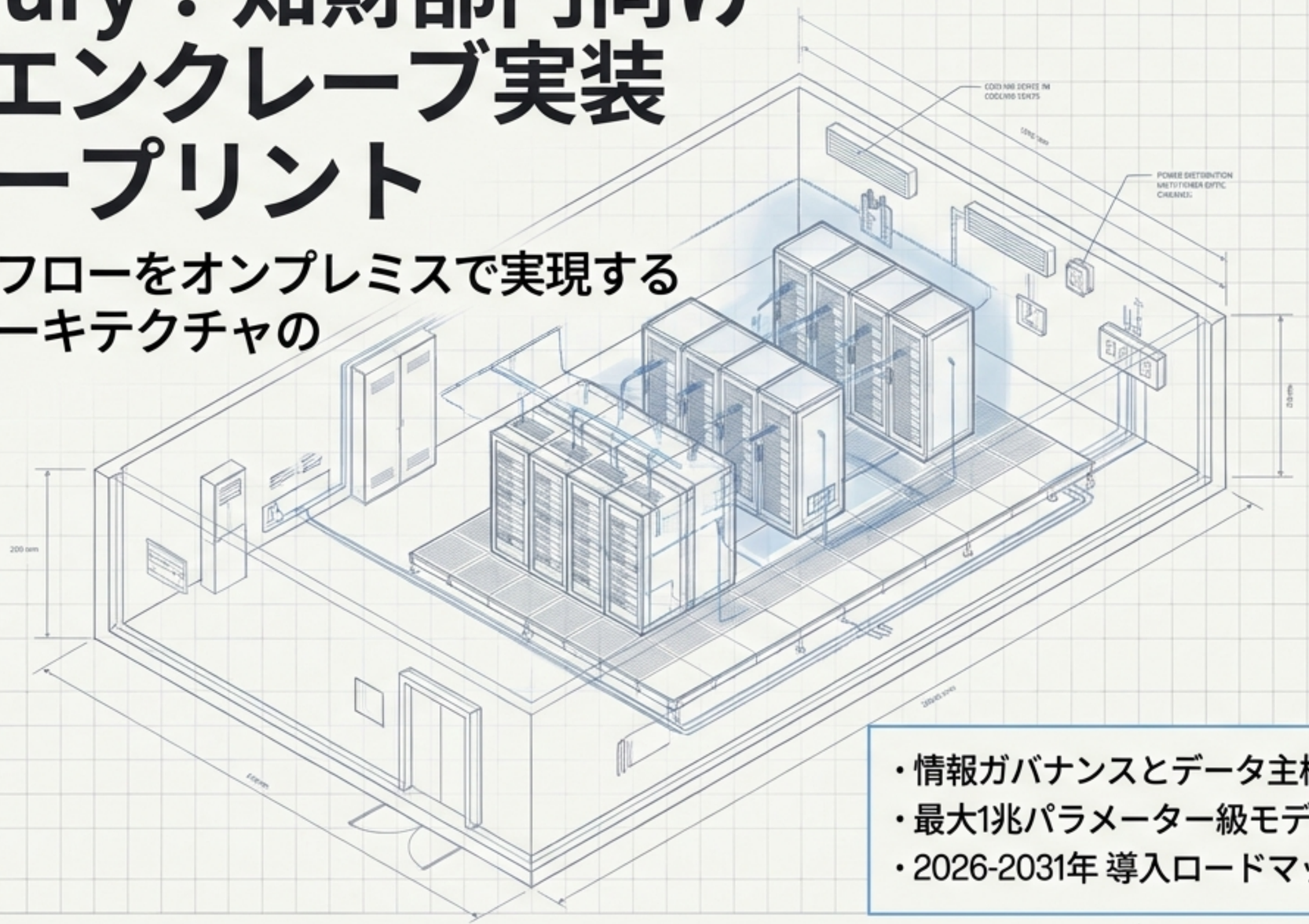
HP ZGX Fury：知財部門向け主権型AIエンクレーブ実装実装ブループリント

