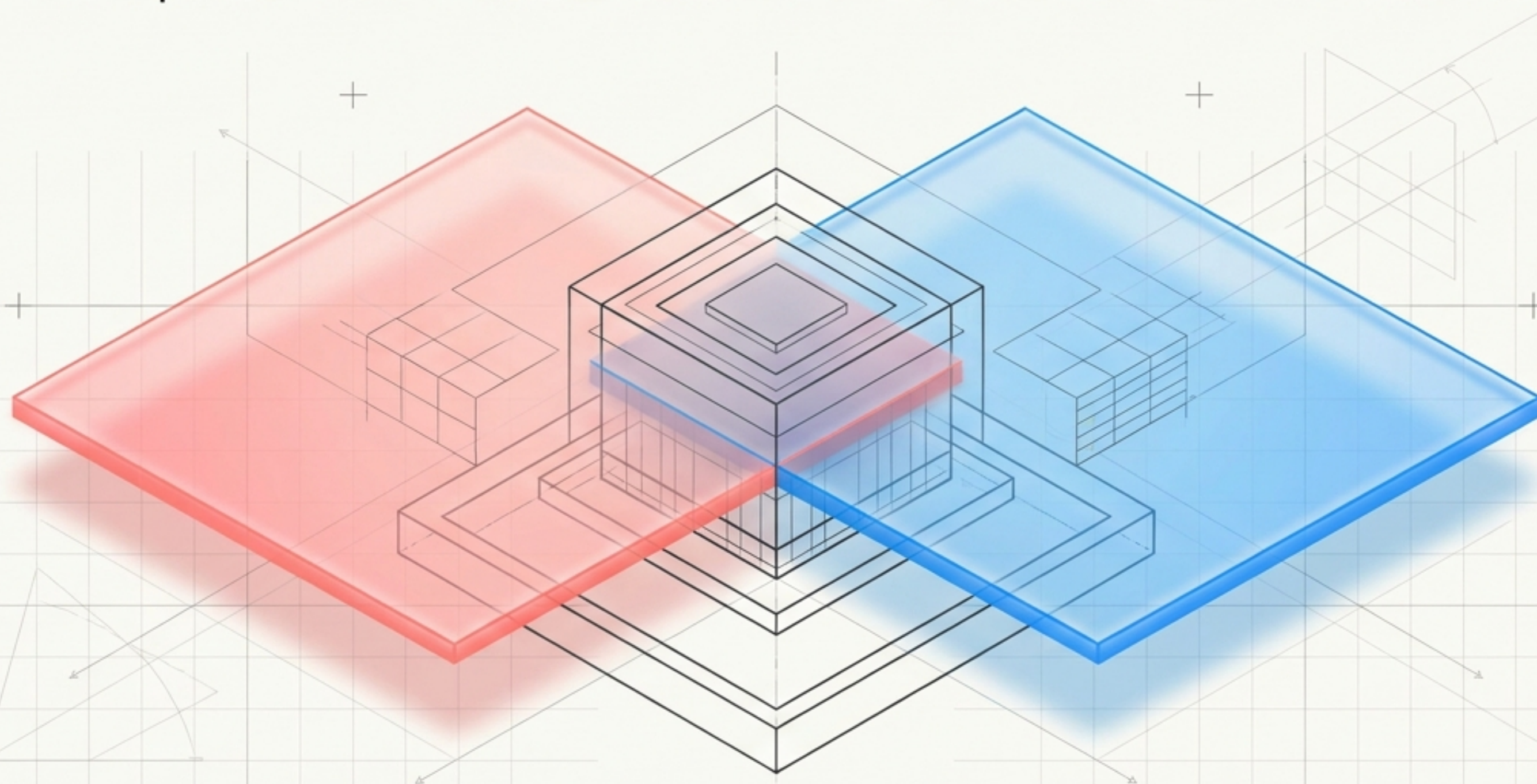
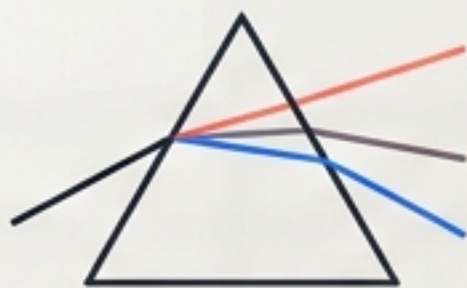


