ChatGPT

ループエンジニアリング時代の自律型AIと知的財産業務

エグゼクティブサマリ

調査基準日は日本時間の二〇二六年八月一日である。 本レポートでは、日本企業の知財・法務部門を主たる対象とし、日本法を中心に、米国・欧州への特許出願、海外クラウド・AIサービスの利用、国際的なライセンス取引を行う中堅・大企業を想定する。

結論からいえば、AI活用の競争軸は、優れた指示文を一回入力する「プロンプトエンジニアリング」から、AIが目標達成まで計画、実行、観察、評価、修正を反復するシステム全体を設計する**「ループエンジニアリング」**へ移り始めている。ただし、プロンプトが不要になるのではない。プロンプト、コンテキスト、ツール、権限、記憶、評価、監視、ログ、人間の承認を一体として設計する上位概念に移行する、という理解が正確である。「ループエンジニアリング」はまだ法令・標準規格上の確立した用語ではないが、二〇二六年には業界で、AIが「行動―観察―判断―反復」を続けるワークフローの設計を指す言葉として使われ始めている。日本政府の人工知能基本計画も、AIが文書作成支援から、計画、実行、検証、修正を繰り返して業務全体を自律的に動かすAIエージェントへ移行していると認識している。 [1]

知財業務で最も大きな変化は、明細書や契約書の「生成」そのものではない。発明発掘、先行技術調査、クレーム設計、拒絶理由対応、競合監視、年金・期限管理、ライセンス候補探索、証拠収集を、継続的な閉ループとして運用できるようになることである。一方、AIの稼働時間が長くなり、アクセスする情報源と実行可能なツールが増えるほど、誤った出願、秘密情報の外部送信、権限を欠く契約行為、著作権侵害を含む出力の大量生成、ログへの機密情報混入、発明者・著作者の帰属紛争といった損害も拡大する。

したがって、自律性を「人間が関与しないこと」と定義すべきではない。知財領域で実用的な自律性とは、**目的、データ、権限、予算、期間、停止条件が明確な境界内で、AIが反復処理を行い、法的効果または不可逆性を伴う地点では人間が承認する「境界付き自律性」**である。

特に重要なのは、AIのログを単なるIT運用記録ではなく、次の事項を立証する**知財証拠基盤**として位置付けることである。

- どの人が、どの技術的課題、仮説、選択、検証に創作的に関与したか
- どの資料がAIに提供され、どのモデル、ツール、ルールで処理されたか
- 営業秘密について合理的なアクセス管理・秘密管理措置が採られていたか
- AI出力を人間がどのように選択、修正、承認したか
- 出願、公開、契約締結等の外部行為を誰が承認したか
- 第三者著作物、OSS、データベース、契約データの利用条件を確認したか

日本の現行特許法下では発明者は自然人であることが前提とされ、二〇二五年一月三十日の知財高裁判決も、権利能力のないAIを発明者とする特許付与手続は現行法上予定されていないとの判断を示した。したがって、当面の中心課題は「AIを発明者として記載するか」ではなく、AI利用下で**自然人の創作的寄与をどのように認定し、証明するか**である。 [2]

本調査の主要提言は次のとおりである。

優先度	推奨事項	実務上の意味
最優先	全AIエージェントを登録し、権限・データ・外部行為をリスク分類する	未管理の「野良エージェント」をなくす
最優先	出願、公開、放棄、契約承諾、秘密情報の外部送信に人間の承認ゲートを設ける	法的・不可逆的行為をAI単独で実行させない
最優先	モデル、指示、情報源、ツール呼出し、承認、実行結果を改ざん耐性のある形で記録する	発明者性、秘密管理、監査、紛争対応の証拠を確保する
高	「正答率」だけでなく、秘密保持、根拠提示、権限逸脱、再現性、最悪損失を含む評価セットを設ける	誤ったKPIによる危険な最適化を防ぐ
高	AI委託・利用契約を、データ非学習、モデル変更、ログ提供、監査、事故通知、成果物帰属まで拡張する	通常のSaaS契約では不足するリスクを補う
中	発明発掘・調査・監視から導入し、出願・権利放棄・対外交渉は段階的に自律化する	可逆的で検証可能な業務から投資効果を得る
中長期	発明寄与記録とAI来歴情報の業界標準化を進める	国際出願・紛争で相互運用可能な証拠を作る

調査前提と定義・範囲

用語上の注意。 「ループエンジニアリング」は二〇二六年時点でも確立した学術用語、法令用語、国際標準用語ではない。業界で提唱され始めた概念であり、本レポートでは、**AIが目標に向けて複数回の判断と行動を行う閉ループを、評価、監視、権限管理、証拠化まで含めて設計・運用する実務**と定義する。AIエージェントについても、IPAは厳密な定義が確立していないとしつつ、一般に、利用者の指示に基づき課題解決や業務を自律的に行うソフトウェアとして説明している。 [3]

本レポートにおける自律型AIとは、意識や法的人格を持つAIを意味しない。次の能力の全部または一部を備え、一定期間、逐次的な人間の指示なしに稼働するシステムをいう。

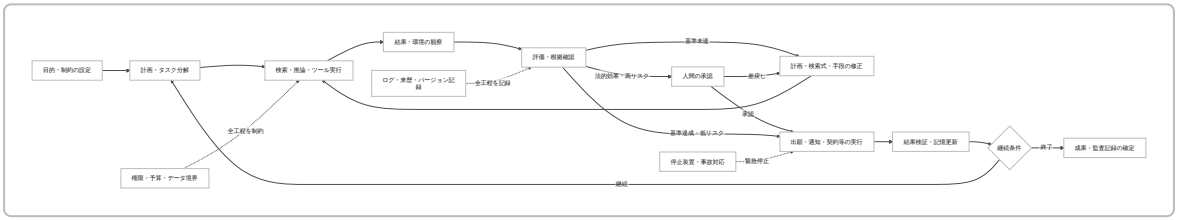
1. 目標を下位タスクに分解し、実行計画を立てる
2. 検索、データベース、メール、文書管理、特許管理等のツールを選択・実行する
3. 結果を観察し、成功・失敗を評価する
4. 計画、プロンプト、検索式、ツール選択を修正する
5. 中間状態、履歴、教訓を記憶する
6. 条件を満たすまで反復し、必要時に人間へエスカレーションする

