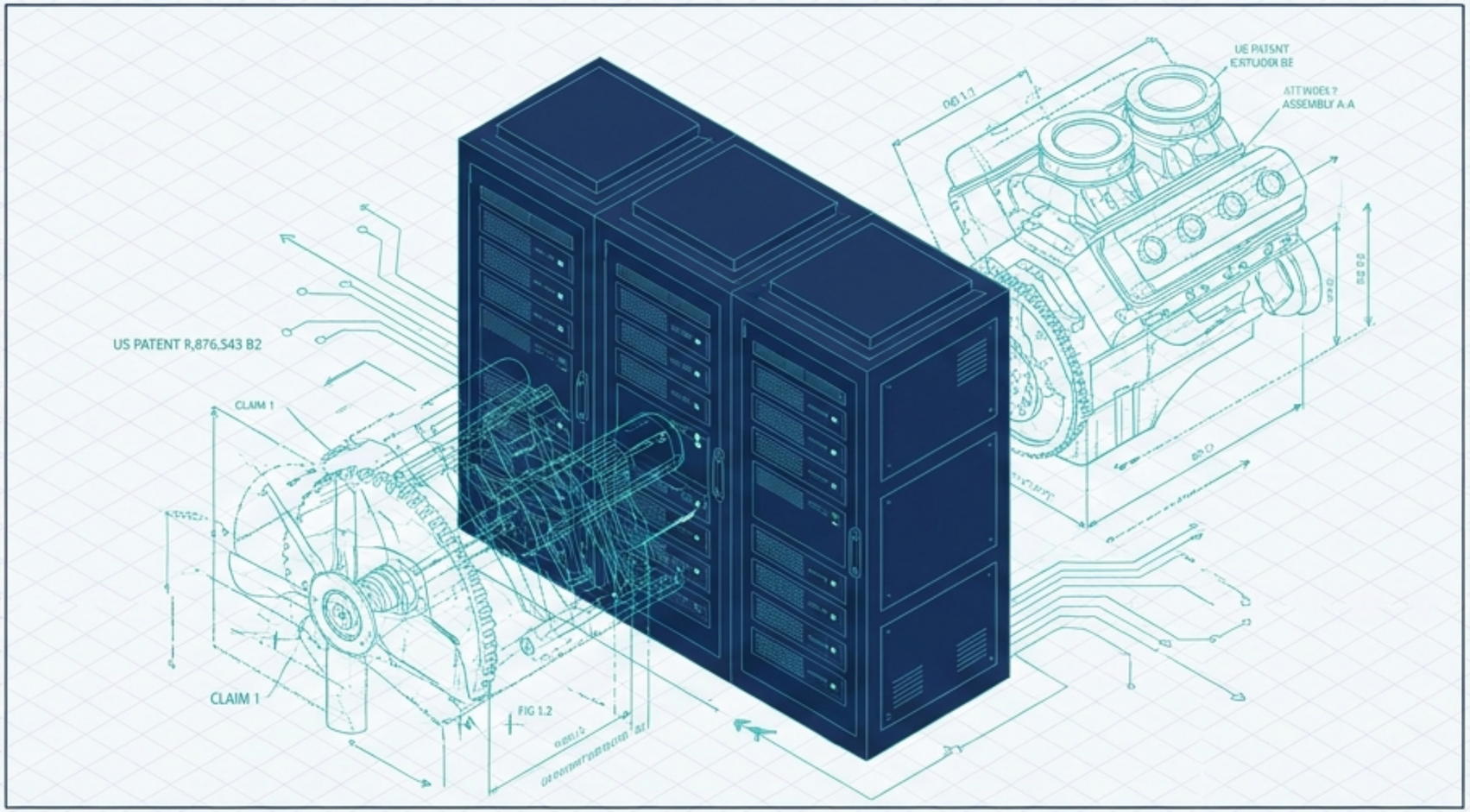
国産LLMの知財活用と課題 2026

汎用AIの試行期から、閉域・業務特化型AIエージェントへの移行



Gemini Notebook

Executive Summary



劇的効率化と審査水準の壁

定型調査業務で**最大94%の効率化**（**22時間→3時間**）を達成する一方、厳密な審査タスク（**GENIAC-PRIZE**）では依然として高い**技術的障壁**が存在。