Claude Fable 5.1 / Mythos 5.1 徹底解剖レポート

Anthropic最新モデルの二層構造、客観的評価、および知財・法務実務への戦略的示唆



エンタープライズ導入における3つのパラダイムシフト



構造 (Structure)

単一ウェイト・二層セーフガード

基盤となるモデル能力は同一。
「一般向け (Fable)」と「審査制 (Mythos)」で適用される安全性網の目 (フィルター) のみを変化させる新アーキテクチャ。



経済性 (Economics)

エージェント型ワークロードの破壊的コスト減

基本単価は据え置きつつ、キャッシュ読み取りコストを75%削減 (\$1.00 → \$0.25)。長文脈・反復処理タスクにおいて最大45%のコスト削減を実現。

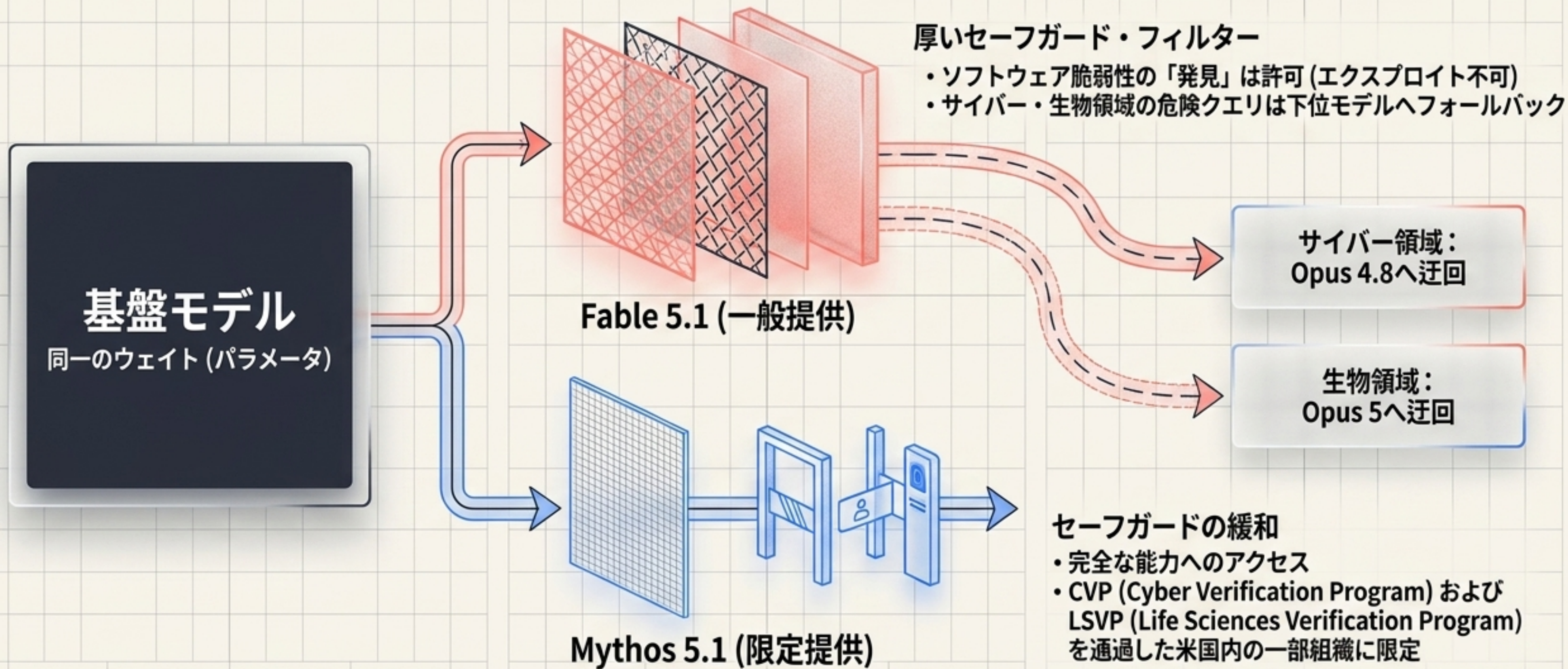


実務 (Application)

