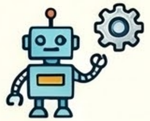
知財部門における「野良AIエージェント」防衛ガイド：リスク管理と7つの統制レイヤー

野良AIエージェントが引き起こす知財特有のリスク

RPA（野良ロボ）との違い



RPA：
定型処理の誤作動

AIエージェント：
「推論・判断」による幻覚（嘘の判断）、
権限逸脱、機密情報の外部送信

⚠ 知財業務への致命的な影響

未公開発明の漏えい
による新規性喪失

誤ったクレーム案による
権利範囲の不備

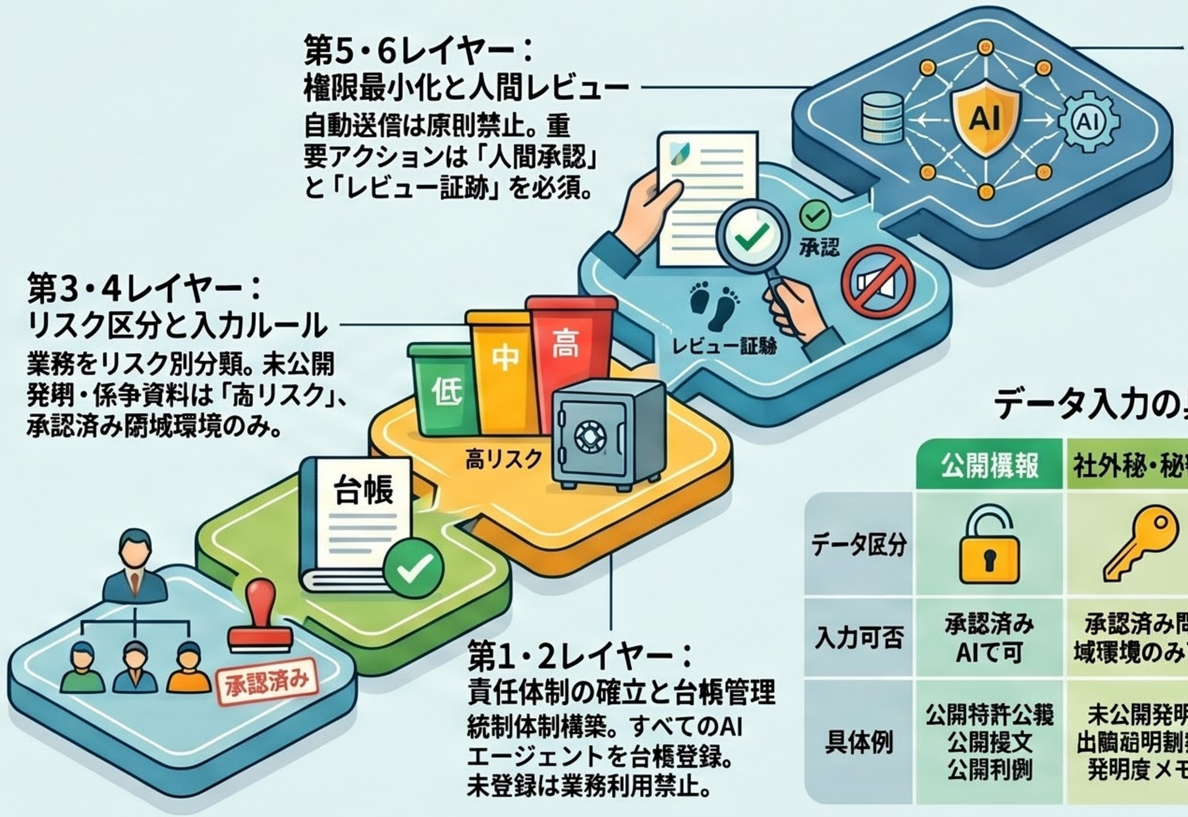
外国サーバへの技術提供
による輸出管理法違反

知財を守る「7つの統制レイヤー」

第5・6レイヤー：
権限最小化と人間レビュー
自動送信は原則禁止。重要アクションは「人間承認」と「レビュー証跡」を必須。

第7レイヤー：
監視・監査・廃止の制度化
定期的な棚卸し。不要なエージェントや権限を確実に削除。

第3・4レイヤー：
リスク区分と入力ルール
業務をリスク別分類。未公開発明・係争資料は「高リスク」、承認済み隔離環境のみ。



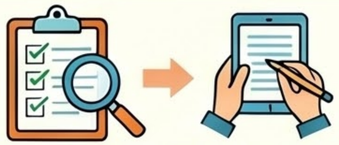
第1・2レイヤー：
責任体制の確立と台帳管理
統制体制構築。すべてのAIエージェントを台帳登録。未登録は業務利用禁止。

データ入力の具体的な可否判断

	公開機密	社外秘・秘密	高度秘密	入力禁止
データ区分				
入力可否	承認済みAIで可	承認済み隔離環境のみ可	個別承認が必要	原則不可
具体例	公開特許公報 公開提文 公開判例	未公開発明 出願前明細書 発明提メモ	係争資料 侵害鑑定 M&A関連 ライセンス交渉	第三者NDAでAI利用禁止の情報 個人情報

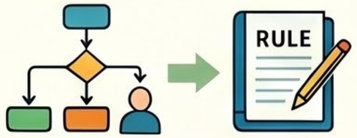
導入ロードマップ（すぐに始めるべきアクション）

0〜30日：現状の棚卸し



既存のスクリプトやRAG、API連携を洗い出し、警告的な台帳作成。高リスクな外部AI利用を停止。

31〜90日：統制ルールの導入



入力ルールやレビュー必須義務を定義。高リスク業務での人間承認をワークフローに組み込み。

3ヶ月以降：継続的改善

台帳管理を正式化、契約にAI利用条項を追加。国際規格を参考に品質管理体制へ統合。