「閉域×HITL」の絶対要請

弁理士法75条（非弁行為）および**新規性喪失リスク**により、AIによる全自動化は不可。専門家が最終責任を負う**Human-in-the-Loop (HITL)** が必須。



階層型アーキテクチャの構築

機密レベルに応じた「**オンプレミス／ゲートウェイ構成**」の選択と、**RAG (検索基盤)** の最適化を通じた実務統合が必要。

Gemini Notebook

なぜ知財分野で「国産・ローカルLLM」が要請されるのか？



新規性喪失・漏洩リスクの遮断

- 出願前の営業秘密を保護
- 外部送信による特許法上の新規性喪失を防止
- NDA・守秘義務（弁理士法30条）違反の回避

日本語特許構造・図表への適合

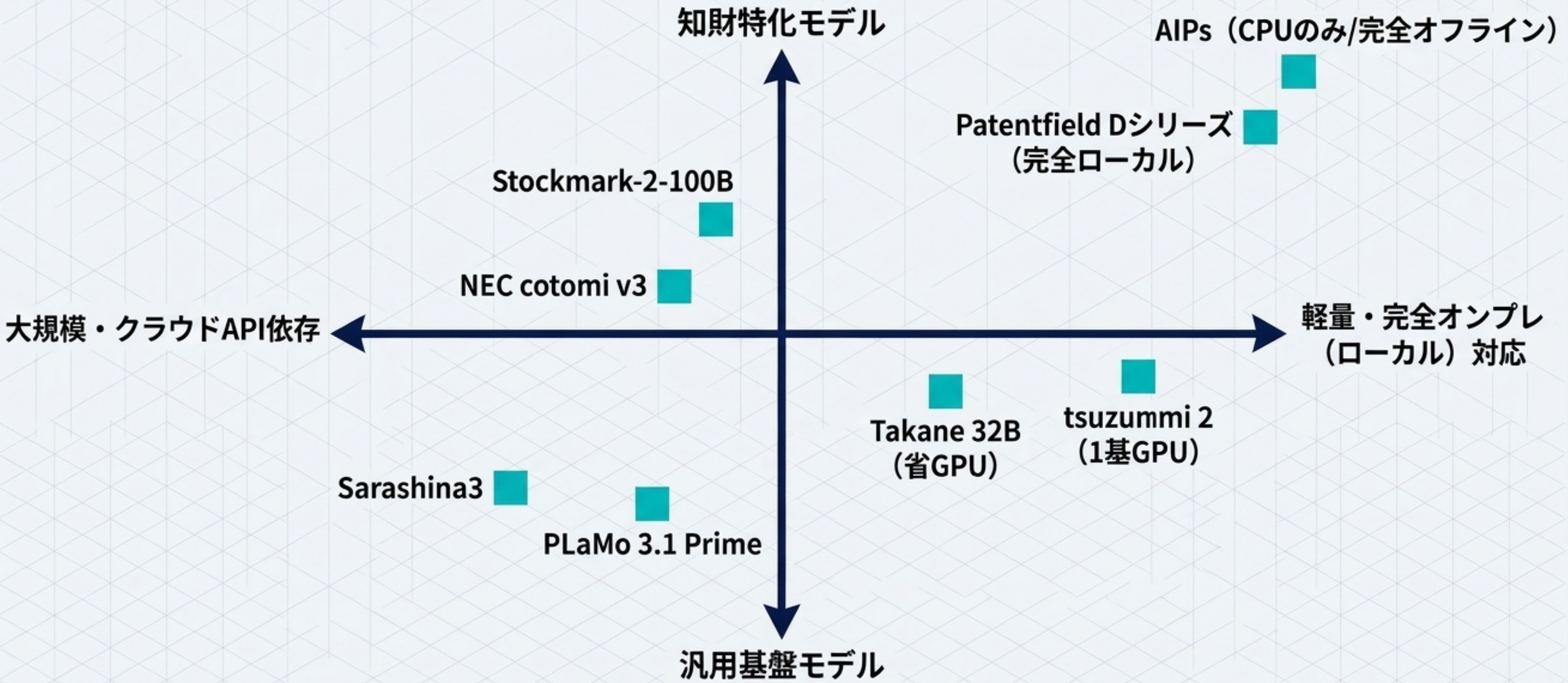
- 独特の修飾構造や上位概念表現の理解
- 特許法第36条に基づく請求項の対応
- 国内特有の図表文化、FI・Fタームとの整合

