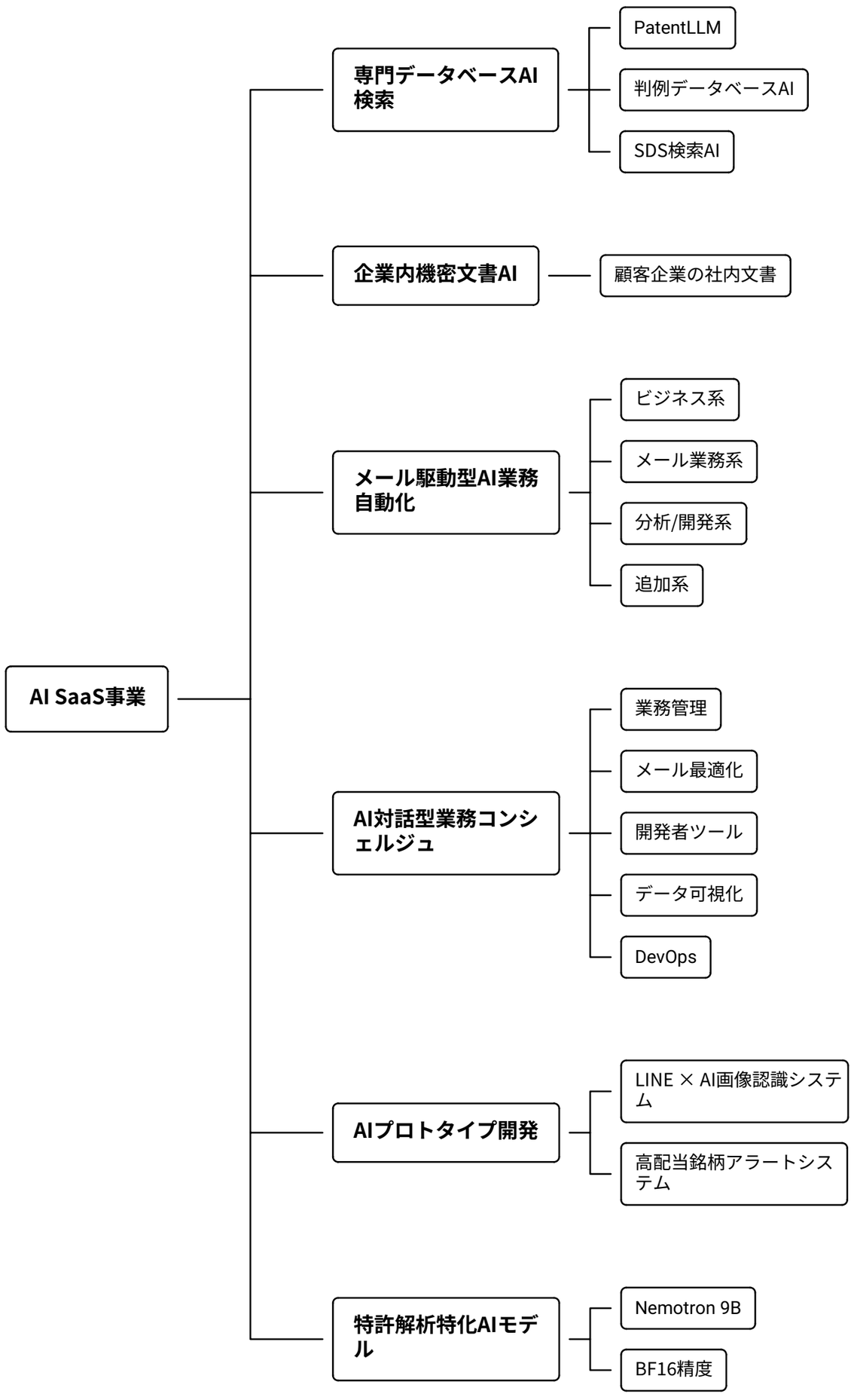
RAG事業

事業者概要
弁護士 × 将棋AI研究者 × フルスタックAIエンジニア。知的財産の専門知識とAI開発技術を兼ね備え、RAG（検索拡張生成）技術を中核に、AIサービスの設計・開発・運用を一人で完結

サービスライン全体像：**6**本柱で企業の業務を丸ごと**AI**化
本事業は特定分野に限らず、企業のあらゆる業務領域をAIで変革するプラットフォームを提供する。

