



国産LLM「Sarashina」とソブリンクラウドが 切り拓く知財業務DXの未来

デジタル主権に基づく自律型AIエージェントの社会実装

ソフトバンク株式会社 / SB Intuitions株式会社
2026年 戦略ブリーフィング

最高機密・特殊言語空間における生成AIの最適解

知的財産 (Intellectual Property)

技術主権

フルスクラッチ開発
の透明性

海外OSSに依存せず、
ゼロから構築。ブラック
ボックスを排除し、産
業用途に不可欠な「説明
可能性」を担保。

インフラ主権

ソブリンクラウドの
堅牢性

国内データセンターに閉
じた環境。API入力デー
タは学習に利用されず、
未公開技術の漏洩リスク
をゼロに。

知財特化性能

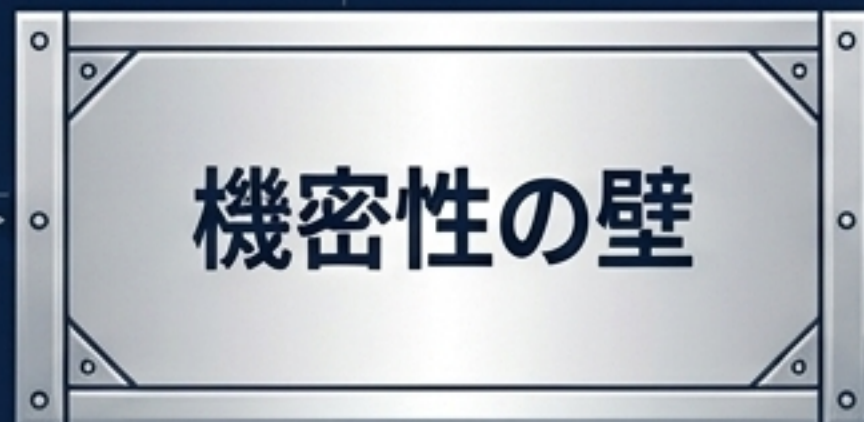
高度な専門日本語・
マルチモーダル処理

JPO基準の論理構造、