オンプレミスでのTCO逆転

- 数万件規模の文献一括処理におけるAPI課金回避
- 1ビット量子化による省リソース化
- CPU駆動モデルによる初期導入コストの抑制

Gemini Notebook

知財AIソリューションの3層構造と市場マップ



Gemini Notebook

ユースケース別 インフラ要件と導入モデルの選択肢

CPUのみ（非ネット接続PC）

モデル: AIPs

- GPU不要、完全オフラインのスタンドアロンPCで稼働
- 審査済み公報のみに学習を絞り込み軽量化
- 未公開発明の外部漏洩リスクを物理的に完全遮断

単一GPU・省リソース環境

モデル: tsuzumi 2, Takane 32B

- A100（40GB）1基等の限られたGPUで稼働
- 1ビット量子化等の技術による劇的なコスト削減
- 自社オンプレミス環境での機密情報の安全な処理

複数GPU・国内専有基盤

モデル: cotomi, Stockmark, PLaMo

- 100Bパラメータ級の大規模モデルや特許特化RAG
- 256K長文文脈をオンプレまたは国内専用クラウドで稼働
- IPランドスケープや日米欧の横断的な複雑探索

Gemini Notebook

民間・公的セクターにおける社会実装の実態

Private Sector (民間企業)

NEC

1,440通りのRAG最適化により特許調査を約22時間から約3時間へ短縮。定型業務で最大94%の効率化。「cotomi Gateway」により閉域環境と外部モデルを安全に振り分け。

Stockmark

製造業のR&D向けAIエージェント。日米欧の特許横断探索を実装し、暗黙知の組織知化を推進。

Public Sector (公的機関)

特許庁 (JPO)