「合成レイヤー」としての再定義

知財・法務領域での網羅的分析力は向上。しかし、自律的完遂能力の低さと引用捏造リスクから、既存データベース上の「分析層」としての運用と人手検証が必須。

アーキテクチャの全貌：1つの頭脳、2つの異なるアプローチ



エージェント型タスクの経済性を根本から変えるプライシング

模似された 1M トークンアジニンググループ

Fable 5

Fable 5.1

基本トークン単価: 入力 \$10 / 出力 \$50

キャッシュ読み取り:
\$1.00

キャッシュ読み取り: \$1.00
→ \$0.25 (75%削減)

キャッシュ読み取り:
\$0.25

Fable 5

Fable 5.1

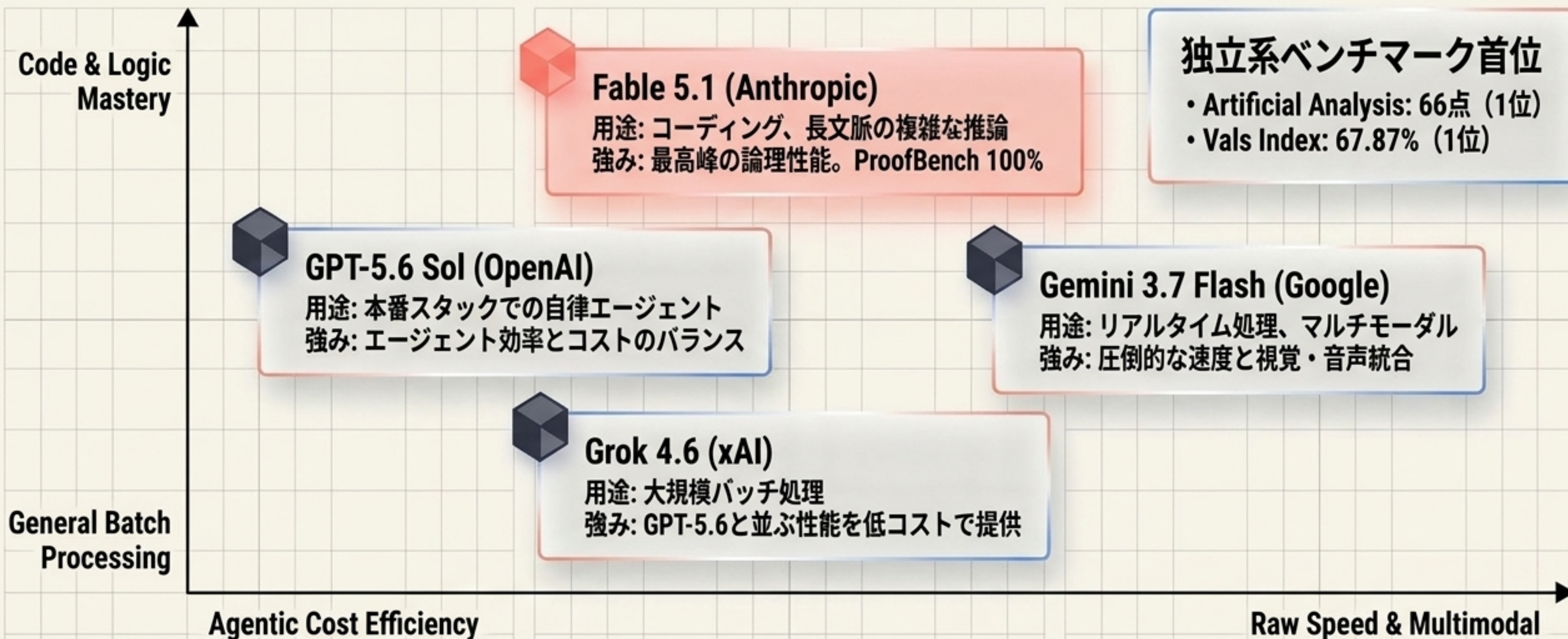
エージェント作業のコスト破壊

キャッシュ再読が頻繁に発生する「自律型エージェント」や「大規模ドキュメントの反復分析」において、最大約45%の総コスト削減。

エンタープライズ対応 (EFS)