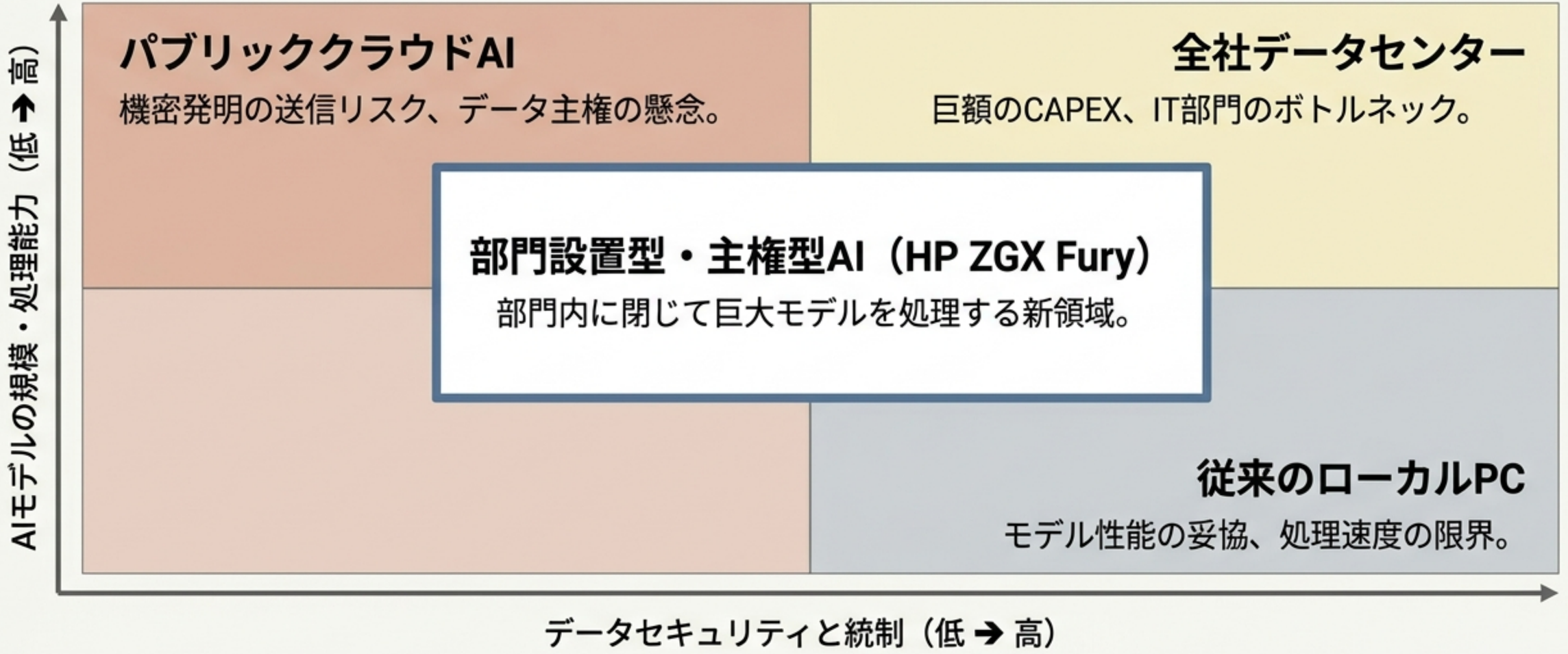
高度機密IPワークフローをオンプレミスで実現するハイブリッドAIアーキテクチャの全容と投資戦略



- 情報ガバナンスとデータ主権の確保
- 最大1兆パラメーター級モデルの実装
- 2026-2031年 導入ロードマップ

Gemini Notebook

知財AIのトリレンマを打破する「第四の選択肢」



知財業務における生成AI利用は、「安全性を妥協してクラウドを使うか」「性能を妥協してPCを使うか」の二者択一でした。ZGX Furyは、部門内でクラウド級の推論能力をセキュアに専有する「第四の選択肢」を提供します。