2026年度にインターネット非接続の「AI基盤サーバ」をオンプレミスで導入し完全閉域化。FI/Fタームの自動付与や内外類似文献抽出を計画。

デジタル庁

ガバメントAI「源内」にて国内5モデル(tsuzumi 2, cotomi v3等) を契約・試用。公的機関が実用性とガバナンスを確認した参照リストに。

Gemini Notebook

精度限界の可視化：GENIAC-PRIZEが示した「審査水準の壁」



国内26年分の特許文献。
XMLのばらつき、意図的な言い回しの違いが存在。

正解文献の9割を包含するには、母集団の0.7%まで絞ってもなお膨大な件数の取得が必要。
単純なベクトル検索（RAG）の限界。

応募37件中。行政の特許審査が求める「網羅性（再現率）」と「論理的対比の厳密さ」の壁。

※完全ローカル構成のPatentfieldのみが技術新規賞を受賞。

Gemini Notebook

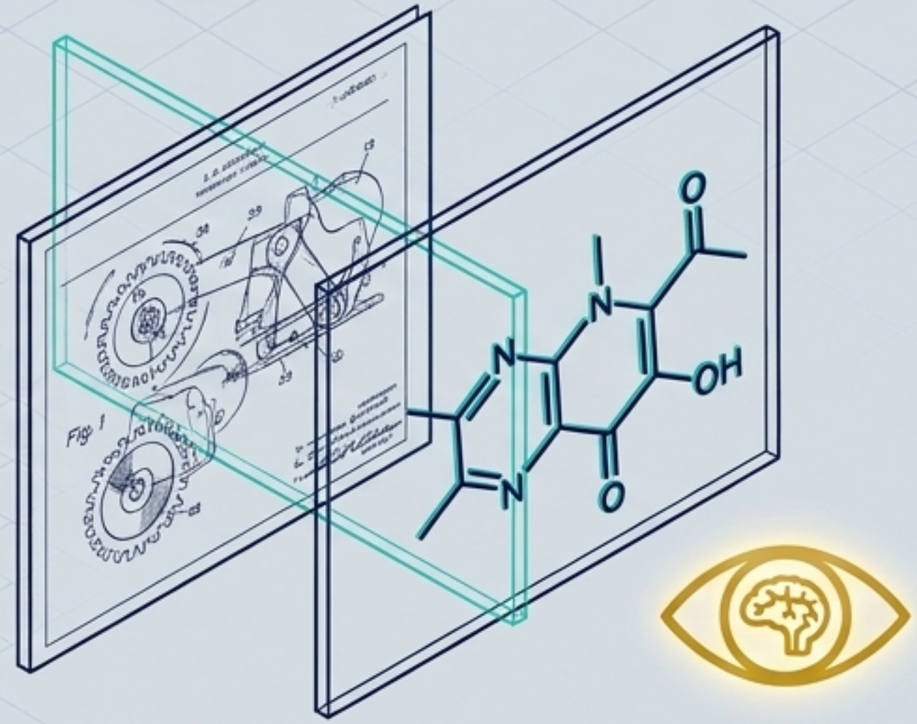
未学習領域（OOD）のジレンマとマルチモーダルの死角

未学習領域（OOD: Out-of-Distribution）



特許出願とは定義上「過去データに存在しない新規技術」。AIの学習済み分布外となるため、もっともらしい誤り（ハルシネーション）を構造的に引き起こしやすい。

マルチモーダル解釈の限界



図面画像や化学式の視覚対応は進展しているが、請求項と図面符号の「立体的・法的な整合性検証」を自動完結させるには、人間の専門家の精査が不可欠。

Gemini Notebook

法的・コンプライアンスの絶対的境界線

1. 弁理士法第75条（非弁行為）	AIによる特許性・侵害成否の鑑定、請求項の最終作成は違法。	【対応】AIは「下書き・整理」に限定し、最終判断は弁理士が行う。
2. 弁理士法第30条（守秘義務）・新規性喪失	外部AIへの未公開発明入力はNDA違反・権利化不能を招く。	【対応】非公開情報入力の厳禁、完全閉域環境の構築。
3. 発明者適格（自然人要件）	現在の法制度下では、AIは発明者になれない。	【対応】プロンプト履歴の保全により、人間の創作的寄与を事後立証可能にする。
4. 著作権法第30条の4 / 47条の5	社内RAGでの非特許文献（論文等）の過度な出力制限。	【対応】軽微利用・付随性の範囲内に収まるようシステム側で表示制御。

Gemini Notebook

シンセシス：なぜ「Human-in-the-Loop」が唯一の解なのか

技術的限界

多段階推論の限界、OODでのハルシネーション、網羅性のボトルネック。
→「技術的に人がカバーせざるを得ない」

結論：Human-in-the-Loop (HITL)

AIは全自動化ツールではない。専門家を支える「オーケストレーター・証拠整理器」として設計し、最終判断を人間が担う「HITL」こそが唯一の法的・技術的最適解である。