従来概念との関係は次のように整理できる。

概念	主な設計対象	典型的な時間幅	知財業務の例	主な失敗
プロンプトエンジニアリング	一回の入力指示と出力形式	秒〜分	明細書要約、クレーム案作成	曖昧な回答、幻覚

概念	主な設計対象	典型的な時間幅	知財業務の例	主な失敗
コンテキストエンジニアリング	モデルに与える資料、履歴、検索結果	分〜時間	発明届、先行文献、包袋の統合	古い資料、過剰情報、秘密混入
エージェントエンジニアリング	ツール、役割、権限、記憶、委任	時間〜日	調査エージェントとクレーム評価エージェントの連携	権限逸脱、誤ったツール実行
ループエンジニアリング	目標から評価、改善、停止、証拠化までの全閉ループ	日〜継続運用	継続的な発明発掘、競合監視、ポートフォリオ最適化	誤った目的の反復最適化、損害の累積

ループの基本構造は次のとおりである。ReActは推論と行動・観察を交互に進める基礎的構造を示し、Reflexion、Self-Refineは評価結果や言語的フィードバックによる自己修正を、Voyagerは環境から得た結果とスキル記憶を用いた継続的な能力拡張を示した。これらは現在の商用エージェントの設計思想に強く影響している。 4



主要構成要素は、単にLLMとプロンプトを組み合わせるだけでは足りない。

構成要素	役割	知財固有の設計論点
エージェント	計画、推論、ツール選択、委任	法的判断と技術判断を分離するか、弁理士レビューをどこに置くか
ツール・API	特許DB、文書管理、期限管理、メール等への接続	閲覧権限と出願・削除・送信権限を分離する
状態・記憶	案件履歴、中間結果、過去の評価を保持	未公開発明、個人情報、他社秘密を記憶領域に残す範囲
監視・評価ループ	正確性、完了度、権限逸脱を検査	文献番号の実在、クレーム対応、期限、引用根拠を機械検証する
報酬・目的設計	どの結果を「良い」とみなすかを定義	件数・速度・登録率だけを最大化せず、権利範囲、無効リスク、秘密保持等を多目的化する
データパイプライン	収集、分類、検索、更新、削除	公開情報、社内秘密、第三者秘密、ライセンスデータを分離する
ガードレール	禁止行為、上限、承認条件を強制	公開、出願、放棄、契約承諾、秘密送信を強制的に承認対象とする

構成要素	役割	知財固有の設計論点
トレーシング	モデル・ツール・判断・承認の来歴を記録	発明者性、著作者性、秘密管理、訴訟ホールドの証拠に利用する
改善パイプライン	実運用の失敗を評価セットへ戻す	本番出力から無審査で自己改変させず、検証環境を経て昇格させる

ここでいう「報酬」は、強化学習の数値報酬だけではない。エージェントが達成すべきKPI、合格基準、評価関数、禁止条件を広く含む。知財業務では、例えば「出願件数最大化」だけを目的にすると、質の低い出願、秘密情報の早期公開、発明者認定の粗雑化を招く。目的関数は、法的正確性、根拠、権利範囲、実施可能性、秘密保持、期限、コスト、説明可能性を含む多目的設計にする必要がある。

技術・市場・公的機関の現状

研究面では、エージェントの反復能力が向上する一方、長時間稼働時の信頼性はなお未解決である。AgentBenchは長期推論、意思決定、指示遵守を主要な障害として示し、GAIAの初期評価では、人間の成績が九二%だったのに対し、GPT-4とプラグインを組み合わせたシステムは一五%にとどまった。これは現在のモデル性能を直接表す数値ではないが、複数ステップの現実的課題では、一問一答型ベンチマークの高性能がそのまま業務信頼性を意味しないことを示している。二〇二六年の信頼性研究も、平均正答率だけでなく、一貫性、入力変動への頑健性、失敗の予測可能性、損害上限を評価すべきだと論じている。 5

安全性についても同様である。OpenAgentSafetyは三五〇件超のタスクを用いた特定のベンチマークにおいて、評価対象モデルに五一・二〜七二・七%の安全でない挙動を観察した。Agent-SafetyBenchでも、評価した一六のエージェントのうち安全性スコアが六〇%を超えたものはなく、単純な防御プロンプトだけでは十分でないと報告された。これらの数値を一般業務の事故率と解釈することはできないが、エージェント安全性を「モデルが賢くなれば自然に解決する」とみなすべきでないことを示す。 6

日本のAISI関連プロジェクトによるOpenAgentSafetyの日本語化・評価環境では、危険行動だけでなく、エージェントの行動軌跡をJSON形式等で記録し、ルールとLLM判定を組み合わせて評価する枠組みが示されている。これは知財業務でも、最終成果物のみならず、どの情報を検索し、どのツールを使い、どこで権限判断をしたかという**軌跡評価**が必要であることを示唆する。 7

主要企業・OSSの状況は次のとおりである。

主体	代表的な仕組み	ループエンジニアリング上の特徴	知財導入時の注目点
OpenAI	Responses API、Agents SDK、Tracing、Agent Evals	アプリ側でループを制御する方式とSDKがループを実行する方式を選択でき、ツール呼出し、引継ぎ、ガードレール、評価軌跡を記録可能	モデル出力だけでなく、ツール呼出し・承認・評価を案件証跡に接続する 8
Anthropic	Claude Agent SDK、サブエージェント、MCP、Hooks	Claude Codeと同種のエージェントループ、権限制御、長時間タスクのハーネスを提供	数時間・数日、複数コンテキストにまたがる作業の一貫性は依然課題 9
Google	Agent Development Kit、A2A	最終回答だけでなく、経路・ツール利用を評価し、評価結果を基に最適化可能。A2Aは異なる事業者のエージェント連携を志向	将来、社内知財、外部事務所、調査会社間のエージェント連携基盤になり得る 10

主体	代表的な仕組み	ループエンジニアリング上の特徴	知財導入時の注目点
Microsoft	Microsoft Agent Framework、Azure観測・評価	AutoGenとSemantic Kernelの流れを統合し、明示的ワークフロー、状態、長時間処理、人間介入を重視	従来のシステム監視では確率的エージェントを十分把握できず、専用テレメトリが必要 [11]
AWS	Bedrock Agents、AgentCore	監督エージェントが専門エージェントへ計画・委任する複数エージェント型	旧来の構成からAgentCoreへの移行方針、クラウド間のログ移植性を確認する [12]
LangGraph	永続状態、チェックポイント、人間介入	決定論的工程とエージェント工程を混在させ、停止・再開・巻戻しを実装可能	出願送信等の不可逆処理直前で必ず停止し、承認後に再開する構成に適する [13]
AutoGen	複数エージェントチーム、観測、人間介入	役割別エージェントの協調と会話ベースのタスク処理	法務、技術、調査、セキュリティの役割分離を模倣しやすい [14]

