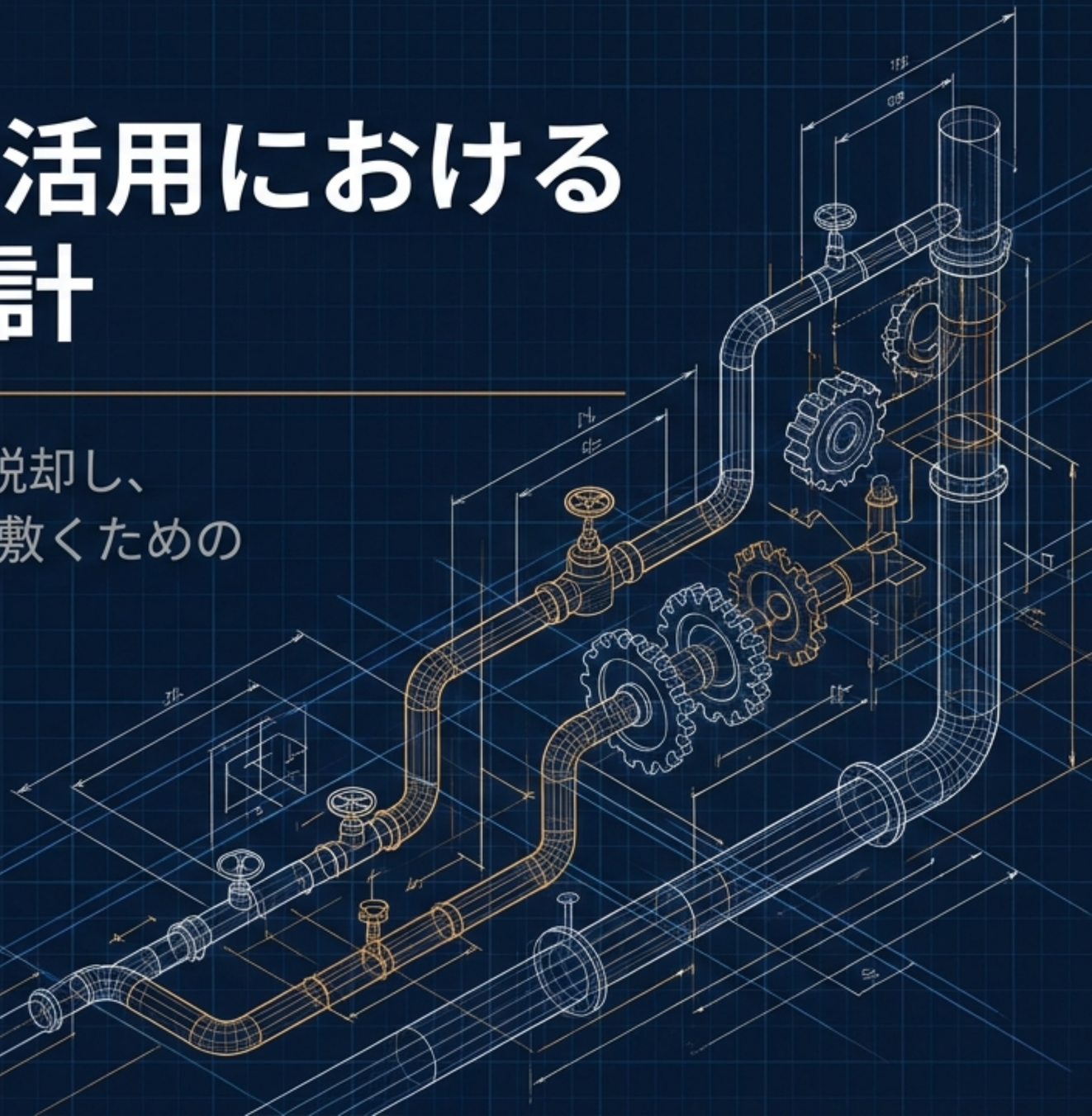
生成AIの知財業務活用における人間関与モード設計

「使うか・使わないか」の二元論から脱却し、
法的リスクと業務特性に応じた統制を敷くための
実践的アーキテクチャ



二元論からの脱却：知財AIは「業務ごとのモード切り分け」がすべてである

知財業務の厳格な要件

- 特許クレーム生成や審査対応において、LLMは有用だが専門家の検証が不可欠。
- 一貫性・法的堅牢性の確保に「人間の責任」は代替できない。

3つの関与モード

- HITL（事前承認必須）：権利範囲・外部提出など非可逆な工程
- HOTL（例外監督中心）：大量処理・例外監督が可能な工程
- HOOTL（事後監査中心）：定型・低リスク処理