1. 専門データベースAI検索（公開データ型RAG）
   - PatentLLM：米国特許400万件（→2000万件拡張計画）をAIが横断検索・分析
   - 判例データベースAI：最高裁判例30,000件をAI検索
   - SDS検索AI：化学物質安全データシートを横断検索
2. 企業内機密文書AI（機密データ型RAG）
   - 顧客企業の社内文書（議事録・マニュアル・提案書等）を取り込んだ「自社専用AI」を構築
3. メール駆動型AI業務自動化（41種類・24時間常駐）
   - ビジネス系10種：経費精算・日報・議事録・見積・請求・顧客管理・在庫・承認・KPI・面談報告
   - メール業務系10種：翻訳・要約・返信案生成・クレーム対応・リサーチ・スケジュール・タスク・フォロー・FAQ・週報
   - 分析/開発系10種：データ分析・競合調査・特許調査・契約チェック・採用管理・バグ報告・デプロイ記録・監視アラート・コードレビュー・ダッシュボード
   - 追加系8種：受注管理・出退勤・営業メール・署名・リード管理・社内通知・テンプレート・ニュースレター
   - Gmail RAG：過去メール全文をAI検索し、経緯や意思決定の流れを自然言語で回答
4. AI対話型業務コンシェルジュ（100種類以上のWebアプリ）
   - 業務管理30種：タスク管理・経費・請求書・タイムトラッカー・カンバン・CRM等
   - メール最適化30種：受信トレイ管理・テンプレート・スケジューラー・A/Bテスト等
   - 開発者ツール20種：JSONビューア・正規表現テスター・SQLビルダー・APIテスター等
   - データ可視化10種：チャート・ダッシュボード・フローチャート・ER図・マインドマップ等
   - DevOps 10種：Webhookテスター・Docker Composeビルダー・CI/CDパイプライン等
5. AIプロトタイプ開発
   - LINE × AI画像認識システム（顧客納品実績あり）
   - 高配当銘柄アラートシステム等、個別ニーズに応じた開発
6. 特許解析特化AIモデル
   - Nemotron 9BをBF16精度でファインチューニングし、特許文書解析に特化した独自AIモデルを構築