法的制約

非弁行為の禁止、未公開情報の守秘義務、自然人による発明者性の担保。
→「法的に人が責任を負わねばならない」

Gemini Notebook

アーキテクチャの青写真：機密レベルに応じたデータの階層化

高機密層

完全オフライン・閉域オンプレミス。未公開発明メモ・実験データをCPU駆動モデルや自社専有基盤で処理し、新規性喪失を遮断。

中間層（マスキング）

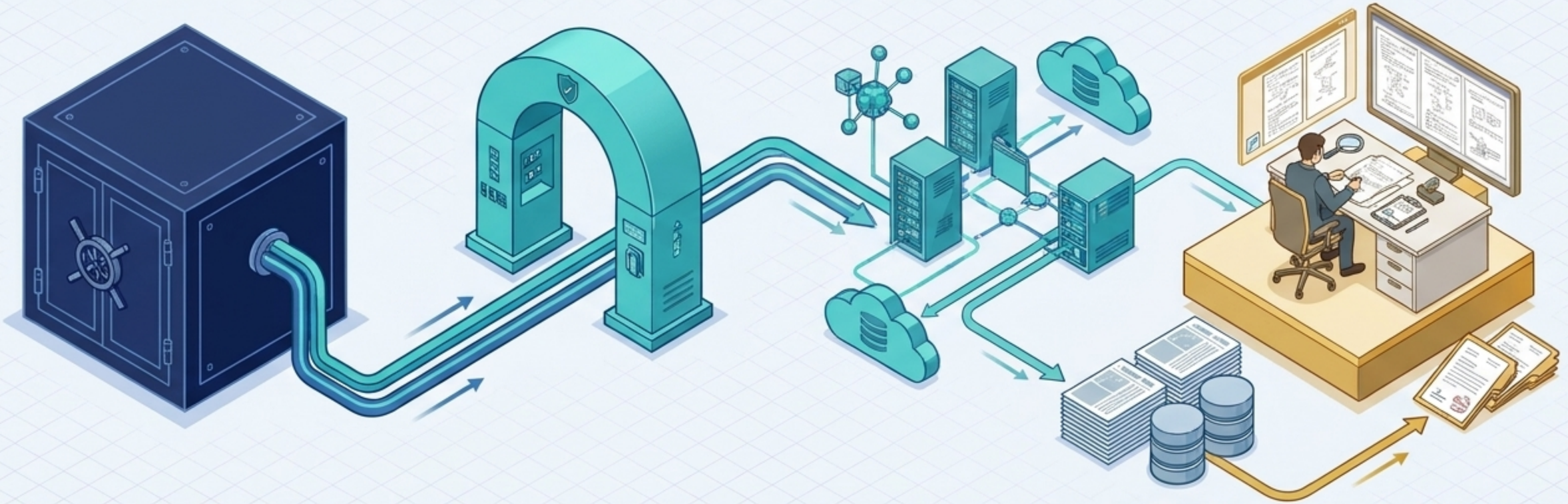
ゲートウェイ・MCP制御。機密箇所を伏せ字化して外部モデルや特許DBへ安全に接続。

