner>
- [34] USPTO「USPTO launches new AI pilot for pre-examination utility application search」2025 年 10 月
<https://www.uspto.gov/about-us/news-updates/uspto-launches-new-ai-pilot-pre-examination-utility-application-search>
- [35] PatentNext「USPTO's New Pilot Program Offers an Early Look at AI-Generated Search Results」2025 年 10 月
<https://www.patentnext.com/2025/10/usptos-new-pilot-program-offers-an-early-look-at-ai-generated-search-results/>
- [36] PatRe（特許審査ライフサイクルのベンチマーク）、arXiv:2605.03571, 2026 年 5 月 5 日投稿
<https://arxiv.org/abs/2605.03571>
- [37] arXiv:2508.14064（特許応答システム PRO への言及を含む論文） <https://arxiv.org/pdf/2508.14064> ※PRO の数値は同論文の先行研究紹介による。PRO 原論文は未確認
- [38] 「Vibe Patenting」arXiv:2609.13422 <https://arxiv.org/pdf/2609.13422> ※本文詳細は未確認
- [39] Nixon Peabody「USPTO extends AI-driven prior art search pilot and waives petition fee」2026 年 4 月 22 日
<https://www.nixonpeabody.com/insights/alerts/2026/04/22/uspto-extends-ai-driven-prior-art-search-pilot-and-waives-petition-fee>
- [40] 特許庁「AI 技術の活用に向けたアクション・プラン（令和 7 年度改定版）」
https://www.jpo.go.jp/system/laws/sesaku/ai_action_plan/ai_action_plan-fy2025.html

- [41] EPO の口頭審理議事録 AI 支援パイロットに関する解説：Solve Intelligence、Mondaq
<https://www.solveintelligence.com/blog/post/epo-to-use-ai-tools-minuting-oral-proceedings>
<https://www.mondaq.com/unitedstates/patent/1610024/epo-to-use-ai-tools-to-prepare-minutes-from-vc-oral-proceedings>
※二次情報
- [42] NTT ドコモビジネス ニュースリリース（2026 年 7 月 30 日）：アイデアシート作成支援 AI エージェントの提供開始 <https://www.ntt.com/about-us/press-releases/news/article/2026/0730.html>
- [43] AWS builders.flash、オムロンの知財 AI エージェント事例（2025 年 11 月 4 日）
<https://aws.amazon.com/jp/builders-flash/202511/omron-intellectual-property-ai-agent> ※本稿では内容未精査
- [44] ストックマーク株式会社 プレスリリース（PR TIMES、2026 年 3 月 5 日）：Aconnect 特許調査エージェントの機能追加 <https://prtimes.jp/main/html/rd/p/000000362.000024407.html>
- [45] Mata v. Avianca, Inc., No. 22-cv-1461 (S.D.N.Y., 2023 年 6 月 22 日) ※判決原文は未取得。複数の二次トラッカーで確認
- [46] Damien Charlotin (HEC Paris) 「AI Hallucination Cases Database」 (2026 年 10 月 25 日更新時点)
※URL 未検証 (<https://www.damiencharlotin.com/hallucinations/> と推定)