検証可能な自動化

- 「どこまで自動化できるか」ではなく、「どこまで検証可能な形で自動化するか」が真の設計課題。
- あとから根拠付きで説明・検証できる統制基盤の構築。

一般論より厳しい、知財実務特有の「5つの防衛ライン」

著作権と学習データ汚染

文化庁ガイドラインに基づく、学習データや生成物に関するリスク低減要件

未公開発明と営業秘密

不用意な入力による漏えいリスク（経産省·秘密情報保護ハンドブック準拠）

将来の権利範囲への直結

単なる情報処理ではなく、特許クレームや紛争ポジションを左右する重み

個人情報・契約責任

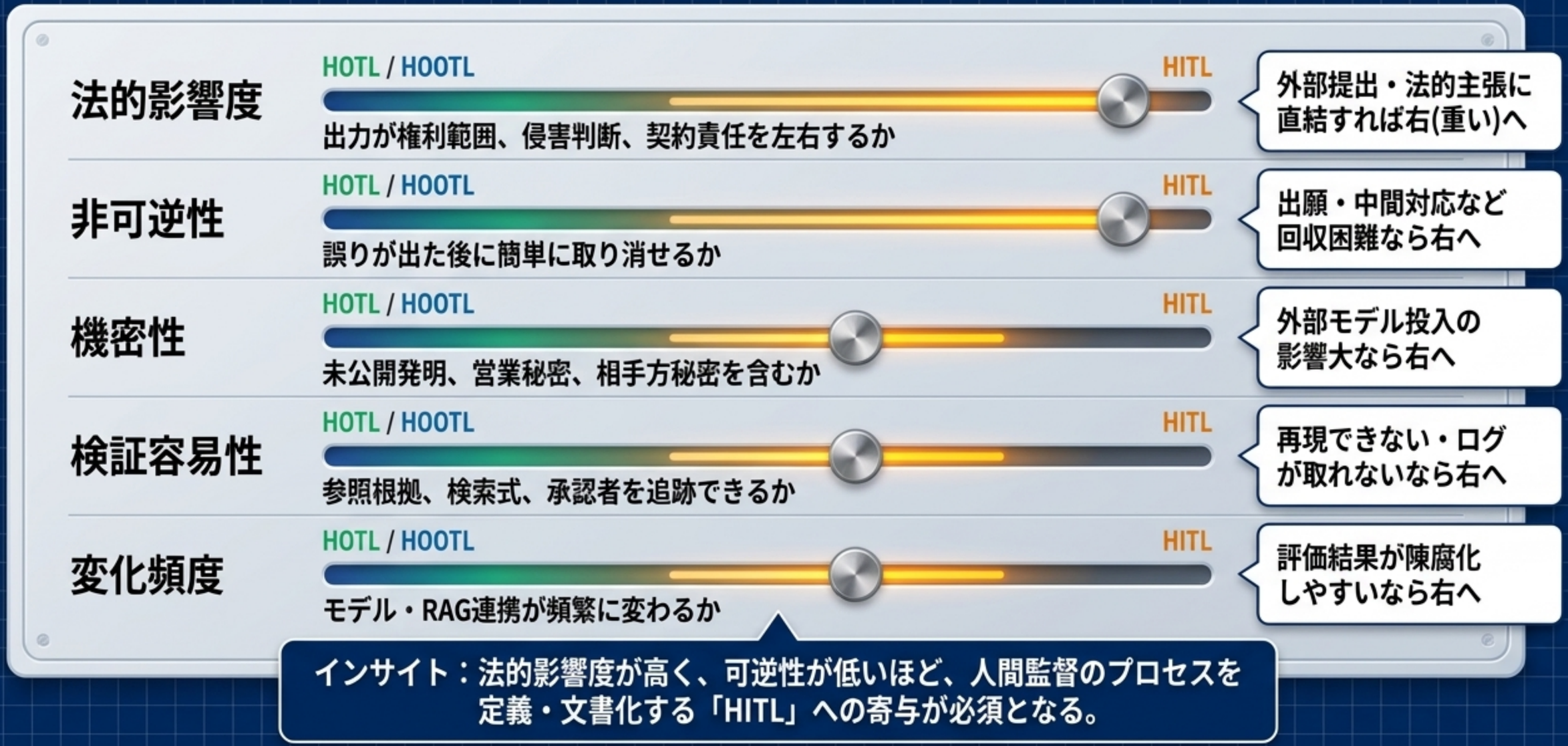
第三者連携やAPI経由での越境移転・契約逸脱リスク

事後的な説明責任

AI事業者ガイドラインが求める「トレーサビリティ」の厳密な実証

導入判断は「性能が良いか」より先に、「機密情報をどの境界で処理するか」から始まる。