公開情報層

クラウドLLM・外部DB。公開特許公報や論文を大規模モデルを用いてマクロ分析や多言語探索。

終着点

弁理士・審査官の検証。HITLによる最終出願書類・審査判断の確定。



Gemini Notebook

アーキテクチャの最適化：パラメータ数より「検索基盤（RAG）」



Step 1. 前処理（ノイズ除去）

特許XML特有の記述のばらつき、意図的な広い権利解釈の言い回しを正規化。

Step 2. 多段階検索・再ランキング

Bi-EncoderとCross-Encoderを組み合わせ、第一段階で数万件を抽出し、さらに厳密な再ランキングを実行して再現率を担保。

Step 3. GraphRAGによる構造化

単純な類似度検索を超え、特許・技術・課題を単一の知識グラフ上で結びつけ、関係性を抽出。

Gemini Notebook

組織的ガバナンスと監査証跡の構築

チェック1. プロンプト履歴の保全

発明発掘プロセスにおけるAIとの対話履歴を事件記録に準じて保存。人間の創作的寄与（発明者性）を事後立証する証跡とする。

チェック2. 学習オプトアウトと権限分離

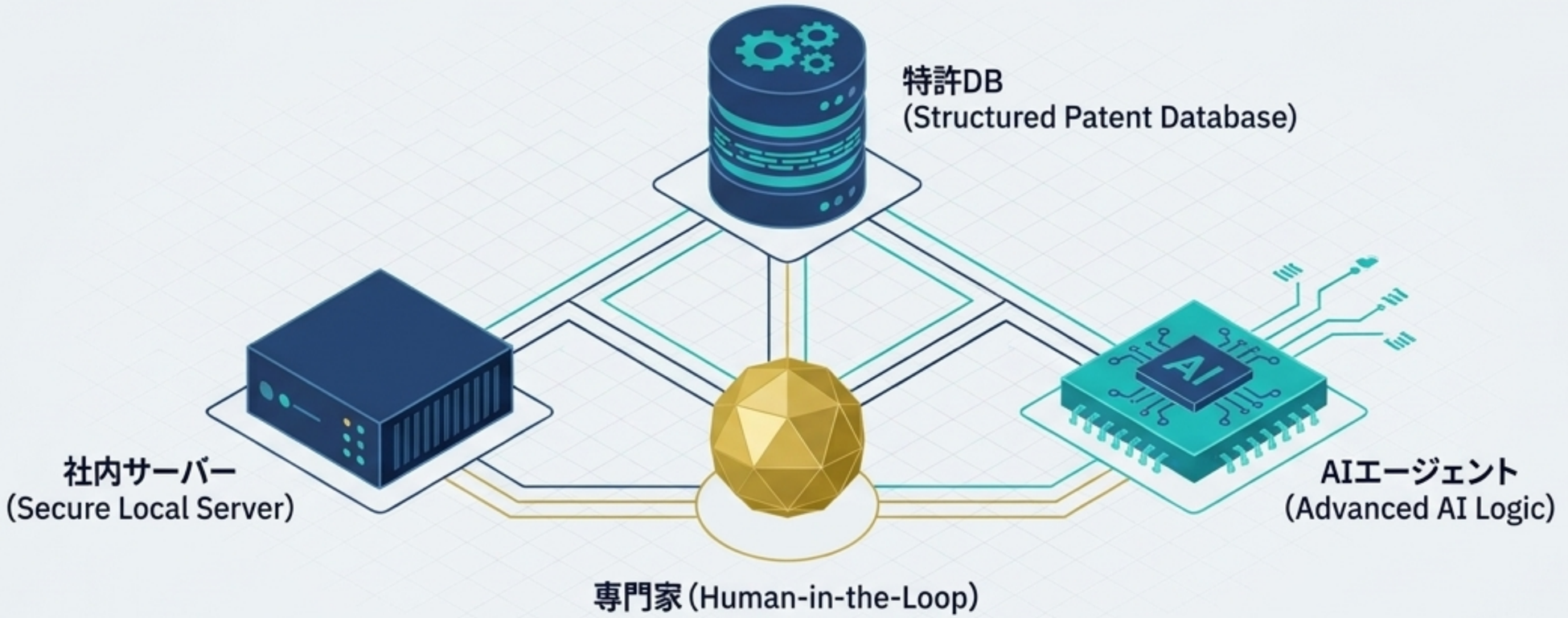
外部AI利用時の学習不可設定の徹底。外部委託先（特許事務所）のデータ分離環境（ハーネス）基準の合意。

チェック3. 役割分掌の明文化（JIPA基準）

AIは「対比表の下書き・技術要素の抽出」まで。請求項の設計と法的判断は「知財担当者・弁理士」に固定する社内規程の整備。

Gemini Notebook

AIを「全自動化ツール」から「専門家のオーケストレーター」へ



国産LLMとローカル環境の進化により、知財実務は新たなステージに入りました。
「海外モデルか国産か」という二者択一から脱却し、機密レベル・推論難度・法的責任に応じた『ハイブリッド設計』こそが競争力の源泉となります。
高度な技術基盤（Vault）の上で、専門家の判断（Precision）を最大化するアーキテクチャを今、構築すべきです。

Gemini Notebook