Gemini Notebook

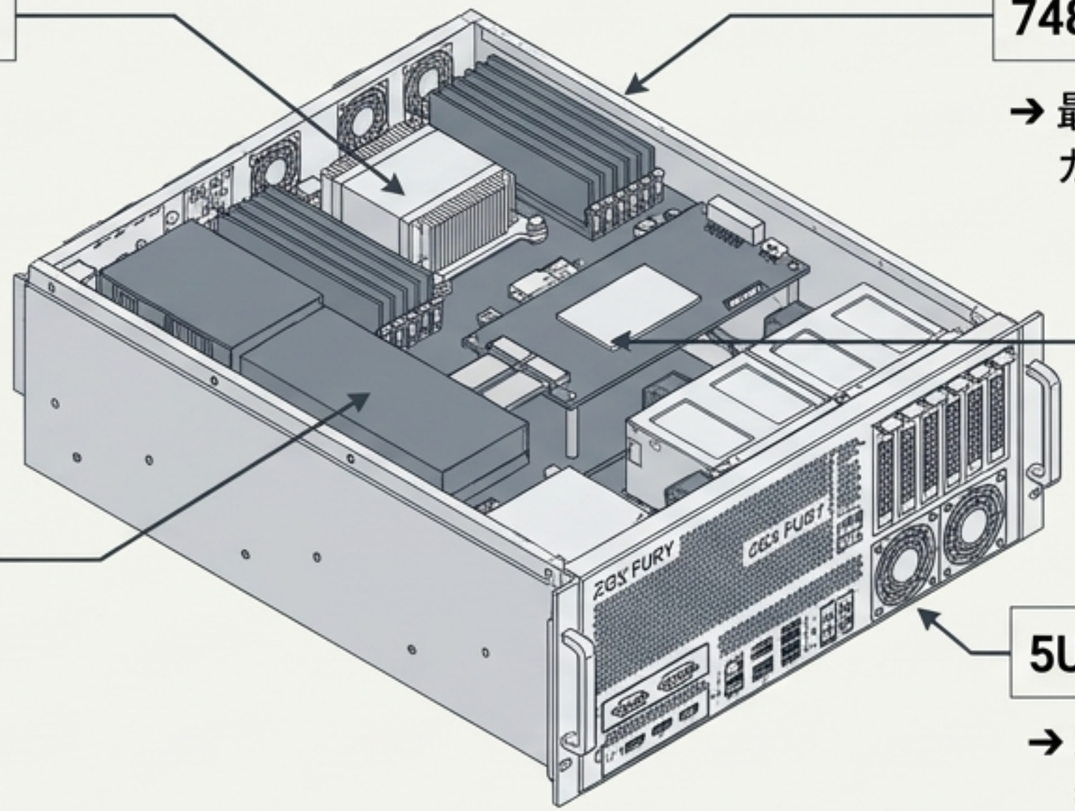
機密データの隣に置けるAIアプライアンス

NVIDIA Grace (72-core Arm)

➔ 大量特許文書の高速な前処理・ベクトル化・RAGインデックス構築。

748GB CPU-GPU コヒーレントメモリ

➔ 最大1兆パラメーター級モデルのローカル推論、1000億超のFine-Tuning。



NVLink-C2C (900GB/s)

➔ 巨大モデルの高速メモリ移動（従来のPCIe構成からの脱却）。

NVIDIA Blackwell Ultra (最大20 PFLOPS FP4)