ベクトル検索 (RAG)、
特許図面解析に対応する
特化型モデル群による圧
倒的な業務適合性。

エンタープライズAI実装を阻む「3つの障壁」

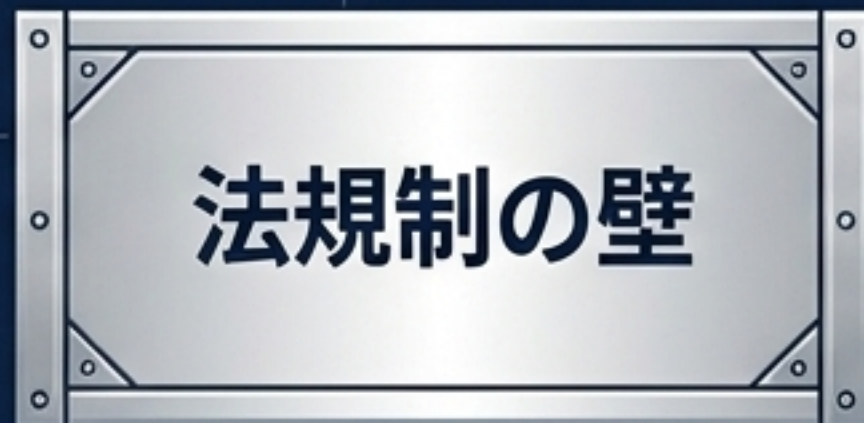


海外製
汎用LLM



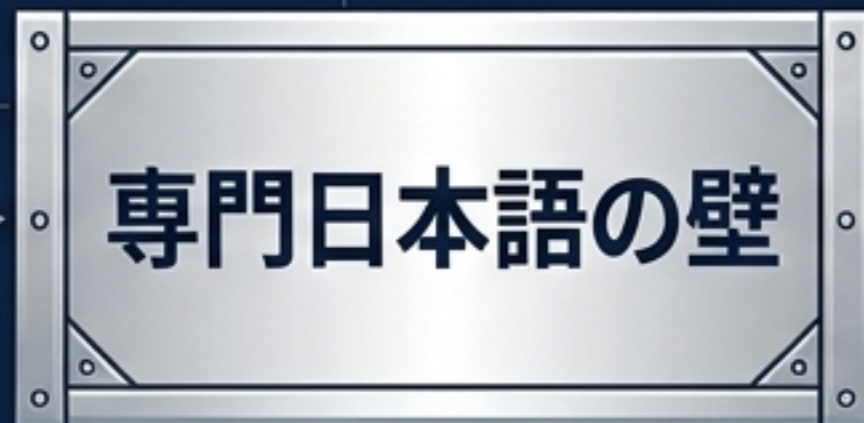
機密性の壁

未公開の発明提案書や営業秘密をパブリッククラウドに送信することによる、新規性喪失・データ漏洩の致命的リスク。



法規制の壁

CLOUD法などによる外国政府のデータアクセス権限や、GDPR等による複雑な域外移転規制への対応コスト。



専門日本語の壁

特許請求の範囲特有の論理構造や厳密な法的・技術的語彙に対する汎用モデルの理解不足。



エンタープライズ
知財データ

デジタル主権を構成する4つの柱



データ主権

物理的保管場所とアクセス権限を日本国内法の下で厳密に管理。外部へのデータ持ち出しを完全に遮断。