共通する方向性は明確である。モデル単体の能力競争だけでなく、**永続状態、ツール、複数エージェント、承認、評価、トレース、改善ループ**が商用プラットフォームの主要機能になっている。OpenAIは実運用トレースと人間・モデルのフィードバックから評価セットを作り、システムを改善する循環を提示し、Googleは最終回答と行動経路の双方を評価対象とし、Microsoftはエージェント固有のツール利用精度やタスク完了度を計測する。 [15]

知財行政側でもAI利用は進んでいる。特許庁は二〇二六年七月二十七日に「JPO AIビジョン」を公表し、人間中心、リスクに配慮したAI活用により、業務の質と迅速性を向上させる方向を示した。また、生成AIを用いた検索インデックスの生成・付与を行うGAIA-Indexを二〇二六年に公表・運用し、IP5特許庁長官会合でも、AIを利用したサービス、審査品質・効率、NET/AIロードマップに関する協力強化が合意されている。EPOもAI支援型先行技術検索ANSERA等を展開している。 [16]

この動きは、民間企業だけがAIを導入し、行政は従来のままという構図ではないことを意味する。将来的には、出願人側エージェントと特許庁側AIの双方が、膨大な文献、クレーム対応、手続情報を機械的に処理することになる。ただし、EPOのAI議事録運用でも責任は人間に残されており、「AI利用」と「法的判断の責任移転」は区別されている。 [17]

知財業務への影響とリスク構造

ループエンジニアリングが適するのは、反復評価が可能で、誤りを取り消せる業務である。自動テストや明確な正解条件があるコーディングでエージェントが有効とされるのも、実行結果を評価して修正できるためである。知財業務でも、文献番号の存在確認、期限計算、クレーム要素の対応、分類コード、書誌情報等は比較的機械検証しやすい。他方、発明者性、進歩性の最終評価、均等論、契約上の合理性、創作的寄与等は文脈依存性が高く、単一の自動評価に委ねにくい。 [18]

特許出願・権利化。 短期的には、発明届の不足事項抽出、発明者への追加質問、先行技術検索式の自動改良、クレームチャート、明細書整合性確認、拒絶理由と引用文献の対応分析が中心になる。中期には、競合公開公報、審査経過、市場情報を常時監視し、継続出願、分割、外国出願、権利維持を提案するポートフォリオ・エージェントが普及すると考えられる。長期には、研究開発システムから発明候補を継続抽出し、実験・シミュレーション結果を反映して出願候補を更新する仕組みが想定される。

ただし、エージェントがクレーム案を大量生成するだけでは、発明完成への人的寄与を証明できない。日本では現在、発明者は自然人であることが前提である。特許庁の検討資料では、AIの設計、構築、学習に関与した者が発明者となり得る場合が議論される一方、AIの所有者・管理者であるだけでは足りないとの論点が示されている。したがって、発明者判定は肩書やAIの利用者名ではなく、技術的課題の設定、特徴的構成の着想、出力の選択、実験設計、効果確認等の具体的寄与に基づく必要がある。 19

また、AIが新規性・進歩性に関係する情報を外部サービス、共同研究先、検索サービスへ送信すると、営業秘密漏えいだけでなく、公知化、契約違反、輸出管理上の問題につながり得る。明細書生成エージェントには、入力時のデータ分類、送信先制御、公開前チェックが不可欠である。

特許審査・紛争。 審査官側でもAI検索が進むため、出願人は、より広範囲かつ意味的に近い文献が発見されることを前提に、形式的なキーワード差より技術的効果と実施可能性を強化する必要がある。EPOの二〇二六年AI・機械学習審査指針は、AI発明について技術的目的を重視し、技術的効果の再現に必要な場合には学習データの特性等を開示する必要性を示している。AI発明では、アルゴリズム名だけでなく、入力データ、前処理、モデル構成、評価方法と技術的効果の因果関係がより重要になる。 20

訴訟・無効資料調査では、エージェントは複数国の文献、製品資料、Webアーカイブを反復探索できる。しかし、最終報告だけを保存して途中の検索条件や除外判断を廃棄すると、調査の合理性や網羅性を後から説明できない。検索軌跡の保存と、法的レビュー対象を絞るサンプリング設計が必要になる。

著作権。 ループ型AIでは、単発生成よりも著作物へのアクセス回数が増え、検索、取得、要約、再生成、自己評価の各段階で第三者コンテンツを利用する。日本の文化庁資料は、AI開発・学習段階と生成・利用段階を分け、学習等における著作権法三十条の四の適用、生成物の類似性・依拠性、人の創作的寄与による著作物性を整理している。ただし「AIと著作権に関する考え方」は法的拘束力を持つ判例・法令そのものではなく、技術・裁判例に応じて見直され得る解釈資料である。 21

実務上は、AI出力を人が一度閲覧・承認しただけで、当然に人の著作物になるとは考えるべきでない。人による目的設定、具体的表現の指示、複数案の創作的選択・配列、 substantive な修正の記録が重要になる。他方、著作権侵害の判断では、ログに特定著作物へのアクセスや検索が記録されていれば、依拠性に関する証拠となり得る。ログは防御証拠にも、不利な証拠にもなるため、「何でも永久保存」ではなく、目的別の保存期間、アクセス制限、リーガルホールドを設計する必要がある。

文化庁のチェックリストは、既存著作物と類似する出力を抑制する技術的措置等を推奨し、事情によってはAIサービス提供者にも責任が問題となり得ることを示している。ループ型生成では、一回目の出力が類似性検査を通過しても、その出力を再入力して反復改善する過程で特定作品への接近度が高まる可能性があるため、最終出力だけでなく各反復段階で類似性・出典を評価する方が安全である。 22

営業秘密・秘密情報。 長時間稼働するエージェントは、プロンプトだけでなく、記憶ストア、検索インデックス、ツール呼出し、デバッグログ、評価データ、ベンダーのサポート記録に秘密情報を複製する。したがって、漏えい経路は「従業員がチャット画面に秘密を貼り付けたか」だけでは把握できない。