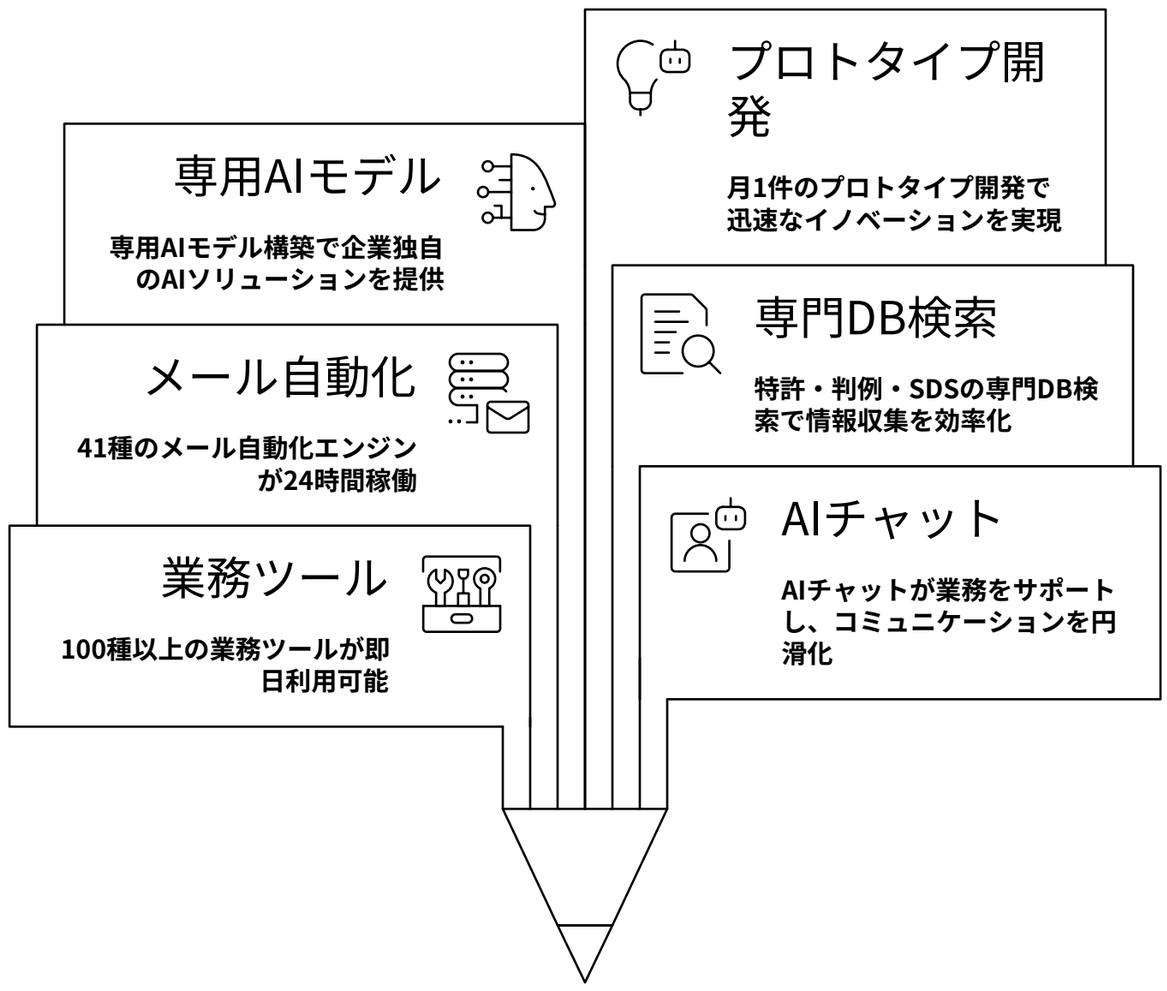
AI SaaS事業のサービスラインナップ



顧客が得る価値：「AI部門を丸ごとアウトソース」

- 100種以上の業務ツール＋AIチャットが即日利用可能
- 41種のメール自動化エンジンが24時間稼働＋特許・判例・SDSの専門DB検索
- AIエンジニア1人分の働き：専用AIモデル構築＋月1件のプロトタイプ開発