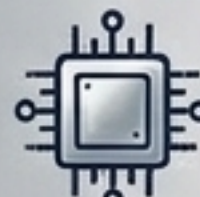
システム主権

重要コンポーネントを国内で構築・維持。海外ハイパースケラーへの過度な依存を排除し、仕様変更リスクを回避。



運用主権

データセンターの運用管理、障害対応、監視体制を国内組織で完結。地政学的リスクによる停止を防ぐ。

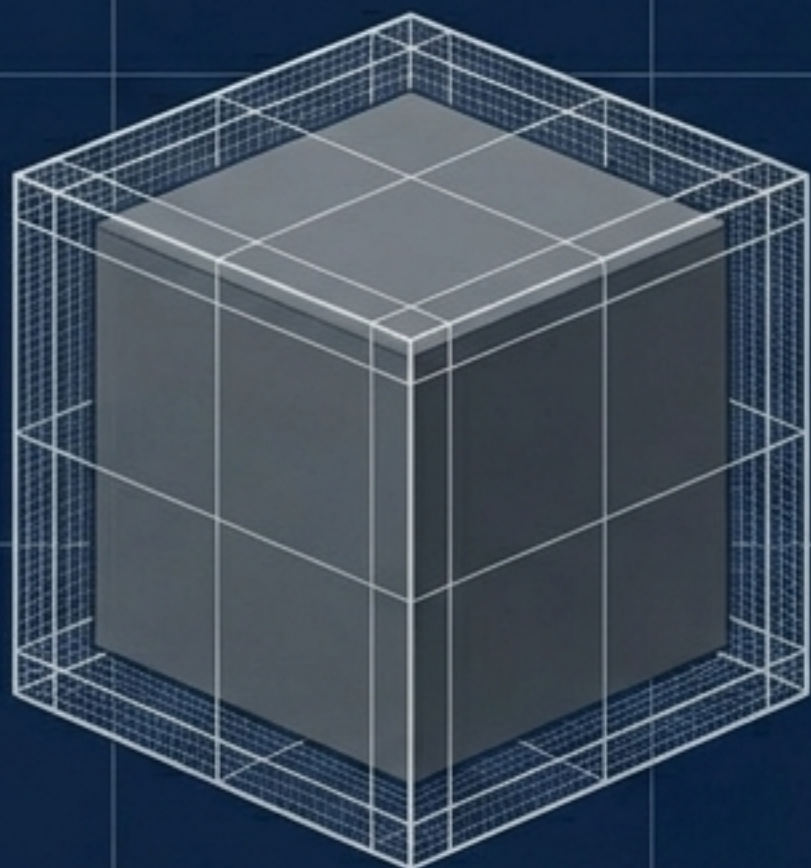


技術主権

Sarashinaのスクラッチ開発によりコア技術の自律性を確保。長期的なAI競争力を国内で維持・発展。

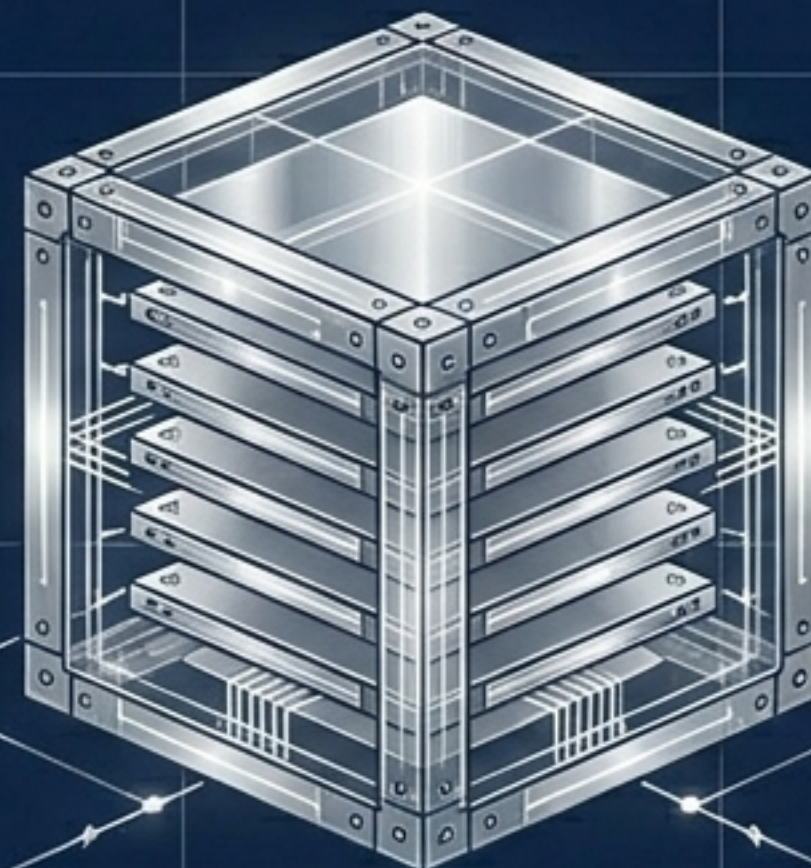
なぜ「フルスクラッチ」が知財業務に必須なのか？

汎用OSS + 追加学習



内部挙動のブラックボックス化によるリスク







Sarashina (フルスクラッチ)