➔ LLM、Embedding、Reranker、OCR後処理を1台で集約・同時実行。

5Uラック・1.6kW・液冷/空冷排熱

➔ 水冷設備不要で既存のサーバールームへ導入可能（希望小売価格2,849万円～）。

ZGXは「高性能PC」ではなく、データセンター技術の一部を1.6kW/5U級へ圧縮した「情報ガバナンス設備」です。

Gemini Notebook

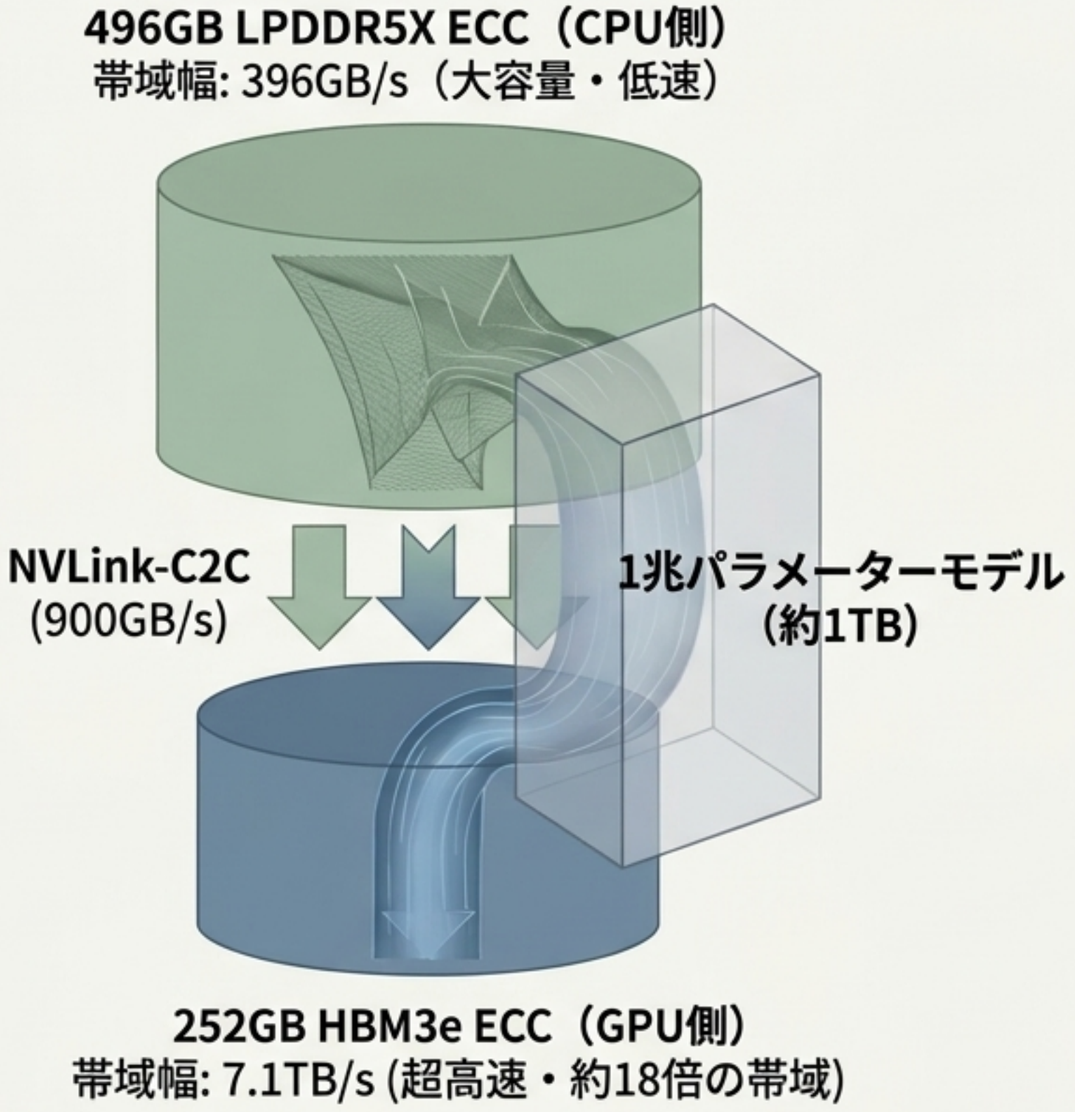
同世代データセンターAI（Blackwell Ultra）とのクラス比較

評価指標	HP ZGX Fury（本製品）	NVIDIA DGX B300	NVIDIA GB300 NVL72
GPU構成	Blackwell Ultra ×1	B300 ×8	Blackwell Ultra ×72
高速メモリ	748GB（コヒーレント）	2.3TB	37TB（Fast memory）
消費電力・設備	最大1.6kW （5U/既存環境可）	14.5kW （データセンター前提）	最大142kW （液冷AI Factory前提）
想定運用単位	知財部門・研究グループ	全社AI基盤	大規模データセンター
RAG・社内Fine-tune	◎ 極めて適合	○ 適合（過剰設備の懸念）	△ 部門用途には過大
同時利用・大量バッチ	○ 中規模	◎ 強い	◎ 圧倒的に強い