モード設計を決定づける「5つの判断パラメーター」



インサイト：法的影響度が高く、可逆性が低いほど、人間監督のプロセスを定義・文書化する「HITL」への寄与が必須となる。

コア・フレームワーク：知財AIにおける3つの運用モード

HITL（事前承認必須）

- 定義：人間が案件ごとの重要出力を確認・承認しないと外部利用に進めない運用。
- 技術的特徴：公式ソース/RAG必須、差分表示、権限分離。
- 主な用途：クレーム案、審査対応主張、個別ライセンス条項。

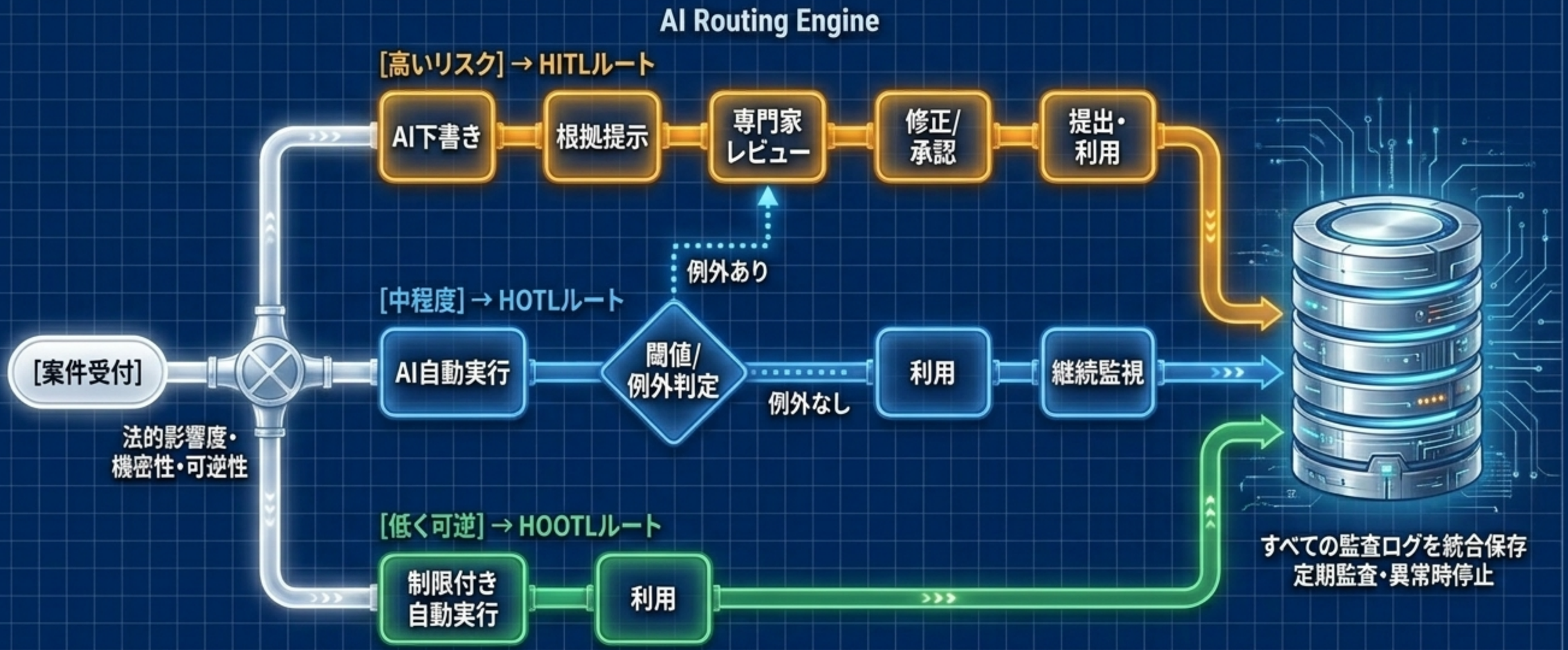
HOTL（例外監督中心）

- 定義：AIが通常処理を進めるが、閾値超過やサンプリング結果に応じて人間が介入する運用。
- 技術的特徴：信頼度閾値、ポリシー判定、例外キュー。
- 主な用途：先行技術調査の一次整理、契約レビュー一次抽出。