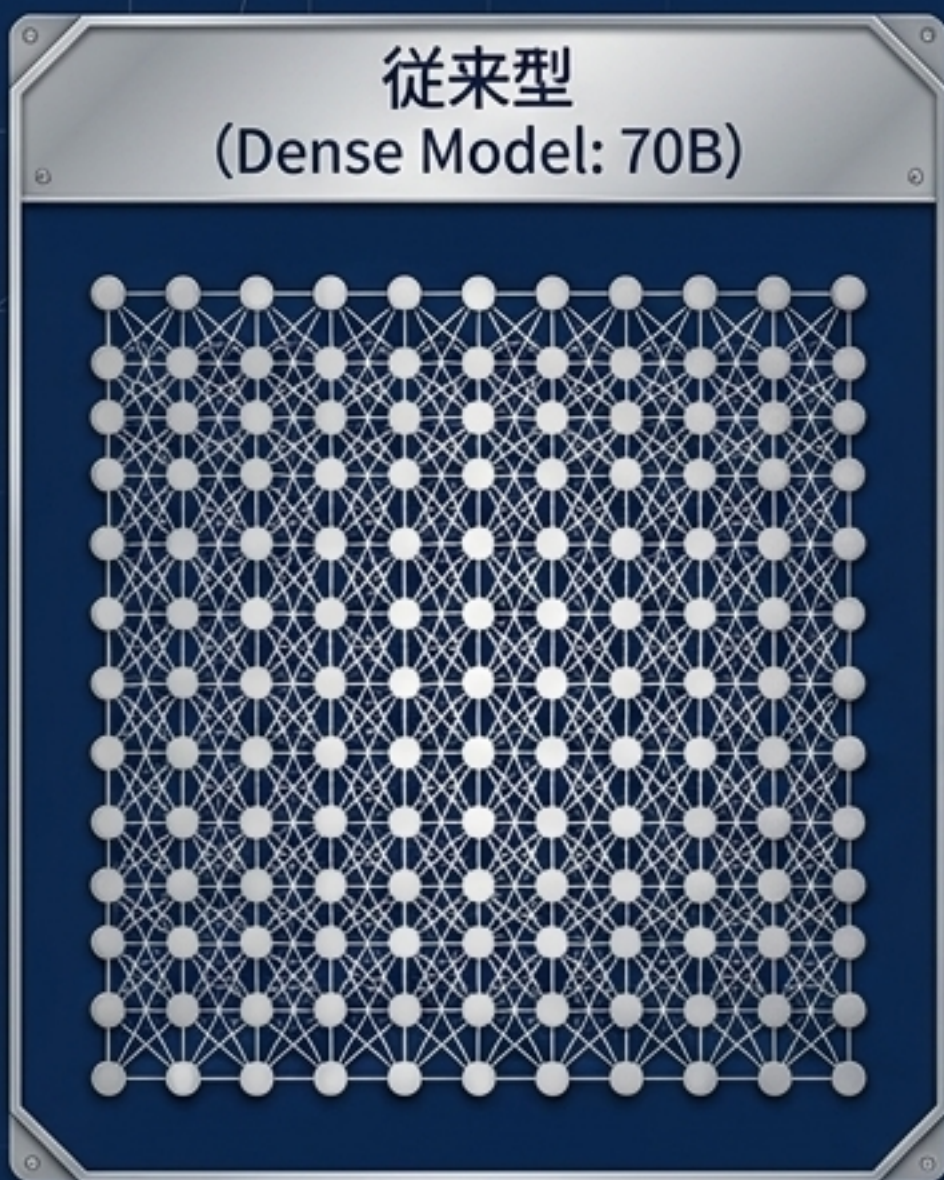
🔒 アルゴリズム・データ構造の完全掌握

- ▶ 初期化からの独自構築：日本語データによる単なる追加学習ではなく、初期段階から独自開発する哲学。
- ▶ 説明可能性の担保：バグ発生時や法的根拠が問われる際、AIの判断プロセスを追跡できる安心感を提供。

Sarashina エコシステム: マルチモーダル対応マトリクス

カテゴリ	代表モデル	パラメータ規模	主要機能・特徴
 テキスト生成	Sarashina2-70B	70B	Llama2アーキテクチャ、102.4kボキャブラリー
 MoE / 高性能	Sarashina2-8x70B	465B	Sparse upcycling手法、日英バイリンガルウェブコーパス
 ビジョン/ マルチモーダル	Sarashina2.2-Vision-3B	4B	Image-to-Text推論
 OCR	Sarashina2.2-OCR	3B	日英ドキュメント解析
 エンベディング	Sarashina-embedding- v2-1B	1B	RAGおよびSTS向けに最適化
 音声	Sarashina2.2-TTS	0.8B	ゼロショット音声生成

「賢さ」と「軽さ」の両立：Sparse Upcyclingのメカニズム



タスク実行時にモデル全体が活性化。
計算負荷が高い。

入力クエリ



特定の専門エキスパートのみが発火。
最高峰の精度を最小限の計算リソースで実現。

4,650億パラメータの巨大な表現力を持ちながら、推論コストを劇的に低減。

絶対的な安全域：ソブリンククラウド「Cloud PF Type A」

国内基盤

東日本データセンター内に構築された圧倒的計算リソース。

完全閉域環境

外部インターネットへのデータ流出経路を物理的・論理的に遮断。

物理的・デジタル境界 (Trust Zone)