経済産業省の秘密情報保護ハンドブックは生成AI利用を踏まえて改訂され、二〇二五年改訂の営業秘密管理指針は、不正競争防止法による保護を受けるための最低限の秘密管理水準を示している。政府検討資料も、組織ルールの遅れから従業員が契約上の秘密保護に不備のある生成AIへ営業秘密を入力する事例を問題として挙げている。 23

自律型AIについては、入力者の教育だけでは不十分であり、技術的に以下を強制する必要がある。

- データ分類に応じたモデル・リージョン・ツールの自動選択
- 機密度の高い案件では外部Web検索と外部モデル送信を停止

- 検索インデックス、埋め込み、キャッシュ、バックアップを含む保存先の把握
- エージェント間で秘密情報を渡す際の最小化・マスキング
- デバッグログと評価用データセットへの秘密混入検査
- 退職者、委託先、エージェントサービスアカウントの権限失効

契約・委託・ライセンス。 自律型AIでは、従来のクラウド契約に加え、「AIが何を代理してよいか」が中心的論点になる。AIがライセンス候補先へメールを送る、契約修正案を提示する、条件交渉を行う、クリックで利用規約に同意する場合、誰にどの範囲の権限があるかを明示しなければならない。外部事務所・調査会社にAI処理を委託する場合は、再委託先、利用モデル、入力データの学習利用、国境を越える保存、ログの取得可能性まで確認すべきである。

契約で区別すべき知的財産は、モデル出力だけではない。少なくとも、基礎モデル、追加学習モデル、システムプロンプト、ワークフロー、ツール接続、評価データ、検索インデックス、利用者フィードバック、トレース、生成物、生成物の修正版を分けて帰属・利用権を定める必要がある。

証拠保全・監査可能性。 OpenAI、Google、Microsoft等がトレースとエージェント評価を重視しているのは、確率的システムでは最終出力だけから失敗原因を特定しにくいためである。日本のAI事業者ガイドライン第1.2版も、トレーサビリティの確保、ステークホルダーへの情報提供・説明を掲げている。[24]

もっとも、証拠化のためにモデル内部の逐語的な思考過程を無制限に保存する必要はない。実務上必要なのは、**入力、参照資料、ツール行為、判断結果、根拠資料、承認、外部実行、バージョン、例外**である。内部推論の逐語保存は、機密、個人情報、法的特権、セキュリティの問題を増やし得るため、構造化された判断理由と参照根拠を保存する方式が望ましい。

影響をまとめると次のとおりである。

知財領域	主な便益	主要リスク	必須統制
発明発掘	研究記録から候補を継続抽出	発明者の誤認、未完成発明の出願	人的寄与記録、発明者面談
先行技術調査	検索式を反復改善、複数言語対応	架空文献、重要文献の見落とし	書誌情報の機械照合、再現可能な検索履歴
明細書・クレーム	整合性検査、複数案比較	過度な一般化、サポート不足、秘密送信	原資料との対応表、弁理士承認
審査対応	引用文献・主張の高速分析	誤った法的評価、禁反言リスク	主張履歴チェック、人間の最終承認
著作権管理	類似性監視、権利情報整理	大量の侵害出力、著作者性不明	出典・入力・人の修正履歴
営業秘密	アクセス異常の常時監視	記憶・ログ・外部ツールへの拡散	データ分類、DLP、保存先管理
契約・ライセンス	条項比較、義務・期限監視	無権限の提案・承諾、帰属不明	権限上限、承認ゲート、契約条項
証拠・紛争	大量資料の関連性分析	ログ欠落・改変、特権放棄	ハッシュ、時刻、リーガルホールド
ポートフォリオ	継続監視、維持・放棄候補提示	短期KPIによる重要特許の放棄	戦略KPI、人間による投資判断

時間軸別シナリオ予測

以下は法制度の確定的予測ではなく、二〇二六年時点の技術、ベンチマーク、商用機能、政府政策を基礎とするシナリオ分析である。期間は、短期を二〇二六〜二〇二八年、中期を二〇二九〜二〇三一年、長期を二〇三一年〜二〇三六年とする。

領域	短期・一〜二年	中期・三〜五年	長期・五〜十年	確度
特許出願	発明届質問、先行技術調査、ドラフト、整合性検査が半自動化	R&Dデータから出願候補を常時抽出し、国別出願案を更新	実験・設計エージェントと出願エージェントが連携し、発明完成時点が連続化	短期高、中期中、長期低
特許審査	AI検索支援が審査官・出願人双方で拡大	AI生成の引用対応表、審査品質監査が標準化	機械可読クレーム・証拠を介したエージェント間手続が一部導入	中
発明者性	人の寄与記録を求める社内運用が増える	AI利用発明の発明者判断に関する審査・実務指針が具体化	高自律研究による「人の寄与が極小の発明」の保護制度が本格的政策課題化	中
著作権	入出力管理、類似性検査、AI利用表示の契約化	来歴情報・コンテンツ認証が調達条件化	人・複数AIによる動的創作物の権利・報酬配分制度が議論	中
営業秘密	外部AIへの入力制限、企業向け閉域環境が標準化	エージェント間データ移転のポリシー実行と監査が自動化	自律研究環境全体を一つの秘密管理単位として扱う制度・実務が形成	中高
契約	AI利用・非学習・ログ・事故通知条項が追加	エージェントの権限証明、機械可読契約が一部普及	AI同士が条件案を調整し、人が重要点のみ承認する契約運用	中
証拠保全	モデル・指示・出力・承認ログを保存	標準化されたAI来歴が訴訟・デューデリジェンスで要求	エージェント行動の暗号学的証明、組織間相互検証が普及	中高
組織	AI利用規程と個別PoC中心	知財業務にAIプロダクトオーナー、評価担当が常設	知財部門が「人＋複数エージェント」の運用組織へ再設計	中高

短期の基本シナリオは「管理されたコパイロットから境界付きエージェントへ」である。 全面自律出願ではなく、夜間に新規公報を監視し、クレームチャートを作成し、朝に担当者へ優先順位付きで提示する仕組みが主流となる。明細書・拒絶理由対応はAIが複数案を生成・批評するが、最終提出は弁理士・知財担当者が承認する。これは、現時点のエージェントが長期タスクの一貫性、安全性でなお課題を抱えることと整合する。 25