絶対性能（FP4）ではNVL72が圧倒的ですが、ZGXの真価は「特許調査が72倍速いこと」ではなく、受電容量や液冷CDUを要するAIデータセンター設備を新設せずに、自部門専用の共有AIノードを確保できる点にあります。

Gemini Notebook

容量と帯域のトレードオフ：1兆パラメーターモデル運用への現実解



【実務上の影響と推奨】

ZGXは748GBのメモリ空間により巨大モデルを「収容」できますが、帯域差（18倍）があるため、72GPUのNVL72と同等の処理スループット（多重度）は出ません。

知財用途のベストプラクティス：常時1兆モデルを動かすのではなく、70B〜数百B級のモデル＋Embedding＋Reranker＋RAGを1台へ並載し、複数担当者で「同時かつ高速に」共有する運用が最も投資対効果を高めます。

Gemini Notebook

TCO分析とクラウドAIとの損益分岐

3-Year Total: **¥45.0M - ¥92.0M**





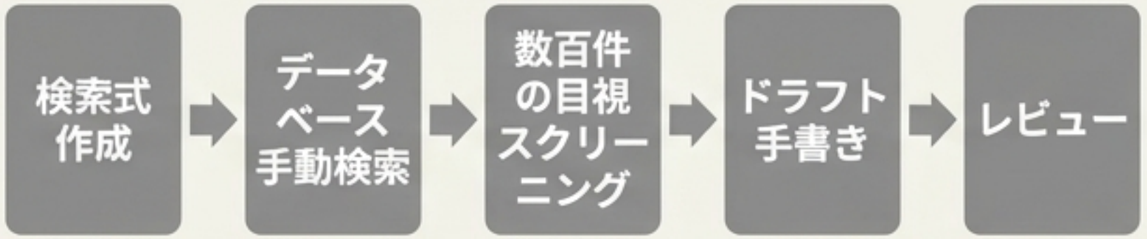
年間2,000時間超の定常利用でCAPEX（ZGX）が有利に

1. 最大コストは電力ではなく「AI統合と運用（人件費）」です。	2. 断続的な利用（年200〜500時間）のピーク負荷や大規模事前学習はクラウド（年間350〜1,020万円）へバーストさせる「ハイブリッド構成」が財務的に最適です。