AIによる業務変革



導入の手間はゼロ — ブラウザを開くだけ。ソフトのインストールもIT機器の購入も不要。

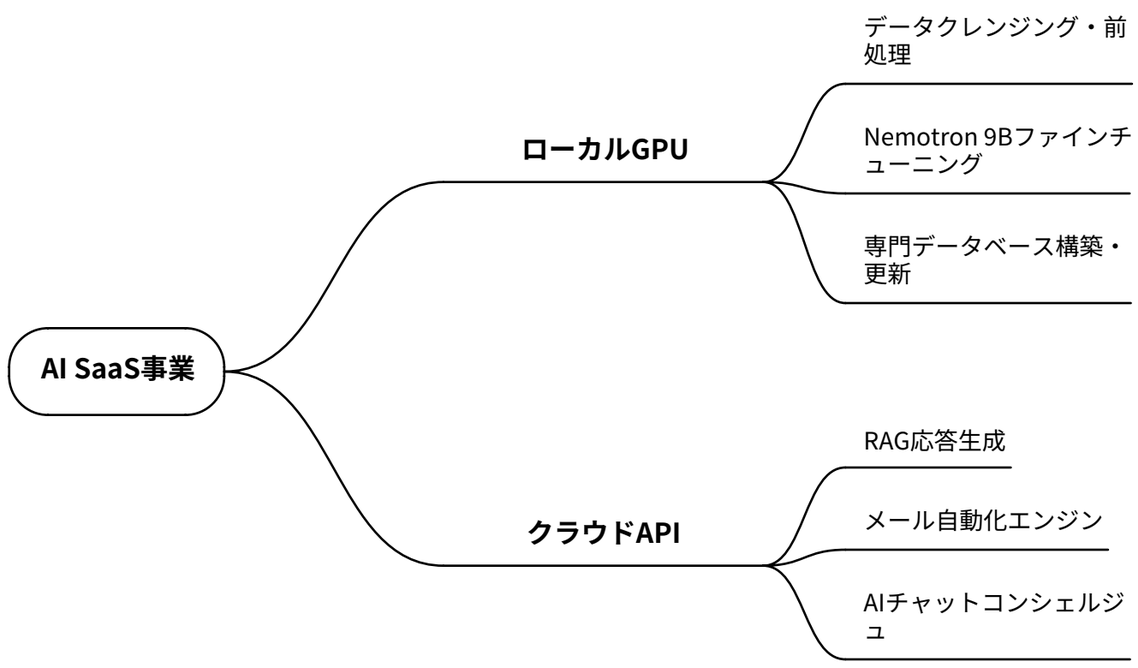
アーキテクチャの革新：ローカル**GPU**とクラウド**API**の最適配置

本事業の技術的革新は「ローカルとクラウドの使い分け」にある。

ローカルGPUの役割：
- 数百万件〜数千万件規模のデータクレンジング・前処理（クラウドAIでは数百万円かかる処理を自社設備で実行）
- Nvidia Nemotron 9Bのファインチューニング（特許解析特化の独自モデル構築）
- 専門データベースの構築・更新（数週間に及ぶ連続バッチ処理、ECC搭載で正確性保証）

クラウドAPI（Gemini / Claude Opus等）の役割：
- 顧客向けRAG応答生成（最新・最高品質のモデルで回答）
- メール自動化エンジンの知的処理（41種類のAI判断）
- AIチャットコンシェルジュの対話処理

AI SaaS事業におけるローカルGPUとクラウドAPIの役割



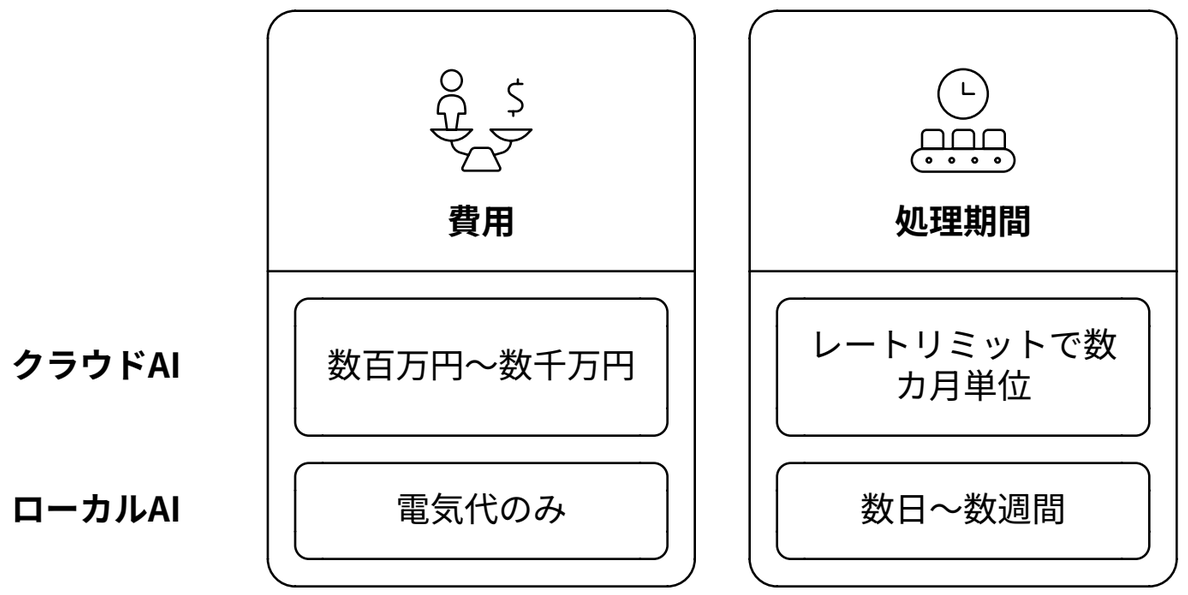
この組み合わせにより「大量データ処理はローカルで低コスト」「顧客体験はクラウドAPIで最高品質」を両立する。