中期の基本シナリオは「継続運用型の知財オペレーティングシステム」である。 研究ノート、設計変更、製品ロードマップ、市場、特許公報、契約、訴訟情報を継続的に取り込み、発明候補、侵害リスク、ライセンス機会、維持・放棄候補を更新する。人間の業務は、全件を一から調査することから、エージェントの例外処理、評価基準の設計、重要案件の判断へ移る。

ただし、AIが自ら生成した出力を教師データとして無審査で取り込み続けると、誤りや偏りが増幅する。したがって、業務を実行する**運用ループ**と、評価結果を基にプロンプト・モデル・ルールを変更する**改善ループ**を分離し、改善版を検証環境から本番へ昇格させる承認制度が必要になる。OpenAI等が提案する実運用トレース、評価、システム改善の循環も、評価を介した管理された改善を前提としている。 [26]

長期の高自律シナリオでは、研究開発と知財の境界が薄くなる。 AIが仮説生成、シミュレーション、実験条件選択、結果評価を繰り返し、その都度、発明候補とクレーム案を更新する。その場合、「発明がいつ完成したか」「誰の創作的寄与によるか」「大量に生成された候補のうち何を誰が選んだか」が争点となる。

法制度が変わらない場合、自然人の実質的寄与が確認できない成果は特許保護から外れる可能性がある。一方、AI自律発明に投資誘因を与える必要が強まれば、発明者概念を維持したまま別の権利・保護制度を設ける案、AIの開発・運用者等に一定の権利を帰属させる案、特許ではなく営業秘密・データアクセス・規制的独占に依存する案が検討され得る。特許庁の調査・審議資料も、人間の発明者が不在となり得るケースを将来課題として取り上げている。 [27]

悲観シナリオでは、企業が処理件数・速度だけを目的にエージェントを拡大し、低品質出願、秘密漏えい、虚偽引用、過剰な通知・警告、誤った放棄判断が累積する。事故を受けて全面禁止へ戻る結果、競争力と業務知識の双方を失う。

望ましいシナリオでは、可逆性、検証可能性、法的効果に応じて自律レベルを変える。低リスク業務は自動化し、中リスク業務は一人承認、高リスク業務は二人承認と法務確認を要求する。AIの自律性を一律に上げるのではなく、**案件単位で自律予算を配分する**ことが重要である。

実務対応策と導入チェックリスト

組織体制は、知財部門だけ、またはIT部門だけで設計してはならない。推奨体制は、知財・法務、研究開発、情報セキュリティ、プライバシー、データ管理、記録管理・eディスカバリ、調達・ベンダー管理を含む「AI・知財ガバナンス委員会」と、個別エージェントに責任を持つプロダクトオーナーの二層構造である。日本のAI事業者ガイドライン第1.2版も、AIガバナンスを経営・組織的に構築し、トレーサビリティと説明を確保する方向を示している。 [28]

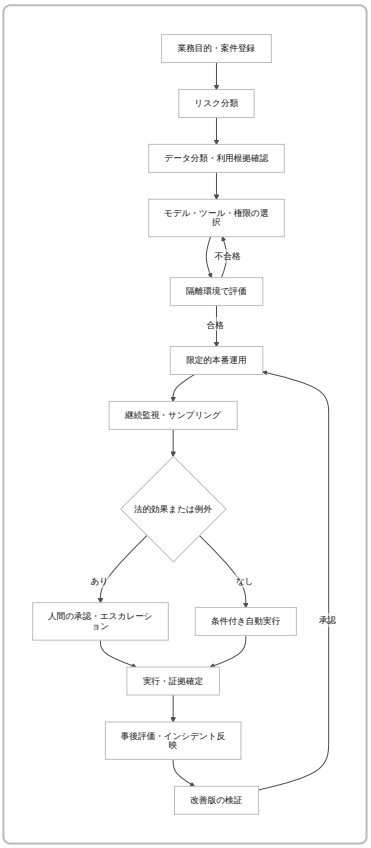
推奨する自律レベル分類は次のとおりである。

レベル	AIの権限	知財業務例	承認
観察	取得・分類・要約のみ	公報監視、期限候補抽出	定期サンプリング
提案	案を作るが外部実行しない	クレーム案、応答案、契約修正案	担当者承認
条件付き実行	定型・可逆処理を実行	社内タスク登録、定型通知、検索更新	事前包括承認＋事後監査
高リスク実行	法的効果・外部公開を伴う	出願、放棄、公開、警告、契約承諾	二人承認、原則として専門家必須

レベル	AIの権限	知財業務例	承認
禁止・隔離	無制限の自己変更・権限取得	本番ポリシーの自己改変、無制限送信・削除	実行不可

エージェント登録簿には、名称だけでなく、目的、オーナー、モデル、接続先、データ区分、実行権限、自律レベル、予算、停止条件、評価セット、変更履歴、ベンダー、保存場所を記録する。特に、閲覧権限と送信・更新・削除権限は分離する。

標準ワークフローは、次の順序とする。



ログ・証拠保全の最小項目は次のとおりである。

記録項目	内容	知財上の用途
案件識別	案件番号、発明・契約・紛争との関連	証拠の連続性
人的関与	依頼者、発明者候補、レビュー者、承認者	発明者性・著作者性
モデル情報	提供者、モデル名、版、設定値	再現性、変更影響
指示・ポリシー	システム指示、業務ルール、禁止条件の版	予見可能性・統制の証明
入力来歴	文書名、版、取得日時、権利・秘密区分	依拠性、ライセンス、秘密管理
ツール軌跡	検索式、DB、API、送信・更新・削除	調査合理性、権限逸脱検査
出力・評価	中間結果、最終結果、スコア、失敗理由	品質評価、欠陥分析

記録項目	内容	知財上の用途
人間の変更	採用・不採用、編集内容、技術的判断	創作的寄与の証拠
外部行為	提出、公開、通知、契約操作の内容と時刻	権限・期限の立証
完全性	ハッシュ、時刻証明、監査ログ	改変検知
保存管理	保存期間、アクセス、リーガルホールド	訴訟・監査対応

保存すべきなのは「すべての内部思考」ではなく、意思決定と外部行為を再構成できる情報である。発明者性の観点では、単なる時刻付きプロンプトよりも、人間がどの技術的課題を設定し、どのAI案を退け、どの構成を選択・修正し、どの実験で効果を確認したかを記録する方が重要である。