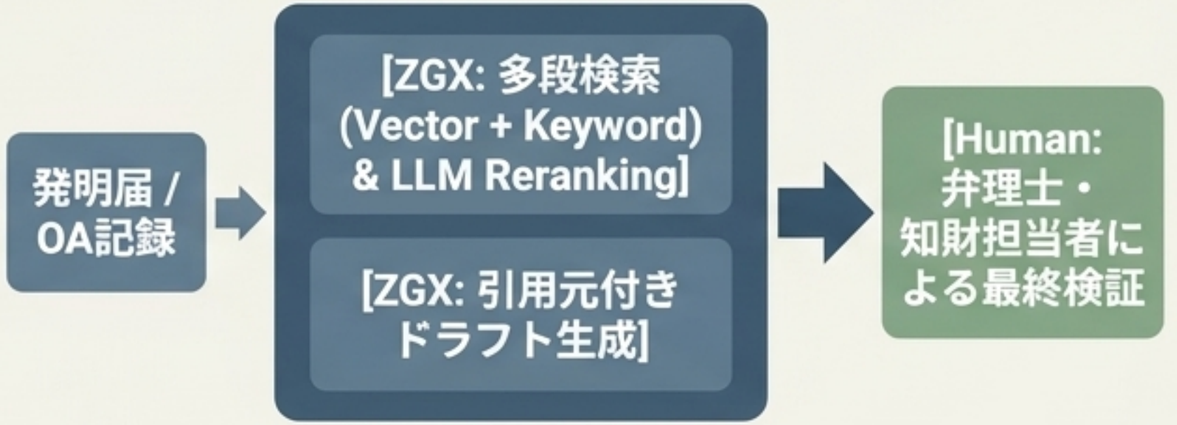
Gemini Notebook

知財ワークフローのAIネイティブ化（Human-in-the-Loop）

【従来プロセス：人がDBを一件ずつ操作】



【ZGXプロセス：AIが収集・ドラフトし、人が検証】

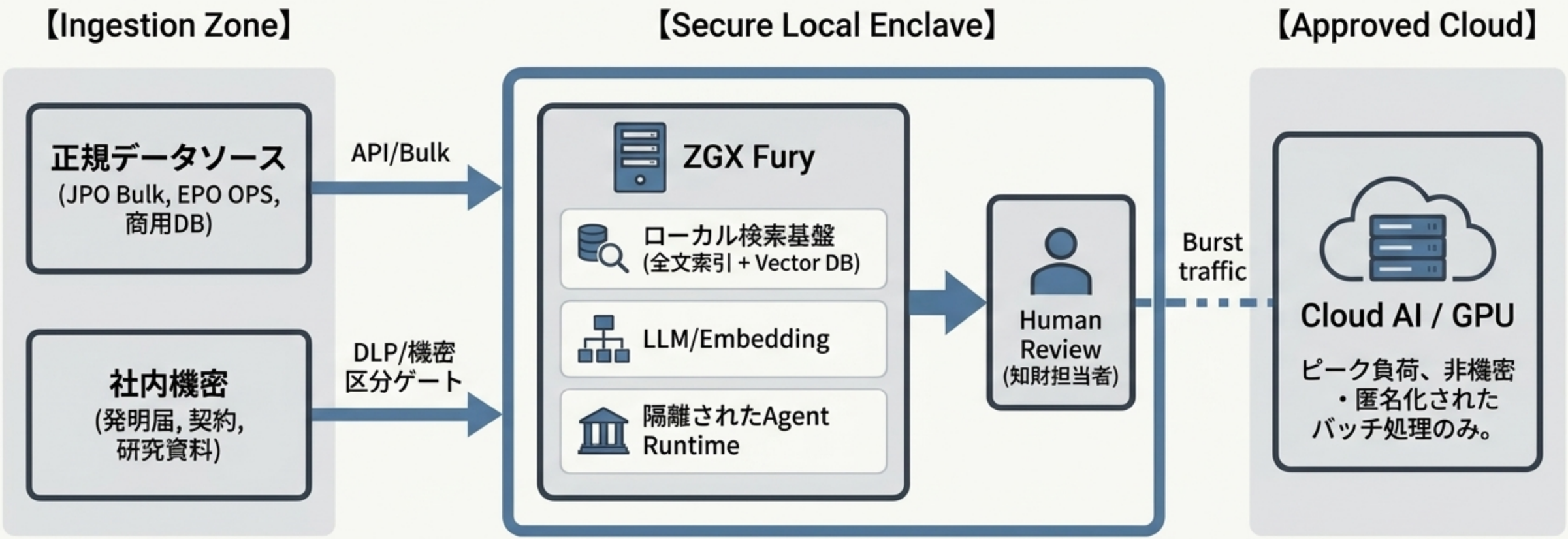


• **特例調査**：概念類似の捕捉（Recall向上）、長文読解の高速化。※最終判断は人間に依存。
• **契約レビュー**：NDA・MTAと社内Playbookの差分抽出。※Governing law等の誤認リスクに注意。
• **機密QA**：未公開出願・社内審査記録の閉域RAG。※API送信ゼロで完全保全。

Stanford RegLabの研究が示す通り、専門RAGでもHallucinationは残ります。
「完全自律化」ではなく「検証可能性の高度化」を目指します。

Gemini Notebook

統括：Confidential-by-Default ハイブリッドアーキテクチャ



最も機密性の高い「未公開発明・契約・研究ノート」をローカルの主権型エンクレーブ（ZGX）へ封じ込め、公開特許の大規模再処理やピーク負荷のみを承認済みクラウドへ逃がす最適解。