1万基超のGPU計算リソース

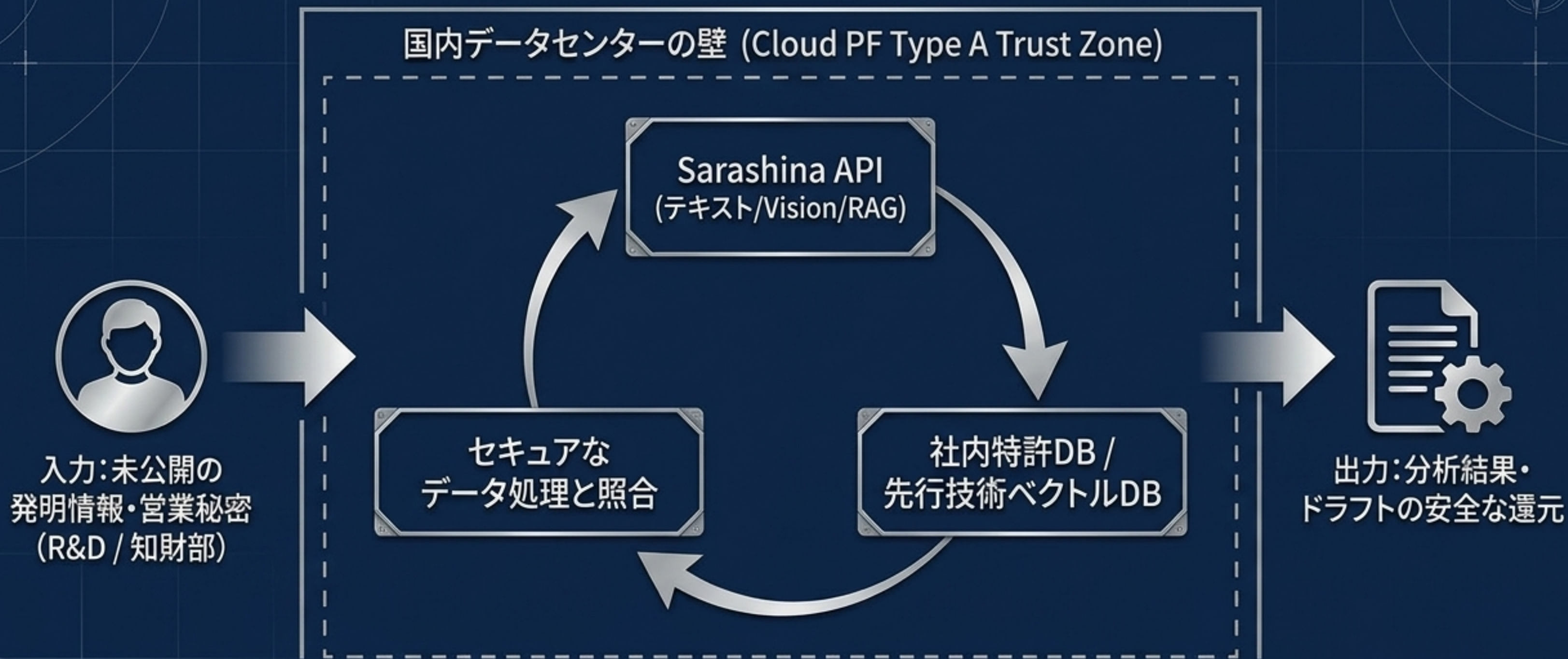
ノー・トレーニング・ポリシー

入力された顧客データは、AIの再学習に一切使用されないことを担保。

顧客主導のアクセス制御

暗号鍵の顧客管理により、ソフトバンク側からもデータ内容にアクセス不可。

Sovereign IP AI Workflow



「最先端のマルチモーダルAI」と「物理的・法的な盾」の融合により、最高機密の本格的な生成AI運用が実現。

ユースケース1: JPO実務に準拠した明細書起案



汎用的なAI

Sarashina (知財特化)