HOOTL（事後監査中心）

- 定義：人間は意思決定ループに入らず、事後的な監査と停止権限のみを持つ運用。
- 技術的特徴：サンドボックス、改ざん困難ログ、キルスイッチ。
- 主な用途：メタデータ付与、期限通知、DLP・保持制御。

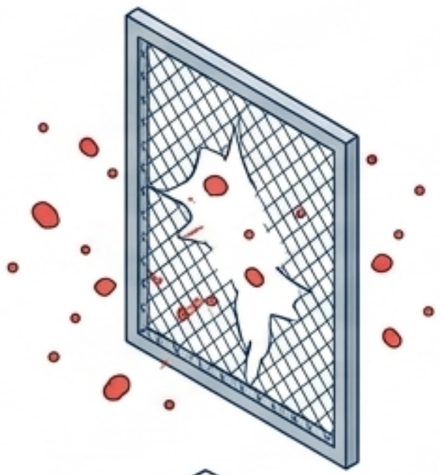
知財案件のAI運用ルーティング・フロー



【警告】機能不全に陥る「名ばかりのHITL」の罠

「人が最後に見る」だけでは統制にならない。自動化バイアスにより、人間はAIの推薦を無批判に追認してしまう。

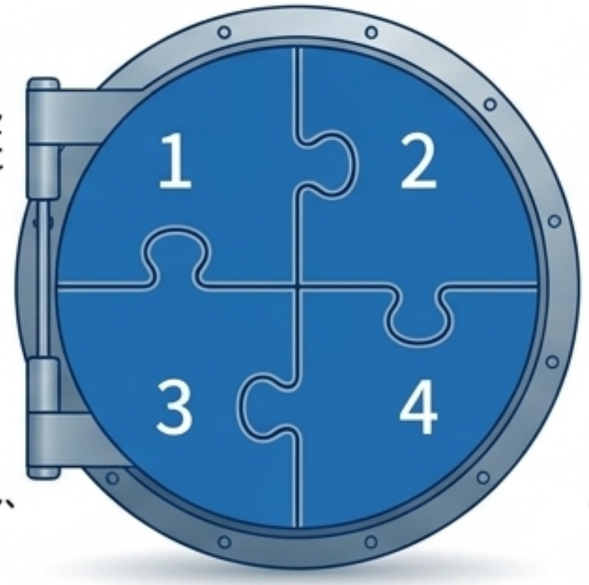
Bad Example (Nominal HITL)



Approve

例：クレーム案レビューに1件3分しか割けない体制は、名前がHITLでも実質はHOTL未満。

The 4 Requirements for True HITL



情報
(Information)
引用元や思考プロセスが画面上で即座に確認できること

時間
(Time)
批判的検証を行うための十分なリソースとスケジュール

権限
(Authority)
AIの出力を差し戻し、プロセスを止める明確な権限

教育
(Education)
AI特有の「もっともらしい嘘（ハルシネーション）」を見抜くリテラシー

業務プロセス別推奨モード (1) 特許出願・調査・審査対応

	HITL（事前承認）	HOTL（例外監督）	HOOTL（事後監査）
特許出願	クレーム・明細書案作成 （主張・記載要件の直接管理が必要）	背景技術・図面説明の一次生成 （下書き速度向上）	書式整形・分類候補 （事務負荷軽減）
	Default: クレーム・対外提出はHITL、周辺はHOTL		
先行技術調査	検索戦略・最終引用文献決定	検索式拡張・クラスタリング・一次要約	重複排除・アラート配信
	Default: HOTL中心で処理し、最終サーチメモはHITL		
特許審査対応	差異主張・補正方針	OA要点抽出・証拠箇所抽出	期限管理・方式応答
	Default: 証拠集約はHOTL、実体判断はHITL		