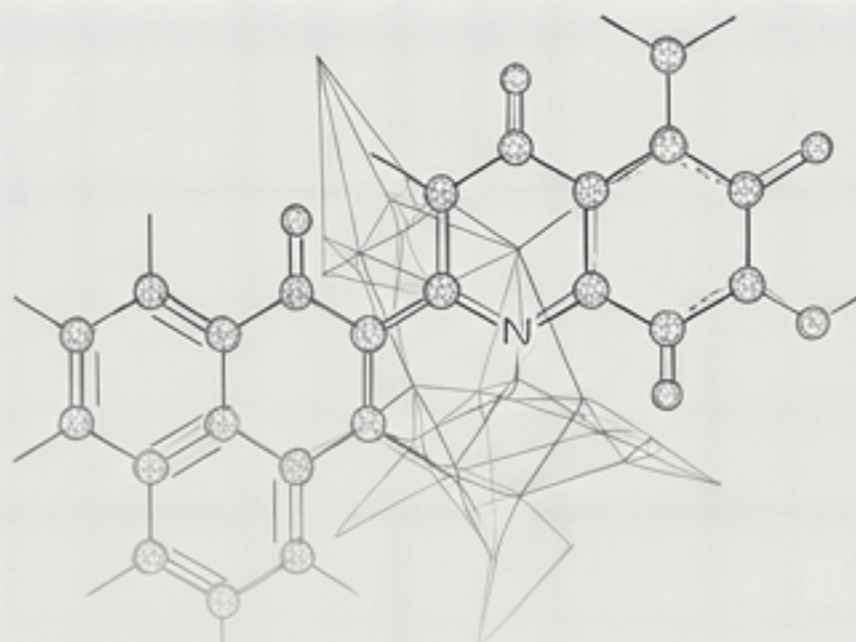
2026年秋より「Enterprise Frontier Safeguards (EFS)」を展開予定。顧客インフラ側でのデータ保持により、エンタープライズ要件である「ゼロデータ保持相当」の環境が構築可能に（現在はデフォルト30日保持）。

独立系評価と競合ランドスケープ：「絶対的な勝者」の不在



戦略的示唆: 用途による使い分け時代。知能指数の数点差よりも、APIのコスト差 (3対1) やタスク適合性が実務上のROIを決定する。

科学的成果の実証：知識労働からR&Dへの拡張



分子設計 (Mythos 5.1)

高親和性バインダーの設計

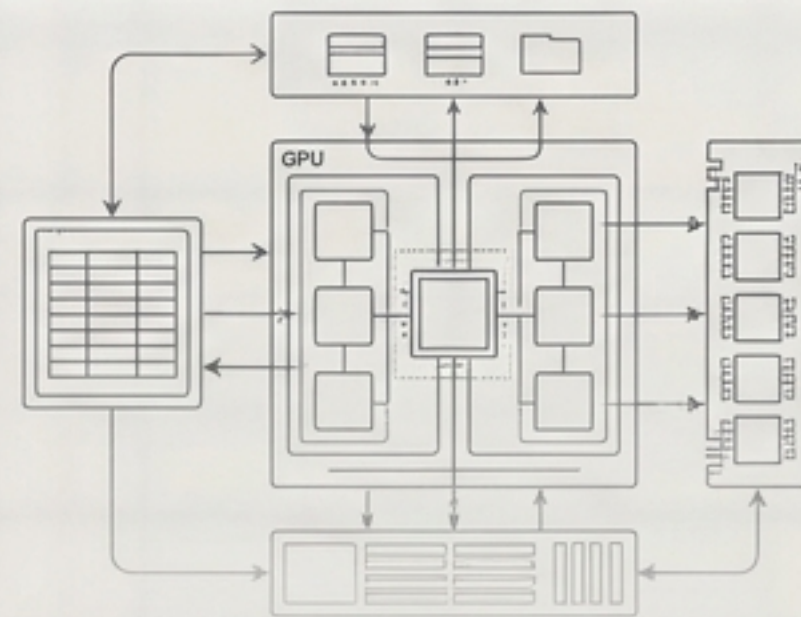
- 12の標的で約50%のヒット率を達成（通常は10～15%）。
- 特定標的では既存の最良設計と比較して10倍の結合親和性を記録。



宇宙科学・マッピング (Fable 5.1)