入力文章

汎用トークナイザ

日常語としての解釈

特許独自の論理構造の破綻

発明提案書

Unigram + 102.4kボキャブラリー

厳密な法的・技術的語彙の正確な解釈

JPO基準のドラフト生成

【重要】弁理士・知財担当者による必須レビュー

権利範囲、サポート要件、禁反言リスクの最終確認を行い、法的完全性を担保。

ユースケース2：ベクトル検索による高度なクリアランス調査



従来型：キーワード完全一致

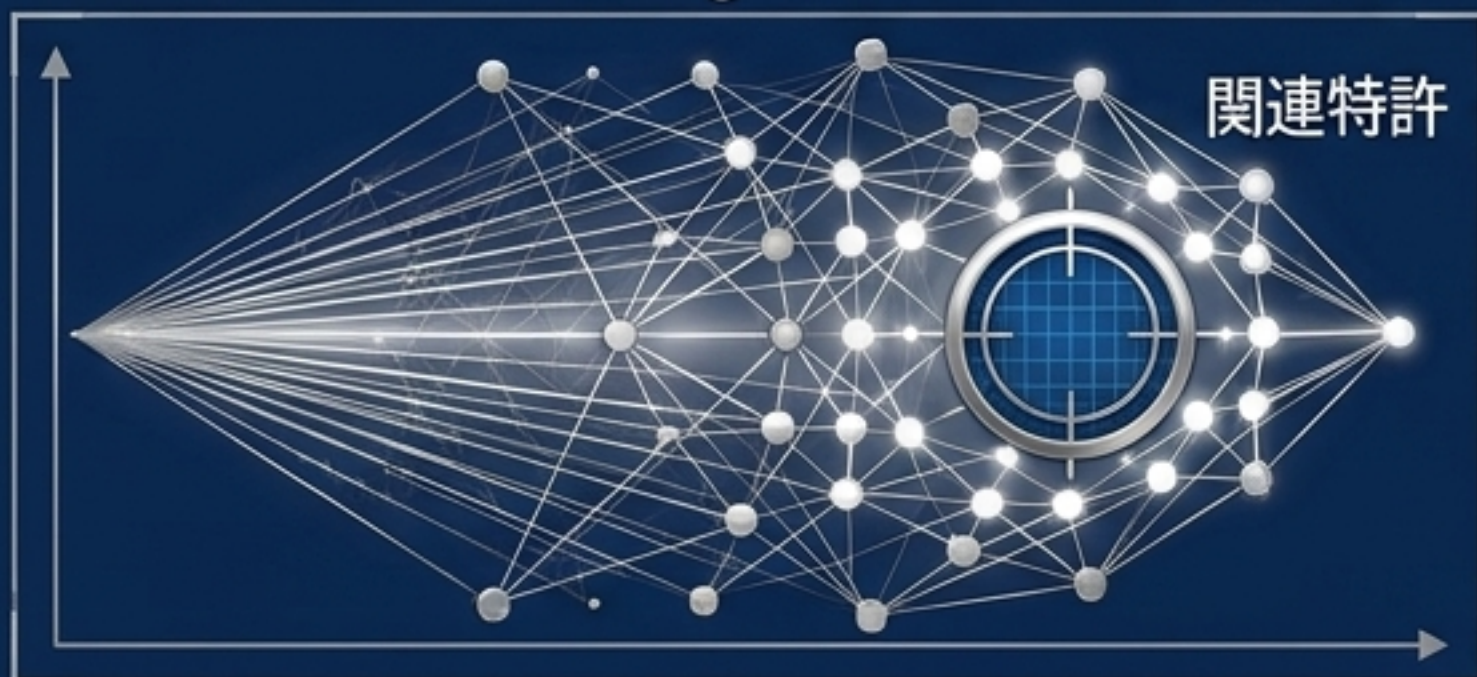


キーワードの揺らぎによる検索漏れが発生。



クレーム対比表の
自動生成補助

Sarashina-Embedding-v2-1B：意味的類似度

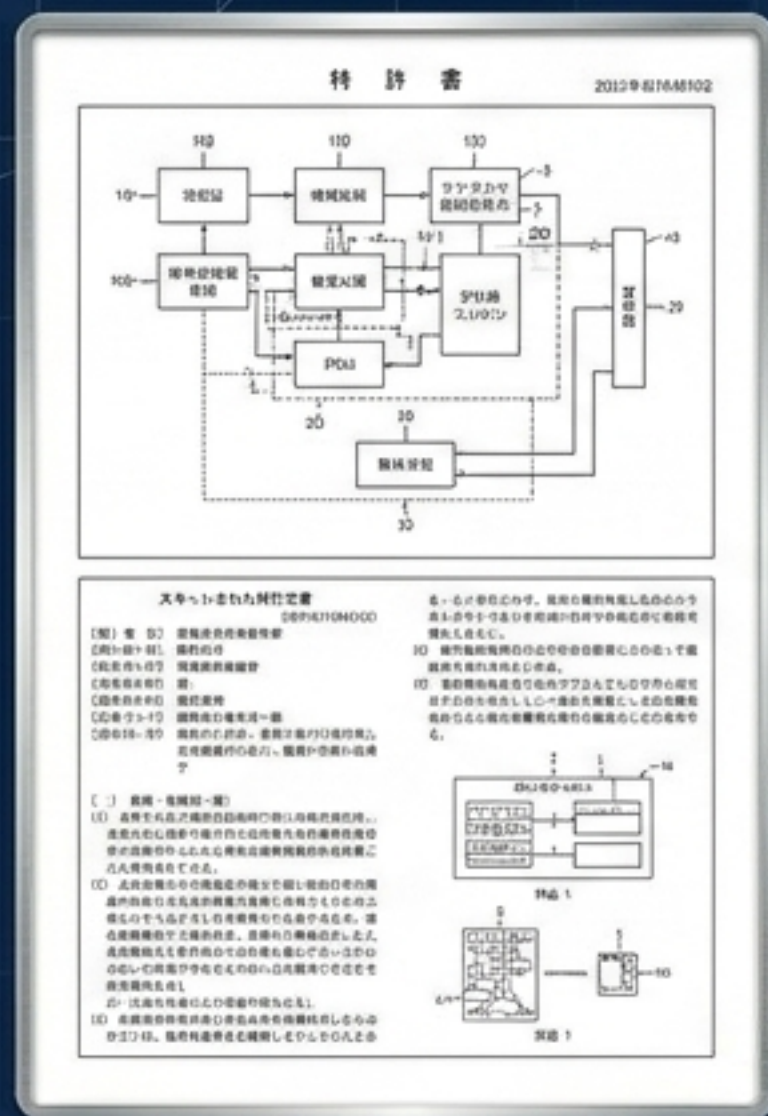


課題解決原理や構造の機能的類似性に基づくRAG検索。精緻なスクリーニングを実現。

- 数千万件の公報から引例を高速抽出。
- 自社発明との「一致点」「相違点」を構成要件ごとに整理。
- 目視スクリーニングの膨大な時間を削減。

ユースケース3：図面とテキストのマルチモーダル統合解析

Sarashina2.2-Vision-3B



スキャンされた特許文書
(非定型)



図面と明細書の
クロスリファレンス解析

Sarashina2.2-OCR (3B)



複雑なレイアウトからの
構造化テキスト抽出



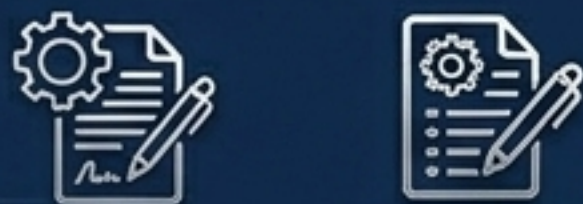
統合化された
知財構造データ

※化学構造式やDNA配列等については、専用の認識・変換ツールとのシステム統合を推奨。

知財AIの未来像：自律型エージェントへの進化