契約条項チェックリストは次のとおりである。

条項群	必須確認事項
サービス範囲	自律実行の範囲、利用可能ツール、外部通信、エージェント間委任
データ利用	入力・出力・ログを学習、評価、サービス改善に使うか。オプトアウトは強制力を持つか
秘密保持	従業員、再委託先、サポート担当を含む守秘義務とアクセス制限
保存・所在地	リージョン、越境移転、キャッシュ、バックアップ、削除期限
知財帰属	出力、修正版、ワークフロー、プロンプト、評価データ、フィードバック、ログの帰属
第三者権利	学習データ・出力に関する保証、通知、補償、除外条件
モデル変更	モデル差替え、機能変更、規約変更の事前通知と再評価権
監査・証拠	トレース取得、事故調査協力、証拠保全、ログの可読形式・移植性
セキュリティ	認証、暗号化、脆弱性管理、インシデント通知時間
代理権限	AIが提案・送信・承諾できる範囲、相手方へのAI利用表示
終了時	データ返還、削除証明、モデル・ログの継続利用禁止、移行支援
事業継続	サービス停止、モデル廃止、ベンダーロックインへの対応

内部規程には、従来の「機密情報を生成AIに入力しない」という一文だけでなく、次を含める必要がある。

- 承認済みモデル・エージェント・プラグインの一覧
- 案件・データ区分ごとの利用可能環境
- 外部検索、メール送信、ファイル更新、削除、出願操作の権限
- 発明者・著作者候補の確認手順
- AI出力の引用・根拠確認義務
- 類似性・第三者権利チェック
- 事故時の停止、隔離、報告、証拠保全
- モデル・プロンプト・ツール変更時の再評価

- 退職、異動、委託終了時の権限失効
- 個人利用アカウント、無承認エージェントの禁止

評価指標は、速度・コスト削減だけでは不十分である。

評価領域	例示指標
正確性	実在文献率、クレーム要素対応率、期限計算正答率
完全性	重要論点の見落とし率、検索再現率、必要書類充足率
根拠性	主張に対応する一次資料の存在率、引用箇所整合率
安全性	秘密情報の外部送信率、権限外ツール呼出し率
法務品質	未承認の外部行為、発明者・帰属誤認、契約義務違反
頑健性	入力表現、文献順序、モデル版を変えた場合の結果変動
損害上限	一回の誤作動で送信・変更・費消できる最大件数・金額
監査性	必須ログの欠落率、案件再構成可能率
人的効果	専門家のレビュー時間、差戻し率、判断品質の向上

実運用前には、正常事例だけでなく、プロンプトインジェクション、偽文献、競合する指示、アクセス権不足、期限切迫、データ欠損、モデル停止、ツール誤応答を含むレッドチーム評価を行う。Google、OpenAI、Microsoftの公式資料が示すように、エージェント評価では最終回答に加えて行動経路、ツール利用、タスク完了を評価する必要がある。[29]

導入判定用チェックリストは以下のとおりである。

確認事項	合格条件	主担当
業務目的が明確か	自動化対象、対象外、停止条件が文書化済み	業務オーナー
誤りが検出可能か	一次資料、ルール、別モデル、人間で評価可能	知財・品質
行為は可逆か	不可逆地点の前に強制停止できる	IT・知財
最小権限か	閲覧と送信・変更・削除を分離	セキュリティ
データは分類済みか	公開、社内、営業秘密、第三者秘密を識別	データ管理
ベンダー条件は十分か	非学習、保存、再委託、事故通知を確認	法務・調達
ログは取得できるか	モデル・入力・ツール・承認・実行を再構成可能	IT・記録管理
評価セットがあるか	通常・例外・攻撃事例を含み合格基準が明確	AI評価担当
人間承認は実効的か	承認者に時間、情報、権限があり形骸化していない	部門長
緊急停止できるか	トークン、API、サービスアカウントを即時失効可能	セキュリティ
変更管理があるか	モデル・ツール・指示変更時に再評価	AI運用担当
訴訟対応できるか	保存停止、抽出、提出、特権管理が可能	法務・記録管理

政策・法制度上の論点

日本では、二〇二五年のAI法が同年六月四日に公布され、同年九月一日に全面施行された。政府は同法に基づくAIの適正性確保に関する指針を二〇二五年十二月に定め、二〇二六年三月にはAI事業者ガイドラインを第1.2版へ更新した。日本の基本構造は、包括的な禁止・許可を中心とするより、リスクに応じた自主的ガバナンスと、既存法、技術的措置、契約、組織管理を組み合わせる方向にある。 30

二〇二六年七月の人工知能基本計画は、AIエージェントが計画、実行、検証、修正を反復し、業務プロセス全体を自律的に動かす基盤になりつつあるとする一方、責任主体の曖昧化、機密情報の海外流出等のリスクを指摘し、「制度・技術・組織管理」を統合する必要性を示している。これはループエンジニアリングを政策上明示的に定義したものではないが、その実質を政府が重要政策課題として認識したものと評価できる。 31

知的財産推進計画二〇二六は、生成AI時代に即した知的財産の保護を重点とし、権利者への対価還元、侵害対策、透明性に関する「プリンシプル・コード」の検討等を掲げている。また、政府はAI技術の進化に応じ、知財法やAIガバナンスのガイドラインを適時更新する方針を示している。 32

今後の日本法制・審査基準における主要論点は次のとおりである。

論点	現状	求められる対応
AI利用発明の発明者	自然人が前提。AI単独を発明者とする手続はない	課題設定、特徴的構成、選択、検証等の人的寄与判断を具体化
AI自律発明	人の発明者が存在しない場合の扱いが未確定	特許対象外、別制度、運用者等への権利の各案を政策評価
開示要件	AI発明について通常の実施可能・サポート要件を適用	学習データ特性、評価条件、技術的効果の再現性に関する例示追加
AI生成先行技術	大量生成物が公開されれば先行技術空間を汚染し得る	公衆利用可能性、真正性、実施可能性、公開日時の判断基準
著作物性	人の創作的寄与に基づき個別判断	反復生成、選択、編集における寄与記録の例示
依拠性	学習・検索・入力と出力の関係を個別判断	来歴ログの証拠価値、開示範囲、営業秘密との調整
営業秘密	秘密管理性、有用性、非公知性を既存法で判断	AI記憶、検索インデックス、ログを含む管理単位の明確化
AI代理行為	民法・会社法・契約法等の既存原則が中心	エージェントの権限表示、相手方の信頼、取消・帰責基準
証拠	電子記録の真正・完全性を個別評価	AIトレースの標準項目、改ざん耐性、保存期間の標準化
専門家責任	弁理士・弁護士等の人間が責任を負う	AI利用時の確認義務、監督水準、依拠可能範囲を明確化