Gemini Notebook

データソースと法務コンプライアンスの境界

J-PlatPatからのスクレイピング
➡ INPIT規約違反リスク。AI学習目的の自動クエリは厳禁。

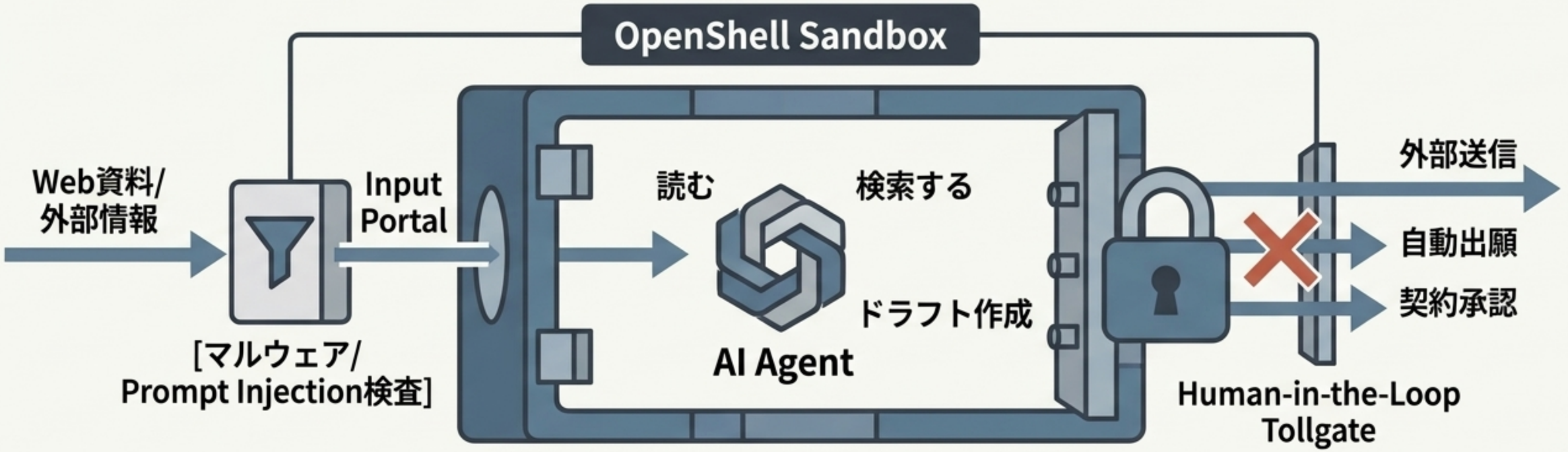


JPO正規バルクデータ / 契約済み商用DB / EPO OPS
➡ バルクダウンロードやREST API（週4GB等の制限遵守）による構築。

1.	2.	3.
営業秘密（経産省指針）	**著作権（文化庁ガイダンス）**	**個人情報（個情委）**
ZGXへの格納だけでは「秘密管理性」は満たされない。アクセス権（RBAC）、秘密表示、Need-to-Knowの継続管理が必須。	公開特許の明細書・図面にも著作権が存在。学習段階と生成・利用段階のリスクを切り離したRAG設計。	ローカル運用で第三者提供リスクは低下するが、利用目的の管理義務は存続。