なぜローカル**GPU**が必要か

2000万件の特許データをクラウドAPIで[クレンジング+タグ付け+付加情報]処理した場合とのコスト比較：

処理方法	費用	処理期間
クラウドAI（GPT-4等）	数百万円〜数千万円	レートリミットで数カ月単位
ローカルAI	電気代のみ	数日〜数週間

コスト比較

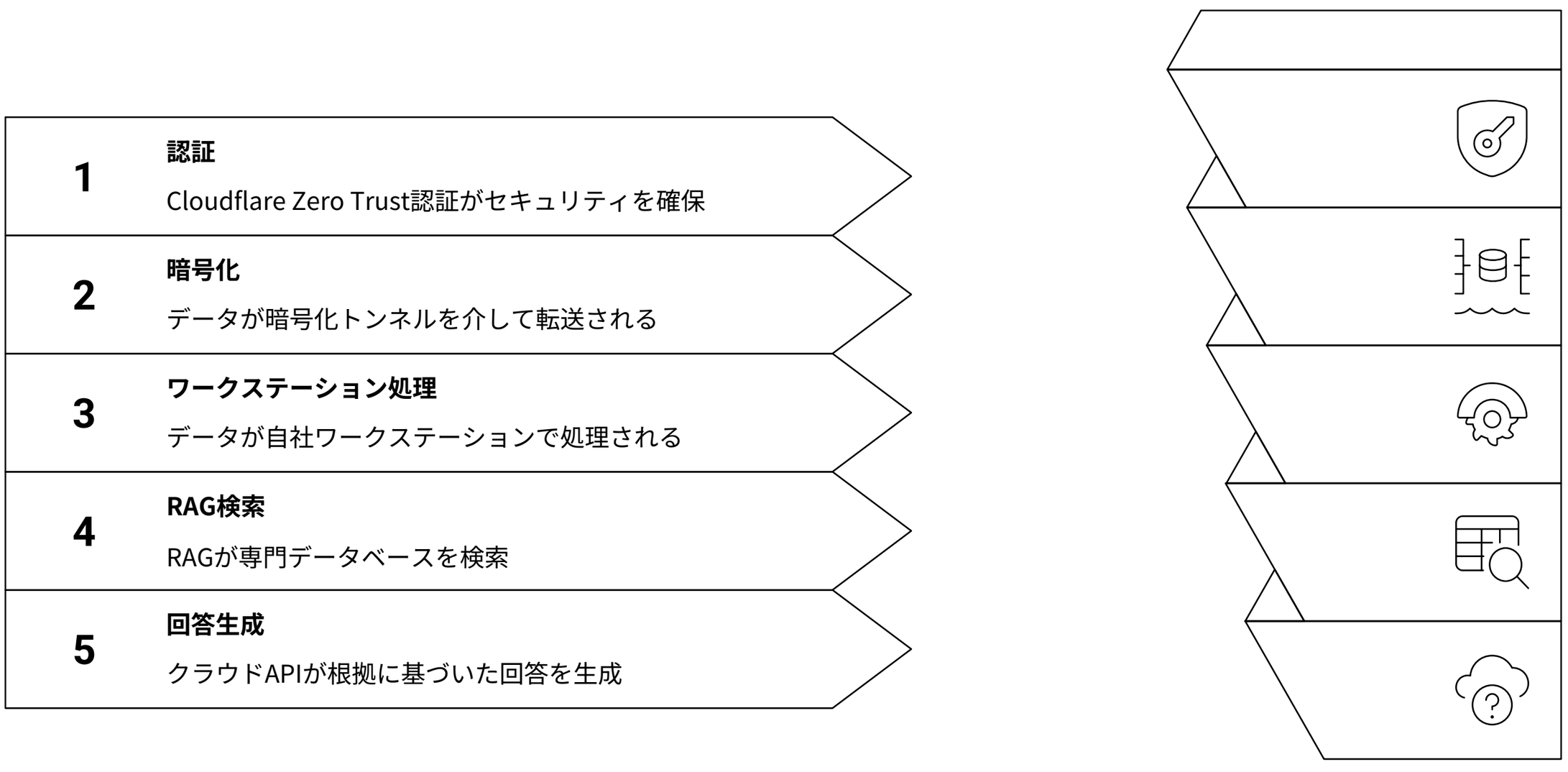


サービス提供の仕組み

顧客（ブラウザ）→ Cloudflare Zero Trust認証 → 暗号化トンネル → 自社ワークステーション → RAGが専門DBを検索 → クラウドAPI（Gemini/Opus）が根拠に基づき回答生成

バックアップ：冗長構成で自動フェイルオーバー

顧客はブラウザからアクセスするだけ。特別なソフトウェアのインストールやIT機器の購入は不要。



競合比較

	当事業者	一般的な競合（クラウドAI再販）
サービス範囲	6本柱・200種以上の機能	単機能チャットボット
専門DB	特許2,000万件・判例3万件・SDS	なし
メール自動化	41種類が24時間常駐稼働	なし or 単純転送のみ
AIモデル	用途別に最適モデルを選択（FT済み独自モデル＋Gemini/Opus）	汎用モデル1種のみ
課金体系	月額固定（使い放題）	従量課金（利用量で増加）
プロトタイプ開発	月1件まで含む（エンタープライズ）	対応不可
参入障壁	GPU + RAG技術 + 専門DB + 弁護士知識	APIキーで誰でも参入可能