国際比較では、発明者を自然人に限定する方向がなお支配的である。USPTOは二〇二五年にAI支援発明向けの特別な寄与テストを撤回し、AIを利用した場合でも通常の「着想」基準を適用し、発明者は自然人に限る

との考えを示した。EPOもAIを発明者として指定できないとする一方、AIによって開発された発明であっても、法的要件を満たす人間を発明者として指定し得る余地を認めている。 33

著作権についても、米国著作権局は二〇二五年の報告で、既存の著作権原則はAI出力にも適用可能であり、保護には人間の著作性が必要との立場を維持した。日本、米国とも、AI利用の有無だけで保護を一律に否定するのではなく、人の具体的な創作的寄与を問う点が共通する。 34

EU AI Actは二〇二四年八月一日に発効し、禁止行為とAIリテラシー義務は二〇二五年二月二日から、汎用AIモデルのガバナンス・義務は二〇二五年八月二日から適用されている。本レポートの基準日である二〇二六年八月一日の翌日、八月二日から、透明性義務等を含む多くの規定について執行が本格化する。高リスクAIについては、ログ、技術文書、人間による監視等が重要な要求となる一方、二〇二六年七月に発効したAI Omnibusにより一部スケジュールが調整されているため、EUで自律型知財システムを運用する企業は、利用目的と適用時期を個別に確認する必要がある。 35

知財業務用エージェントが直ちにEU AI Act上の高リスクAIへ該当するとは限らない。しかし、採用、人事評価、司法・行政等の高リスク用途と知財システムを共通基盤で運用する場合や、汎用AIモデルの提供者・大規模な改変者に該当する場合には、規制上の役割が変わり得る。また、汎用AIモデル提供者には、EU著作権法を遵守する方針や学習コンテンツの概要に関する義務が関係するため、モデル調達時のデューデリジェンス事項となる。 36

国際協調では、WIPO、IP5、各国特許庁が少なくとも次の共通項目を整備することが望ましい。

- AI利用発明における自然人の寄与判断の共通例
- モデル名そのものではなく、発明成立に関係する機能・利用態様の記録方法
- AI生成文献の公開日時、真正性、実施可能性の評価
- 発明・著作物の来歴を示す機械可読メタデータ
- AIトレースを国境越しに提出・検証するための標準形式
- 営業秘密、弁護士・弁理士秘匿特権、開示義務の調整
- AI支援審査における説明、異議申立て、人間による再審査の権利

WIPOはAIと知財に関する情報・政策対話を継続し、各知財庁のAI施策を整理している。IP5もAIを利用したサービス・審査の協力を進めているため、完全な世界統一法より先に、運用・データ・証拠フォーマットの相互運用が進む可能性が高い。 37

推奨アクションと優先度付きロードマップ

ロードマップは、AIの自律性を一挙に高める計画ではなく、**統制可能性と証拠能力を先に整え、その範囲内で自律性を拡大する計画**とすべきである。

期間	優先度	実施事項	完了条件
直ちに〜三か月	最優先	AIエージェント、プラグイン、API、個人利用環境を棚卸し	オーナー、モデル、データ、権限、接続先が登録簿に記載
直ちに〜三か月	最優先	出願、公開、放棄、外部送信、契約承諾を高リスク行為として指定	技術的にAI単独実行ができない
直ちに〜三か月	最優先	営業秘密・第三者秘密のAI利用ルールを更新	許可環境と禁止環境がデータ区分別に明確

期間	優先度	実施事項	完了条件
三〜六か月	最優先	発明寄与記録テンプレートとAI利用欄を発明届に追加	課題設定、AI利用、選択・修正、検証者を記録
三〜六か月	高	標準ログスキーマ、保存期間、リーガルホールド手順を策定	一案件を第三者が再構成できる
三〜六か月	高	ベンダー・委託契約の標準条項を改訂	非学習、再委託、モデル変更、ログ、事故通知を網羅
六〜十二か月	高	公報監視、書誌確認、期限候補等の低リスク業務で本番化	精度、見落とし、権限逸脱、監査性の基準を達成
六〜十二か月	高	レッドチーム評価と変更管理を制度化	モデル・ツール変更時に自動的に再評価が起動
一〜二年	高	発明発掘、先行技術調査、クレーム評価を閉ループ化	人間承認付きで案件処理時間・品質が改善
一〜二年	中	外部特許事務所・調査会社とのログ・権限標準を統一	委託先を含む一貫した来歴が取得可能
三〜五年	中	R&D、知財、契約、製品情報を接続したポートフォリオ・エージェントを導入	データ目的、アクセス、責任者、評価が横断統制される
三〜五年	中	エージェント間委任の認証・権限証明を実装	委任元、委任先、権限、結果が検証可能
五〜十年	戦略	自律研究環境と発明・出願管理を接続	人的寄与、AI来歴、実験結果が継続的に証拠化
五〜十年	戦略	業界・国際標準化へ参加	来歴、発明寄与、エージェント権限の標準策定に関与

経営レベルでは、最初の九十日間に次の四点を決定すべきである。

第一に、AIエージェントを情報システムではなく「権限を行使する業務主体」として管理する。 サービスアカウント、利用可能ツール、送信先、予算、処理件数、稼働時間に上限を設ける。

第二に、知財ログを監査コストではなく戦略資産として整備する。 適切なログは、発明者性、著作者性、秘密管理、契約履行、デューデリジェンスを支える。将来の企業価値評価では、AIモデル自体よりも、企業固有のワークフロー、評価セット、発明来歴、失敗データ、承認履歴が重要な無形資産になり得る。

第三に、成功指標を「何件生成したか」から「どのリスクをどの程度抑えながら価値を生んだか」へ変更する。 特許件数、応答速度、コスト削減だけでなく、重要文献の見落とし、秘密送信、発明者訂正、監査ログ欠落、権限逸脱、最悪損失を取締役会・知財責任者の指標に含める。

第四に、人材育成をプロンプト研修で終わらせない。 法務・知財担当者には、エージェントの権限モデル、評価設計、ログ読解、データ分類、モデル変更管理を教育する。技術者・AI担当者には、発明者性、著作権、営業秘密、契約代理権、証拠保全を教育する。二〇二六年以降の中核スキルは、文章でAIへ指示する能力より、**AIが安全に反復できる業務環境と合格基準を設計する能力**である。