業務プロセス別推奨モード (2) 契約・権利行使・機密管理

	HITL（事前承認）	HOTL（例外監督）	HOOTL（事後監査）
権利行使・侵害調査	クレームチャート・無効化反論（紛争耐性確保）	製品仕様からの要素対応一次マッピング	監視アラート・クロール
	Default: 初動トリアージはHOTL、本丸はHITL原則		
契約・ライセンス	個別交渉・責任配分条項のレビュー	プレイブック照合・逸脱箇所検知	標準雛形へのメタデータ差し込み
	Default: 個別交渉はHITL、標準契約運用はHOTL		
機密管理	社外モデル投入の例外承認・越境移転判断	DLPアラート・外部共有審査	自動ラベリング・保持期間経過による削除
	Default: ルール化できる統制はHOOTL/HOTL、例外判断はHITL		

統合的インサイト：知財AIを支える「3層のガバナンス・アーキテクチャ」

どの出力が、どの入力・根拠・モデル版に基づいて生成されたか。

監査ログ保存・キルスイッチ運用

案件をHITL/HOTL/HOOTLのどこに配分し、誰が承認プロセスを担うか。

プロンプト統制・RACIの明確化

どのモデル、接続先、データ領域（社内/公知/機密）が許可されているか。

ベンダー審査・保持期間確認

「AIが説明できること」ではなく、「人間が事後に結果を再構成（Verifiability）できること」が知財実務の真の命綱である。

主要ツールの特性と実務上の注意点（公式仕様ベース）

汎用LLMプラットフォーム（Generic GenAI）

OpenAI / Microsoft Azure

学習利用なし、SOC2監査済。Azureは機密RAGや統合監査に向くが、保存条件の精査必須。

Google Gemini / Anthropic Claude

Claudeは長文に強いが、無期限保持が既定の機能に注意。

国内特化LLM (NTT, NEC, Fujitsu)

機密性要件に応じたオンプレ配置や、日本語の微細な語感重視のPoCに価値。

知財特化ツール・データ管理（IP & Data Tools）

検索・分析 (J-PlatPat, Espacenet, PatSnap, Clarivate)

概念探索やワークフロー統合に強力。ただし、生成された法的推論部分は必ずHITL回帰が必須。

統制基盤 (Microsoft Purview, Box, iManage)

監査証跡、DLP、法的保持運用のコア基盤として機能。

導入時の評価基準：精度だけを追うと失敗する「多角的なKPI」

NISTやISOの文脈に沿い、単一KPIではなく継続的な管理スコアとして測定する。

1. 精度 (Accuracy)

誤差があっても人手補正可能か(HITL)、
安定して例外抽出が機能するか(HOTL)。

2. 再現性 (Reproducibility)

同一入力での出力のばらつきが
閾値内に収まるか。

3. コスト (Cost)

1案件あたりのAPI費用＋
「監査・人件費」の総計。

4. スループット (Throughput)

処理件数の向上、
処理件数の向上と、期限遵守率への寄与。

5. 人的リソース (Human Load)