当事業者 vs. 一般的な競合

特徴	当事業者	一般的な競合（クラウドAI再販）
サービス範囲	6本柱・200種以上の機能	単機能チャットボット
専門DB	特許2,000万件・判例3万件・SDS	なし
メール自動化	41種類が24時間常駐稼働	なし or 単純転送のみ
AIモデル	用途別に最適モデルを選択（FT済み独自モデル＋Gemini/Opus）	汎用モデル1種のみ
課金体系	月額固定（使い放題）	従量課金（利用量で増加）
プロトタイプ開発	月1件まで含む（エンタープライズ）	対応不可
参入障壁	GPU + RAG技術 + 専門DB + 弁護士知識	APIキーで誰でも参入可能

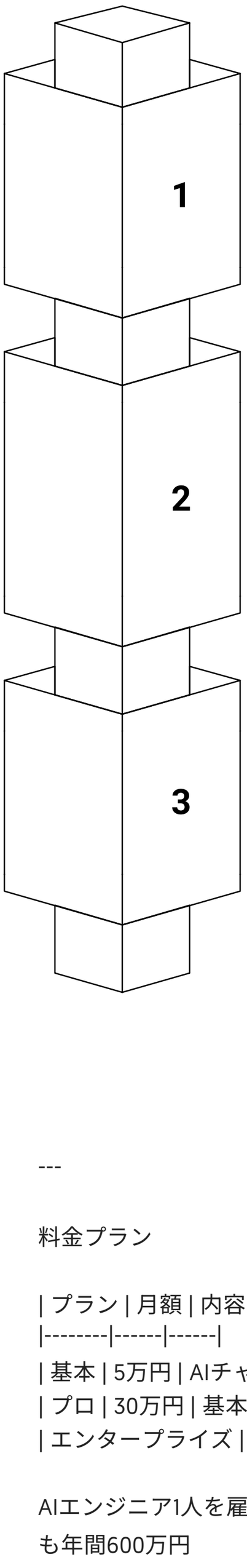
参入障壁

第1層：データ資産 — 特許2000万件・判例3万件のクレンジング済み専門DB。構築に数年かかる資産であり、後発者が短期間で追いつけない。

第2層：技術スタック — RAGシステム設計 + vLLM並列推論 + Nemotronファインチューニング + メール自動化41種 + Webアプリ100種 + Cloudflareセキュア公開。この複合技術を一人で構築・運用する能力。

第3層：専門知識 — 弁護士としての知的財産の専門知識がAIモデルの精度向上とサービス設計に直結。技術と業界知識の掛け合わせ。

AI SaaS事業の参入障壁



データ資産

特許2,000万件・判例3万件のクレンジング済み専門DB。構築に数年かかる資産であり、後発者が短期間で追いつけない。

技術スタック

RAGシステム設計 + vLLM並列推論 + Nemotronファインチューニング + メール自動化41種 + Webアプリ100種 + Cloudflareセキュア公開。この複合技術を一人で構築・運用する能力。

専門知識

弁護士としての知的財産の専門知識がAIモデルの精度向上とサービス設計に直結。技術と業界知識の掛け合わせ。

料金プラン

プラン	月額	内容
基本	5万円	AIチャット＋業務支援ツール100種以上
プロ	30万円	基本 + メール自動化41種 + 専門DB検索（特許・判例・SDS）+ 機密文書RAG
エンタープライズ	50万円	全機能 + 専用AIモデル構築 + 月1件プロトタイプ開発

AIエンジニア1人を雇用する場合の年間コスト1200万円-2400万円に対し、エンタープライズでも年間600万円

料金プラン

プラン	基本	プロ	エンタープライズ
月額	5万円	30万円	50万円
内容	AIチャット＋業務支援ツール100種以上	基本＋メール自動化41種＋専門DB検索（特許・判例・SDS）＋機密文書RAG	全機能＋専用AIモデル構築＋月1件プロトタイプ開発

その他：コンサルティングとして、RAGシステム一式の構築・譲渡にも対応。自社インフラ上で完結するAI環境を構築し、ナレッジトランスファー込みで納品。