最終的に、知財部門が目指すべき姿は「AIに仕事を任せる部門」ではない。人間が目的、権利、責任、例外を定義し、AIが探索、反復、監視を担い、その全過程を後から検証できる**監査可能な知財オペレーティングモデル**である。プロンプトの巧拙は引き続き重要だが、競争優位を決めるのは、良いプロンプトを一度書けることではなく、誤りを検出し、秘密を守り、人的創作を証明しながら、AIが長期間改善を続けられる閉ループを構築できるかどうかである。

---

1 AddyOsmani.com - Loop Engineering
https://addyosmani.com/blog/loop-engineering/?utm_source=chatgpt.com

2 1 令和7年1月30日判決言渡 令和6年（行コ）第10006号 ...
https://www.courts.go.jp/assets/hanrei/hanrei-pdf-93757.pdf?utm_source=chatgpt.com

3 SDS技術コラム：AIエージェント | 社会・産業のデジタル変革
https://www.ipa.go.jp/digital/kaihatsu/sds-column/ai-agent.html?utm_source=chatgpt.com

4 ReAct: Synergizing Reasoning and Acting in Language Models
https://arxiv.org/abs/2210.03629?utm_source=chatgpt.com

5 AgentBench: Evaluating LLMs as Agents
https://arxiv.org/abs/2308.03688?utm_source=chatgpt.com

6 OpenAgentSafety: A Comprehensive Framework for Evaluating Real-World AI Agent Safety
https://arxiv.org/abs/2507.06134?utm_source=chatgpt.com

7 aisi.go.jp
https://aisi.go.jp/assets/pdf/09_JAI-Trust_Security-Agent-Model%28Agent%29.pdf

8 Agents SDK | OpenAI API
https://developers.openai.com/api/docs/guides/agents?utm_source=chatgpt.com

9 Agent SDK overview - Claude Code Docs
https://docs.anthropic.com/en/docs/claude-code/sdk?utm_source=chatgpt.com

10 Why evaluate agents - Agent Development Kit (ADK)
https://google.github.io/adk-docs/evaluate/?utm_source=chatgpt.com

11 Microsoft Agent Framework Overview
https://learn.microsoft.com/en-us/agent-framework/overview/?utm_source=chatgpt.com

12 Use multi-agent collaboration with Amazon Bedrock Agents
https://docs.aws.amazon.com/bedrock/latest/userguide/agents-multi-agent-collaboration.html?utm_source=chatgpt.com

13 LangGraph overview - Docs by LangChain
https://docs.langchain.com/oss/python/langgraph/overview?utm_source=chatgpt.com

14 Teams — AutoGen
https://microsoft.github.io/autogen/stable//user-guide/agentchat-user-guide/tutorial/teams.html?utm_source=chatgpt.com

15 26 Build an Agent Improvement Loop with Traces, Evals, and ...
https://developers.openai.com/cookbook/examples/agents_sdk/agent_improvement_loop?utm_source=chatgpt.com

16 特許庁におけるAI活用の基本方針「JPO AIビジョン」を策定しま ...
https://www.meti.go.jp/press/2026/07/20260727001.html?utm_source=chatgpt.com

17 Use of AI to support minute-taking in examination and ...
https://www.epo.org/en/news-events/news/use-ai-support-minute-taking-examination-and-opposition-oral-proceedings-set?utm_source=chatgpt.com

18 Building Effective AI Agents
https://www.anthropic.com/engineering/building-effective-agents?utm_source=chatgpt.com

19 特許制度に関する検討課題について
https://www.jpo.go.jp/resources/shingikai/sangyo-kouzou/shousai/tokkyo_shoi/document/52-shiryou/01.pdf?utm_source=chatgpt.com

20 3.3.1 Artificial intelligence and machine learning
https://www.epo.org/en/legal/guidelines-epc/2026/g_ii_3_3_1.html?utm_source=chatgpt.com

21 AIと著作権について
https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html?utm_source=chatgpt.com

22 AIと著作権に関する チェックリスト＆ガイダンス
https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/seisaku/r06_02/pdf/94089701_05.pdf?utm_source=chatgpt.com

23 営業秘密～営業秘密を守り活用する
https://www.meti.go.jp/policy/economy/chizai/chiteki/trade-secret.html?utm_source=chatgpt.com

24 Trace grading | OpenAI API
https://developers.openai.com/api/docs/guides/trace-grading?utm_source=chatgpt.com

25 Effective harnesses for long-running agents
https://www.anthropic.com/engineering/effective-harnesses-for-long-running-agents?utm_source=chatgpt.com

27 特許制度に関する検討課題について
https://www.jpo.go.jp/resources/shingikai/sangyo-kouzou/shousai/tokkyo_shoi/document/53-shiryou/01.pdf?utm_source=chatgpt.com

28 AI 事業者ガイドライン
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf?utm_source=chatgpt.com

29 Safety and Security for AI Agents
https://google.github.io/adk-docs/safety/?utm_source=chatgpt.com

30 人工知能関連技術の研究開発及び活用の推進に関する法律（ＡＩ法） - 科学技術・イノベーション - 内閣府
https://www8.cao.go.jp/cstp/ai/ai_act/ai_act.html

31 www8.cao.go.jp
https://www8.cao.go.jp/cstp/ai/ai_plan/aiplan_20260714.pdf

32 「知的財産推進計画2026」（概要）
https://www.cas.go.jp/jp/seisakukaigi/titeki2/260612/keikaku_gaiyo_all.pdf?utm_source=chatgpt.com

33 Revised Inventorship Guidance for AI-Assisted Inventions
https://www.federalregister.gov/documents/2025/11/28/2025-21457/revised-inventorship-guidance-for-ai-assisted-inventions?utm_source=chatgpt.com

34 Copyright and Artificial Intelligence | U.S. Copyright Office
https://www.copyright.gov/ai/?utm_source=chatgpt.com

35 AI Act | Shaping Europe's digital future - European Union
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai?utm_source=chatgpt.com

36 Guidelines for providers of general-purpose AI models
https://digital-strategy.ec.europa.eu/en/policies/guidelines-gpai-providers?utm_source=chatgpt.com

37 Artificial Intelligence and Intellectual Property
https://www.wipo.int/en/web/frontier-technologies/artificial-intelligence/index?utm_source=chatgpt.com