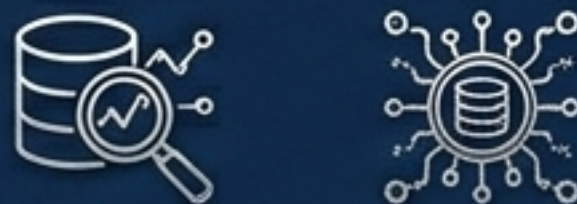
Phase 1: 特定タスク向け



文書作成エージェント

人間が指示した単一のタスク（明細書ドラフト作成、拒絶理由通知の論点整理）を実行。

Phase 2: プロセス拡張



先行技術調査エージェント

特許DBへ自律的にクエリを発行し、関連度の高い引例の抽出からスコアリングまでのワークフローを担う。

Phase 3: 完全自律化



マルチエージェント協調

調査、起案、探索的データ解析など、専門機能を持つ複数のAIが相互連携し、専門家と高度に協調するシステム。

※中外製薬との共同研究モデルに基づく進化の系譜

知財DX実装へのロードマップとガバナンス

Technical Setup (PoC検証項目)

- RAG環境の構築：社内特許公報、審査基準、過去の出願書類のベクトル化。
- 精度検証：検索再現率、クレーン対比精度の測定。
- リスク統制：ハルシネーション率の測定、監査ログ取得の確認。
- セキュリティ：機密情報保護（DLP連携）、厳密な利用権限管理。

Human Governance (人間とAIの協調)



AIの出力は最終判断ではない。
特許請求の範囲の設計、FTO判断、出願可否など、法的・技術的責任を伴うプロセスは『人間の専門家（弁理士等）による最終レビュー』を必須とするガバナンス設計を厳格に実装する。

デジタル主権が、日本の知財戦略を次の次元へ

データ主権

透明性

日本語の深淵な理解

SarashinaとCloud PF Type Aが提供するこの3つの絶対的価値は、単なる業務効率化を超え、未公開技術を安全に資産化する「次世代の知財プラットフォーム」を創出する。



第一歩としてのセキュアPoC環境構築へ。
戦略ブリーフィングのご提案。