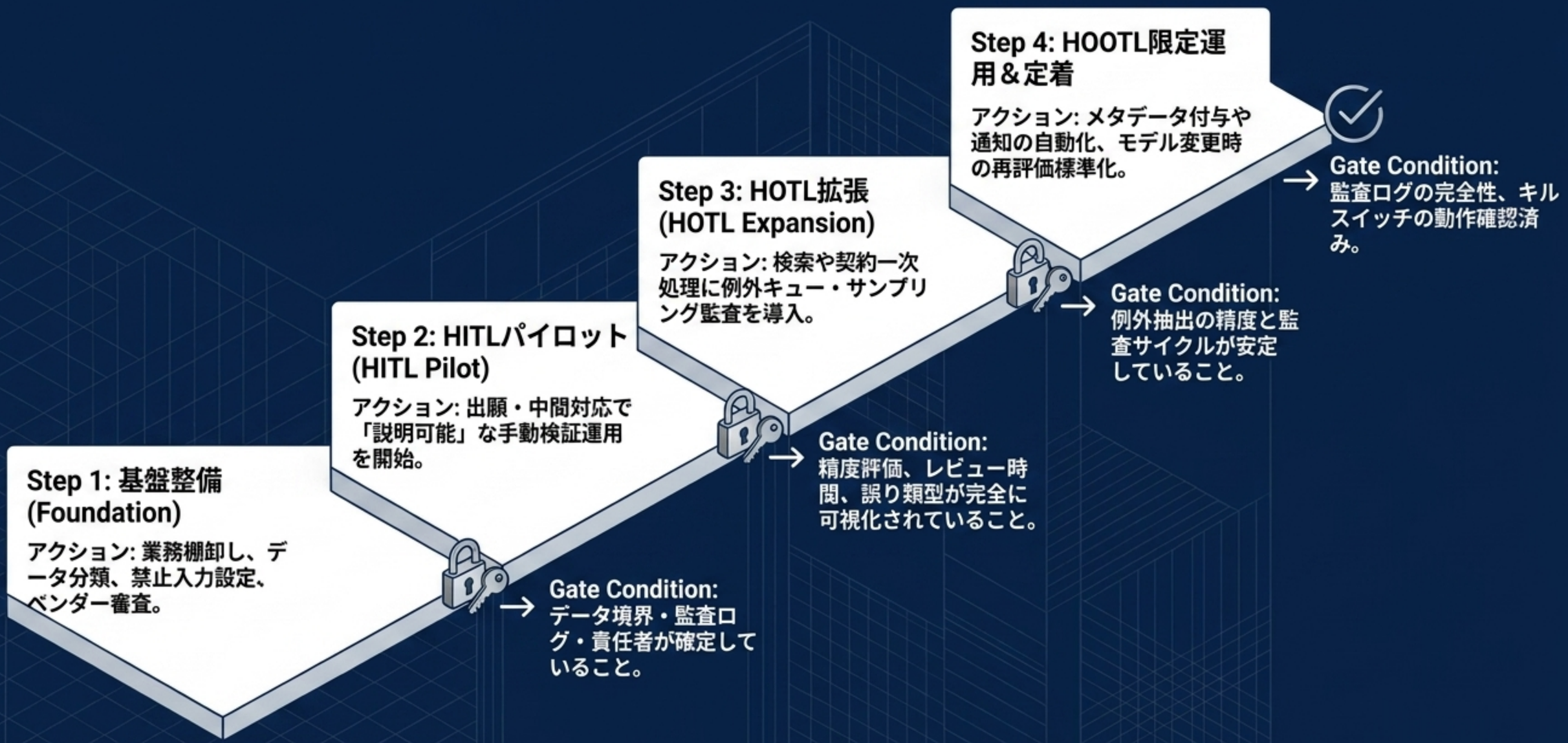
専門家のフルレビューが必要か(HITL)、
サンプルレビューで済むか(HOTL)。

6. リスク・セキュリティ (Risk/Security)

誤り時の法的影響、保持設定、DLP連携、
監査ログの完全性。

**Rule of Thumb:「精度が低いからHITL」ではなく、
「誤りのコストが大きいからHITL」とする。**

知財AIガバナンスの段階的移行ロードマップ



実務担当者向け AI導入・運用ダッシュボード・チェックリスト

導入前（基盤・評価設計）

- ☐ 業務をHITL/HOTL/HOOTLに分類し、RACI（責任分解点）を定義した。
- ☐ 未公開発明・営業秘密の「入力禁止・要承認・許可」の境界を決めた。
- ☐ ベンダーの学習利用、保持期間、ログ仕様、越境移転条件を契約レベルで確認した。

導入中（説明可能な運用）

- ☐ 重要出力には、参照ソースや根拠文書が必ず紐づいている。
- ☐ モデル版、プロンプト版、承認者のログが事後検証可能な状態で保存されている。
- ☐ 監督者に対し、自動化バイアスとハルシネーションに関する教育を実施した。

導入後（拡大とモニタリング）

- ☐ 月次で誤り率、差戻し率、レビュー時間、インシデントを監査している。
- ☐ 新規業務へ拡張する前に、既存業務での「例外処理」と「停止手順」が機能することを確認した。
- ☐ 対外提出案件について、提出後に根拠パックを保全した。

「再現不能な出力は、
権利を動かす工程に使わない。」

係争・監査・説明要求に備え、案件単位で「誰が、どの根拠で、どのAI出力を採用したか」
を再構成できる状態を維持することが、知財部門における最大の防衛線である。

準拠・参照フレームワーク: WIPO ガイドライン | OECD 自動化分類 | NIST AI RMF | ISO 42001 & 23894 | 経済産業省 AI事業者ガイドライン | 文化庁 AIと著作権整理