Gemini Notebook

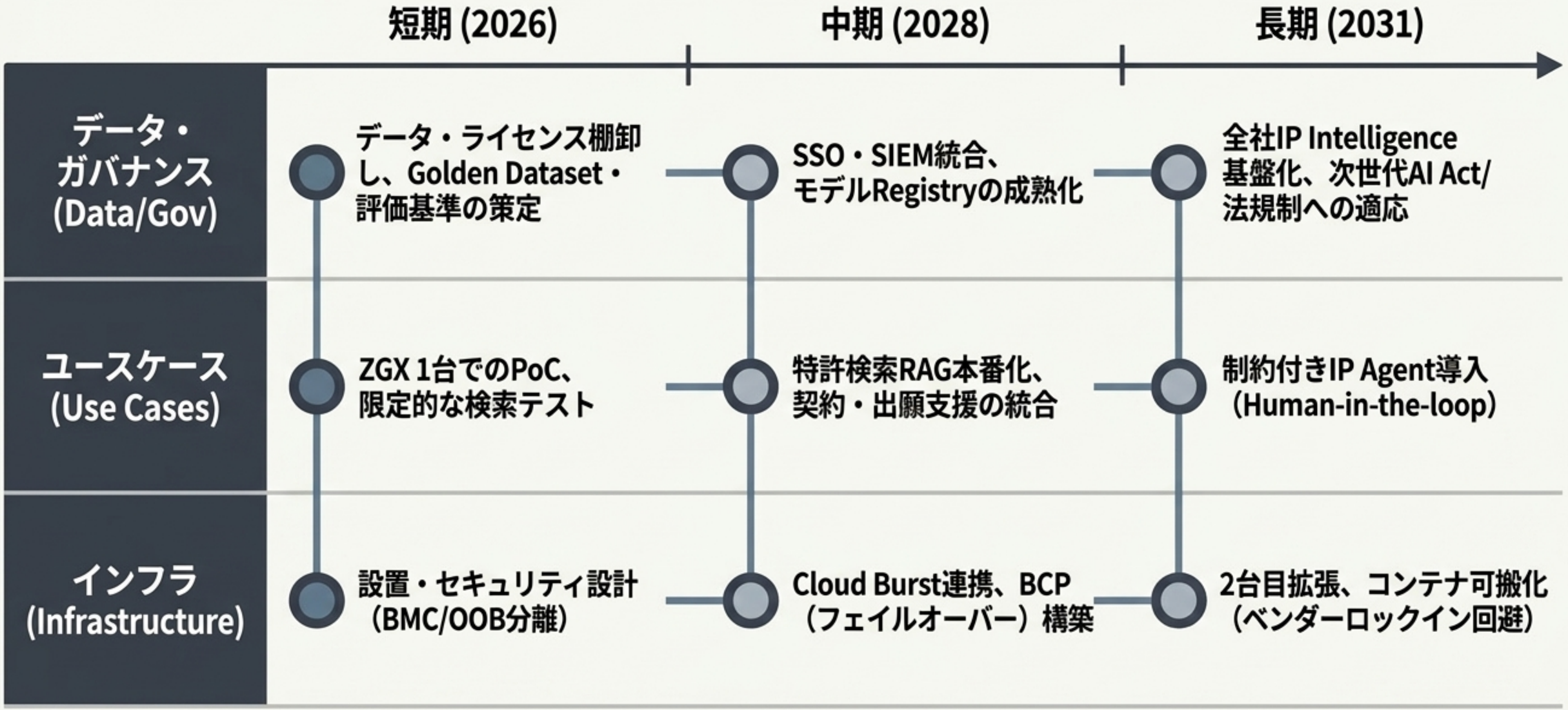
AIエージェントの隔離制御とセキュリティ（OWASP対策）



- **LLM Supply-Chain防御:** 外部モデル取得時は、署名・ハッシュ値・ライセンスを確認した上でStaging環境から昇格（Registry管理）。
- **間接的Prompt Injection防御（OWASP LLM01）:** 侵害調査等で外部Web証拠を取得する際、ZGXから直接Webアクセスさせず、外部収集専用環境からの一方向Importへ制限。
- **権限の最小化:** エージェントのTool-CallをAllowlist化。AIによる決定・送信を物理的にブロック。

Gemini Notebook

知財部門向けローカルAI導入ロードマップ (2026 - 2031)



Gemini Notebook

AIネイティブ知財組織の新たなKPIと結論

✕ 工数削減時間	○ 既知先行技術の再発見率・候補文献Recall
✕ AIの自動処理件数	○ Claim-to-Evidenceの根拠付与率・引用検証合格率
✕ クラウド利用料	○ 1レビュー案件当たりの総コスト・機密送信ゼロ件数

HP ZGX Furyは「単なる高速な特許検索用PC」ではありません。
最重要機密データへ、70B~1T級の高度なAIを“直接近づける”ことを可能にするインフラです。
勝負を分けるのは計算性能そのものではなく、「機密データを安全にAIへ接続し、その出力を検証可能な知財ワークフローへ組み込めるか」という情報ガバナンスの実行力にあります。

Gemini Notebook