金星の高解像度標高マップ

- Magellan探査機のレーダー画像から学習。
- 惑星の約1/3の領域で、解像度を従来の10～20kmから2～3kmへと劇的に改善。



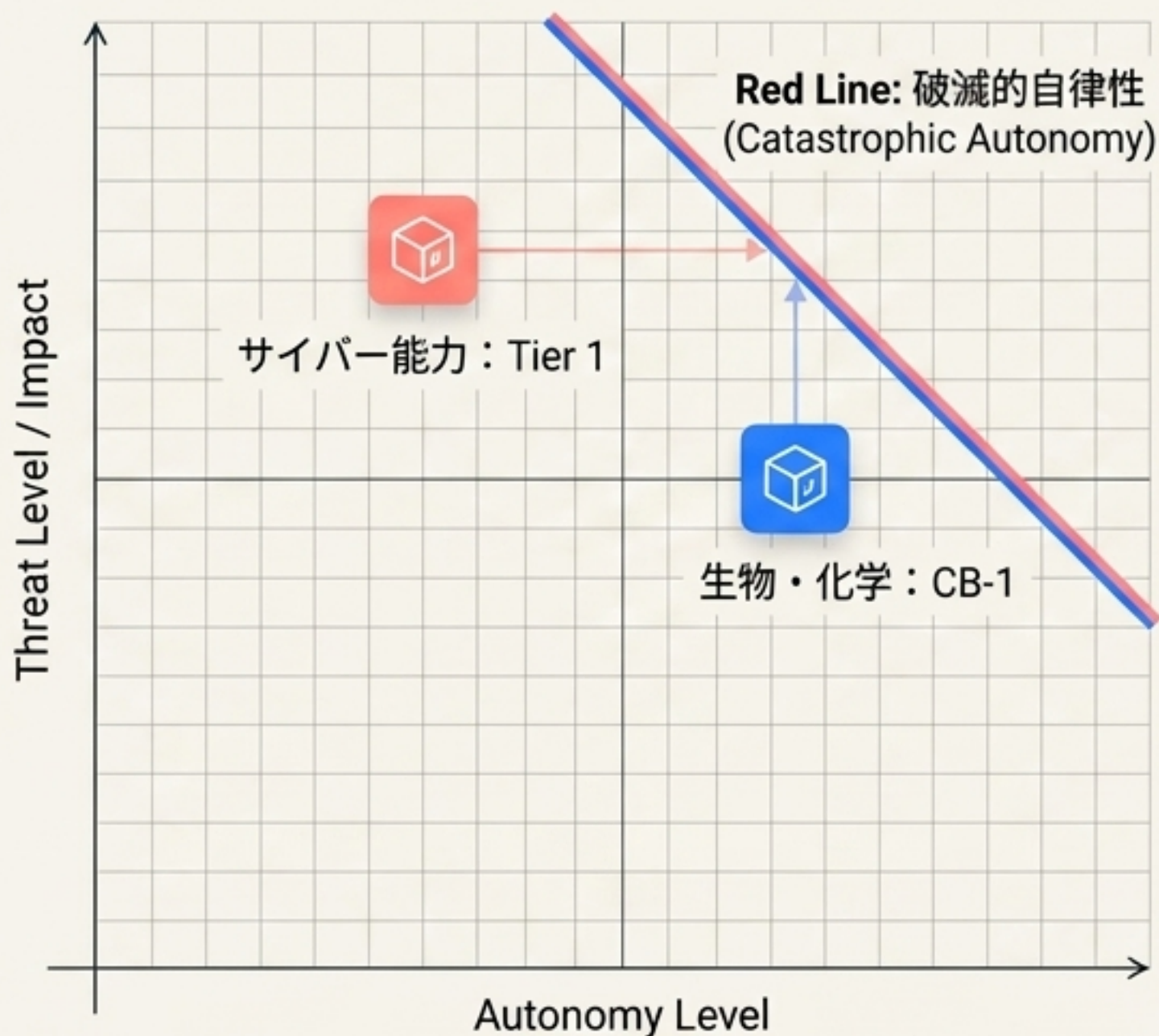
計算生物学 (Fable 5.1)

GPUコストの大幅削減

- オープンソース深層学習モデルにカスタムGPUカーネルを実装。
- 最大2.5倍の高速化を実現し、ゲノムワイド解析のGPUコストを30～60%削減。

System Card評価：脅威モデルの再定義とセーフガード

旧ASL (AI Safety Level) から「FCF (Frontier Compliance Framework)」に基づくティア制へ移行。



サイバー能力：Tier 1

Anthropic史上最強の能力。Mythos 5.1はOSS-Fuzzで17ターゲット満点、Firefox 147で98.0%のフルエクスプロイト達成。ただし、完全自律型の新規攻撃（Tier 2）には未到達。

生物・化学：CB-1

基礎的背景を持つ者の兵器合成を支援し得るが、新規開発の律速となる「希少専門家の代替（CB-2）」には未到達。

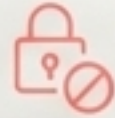


プロンプトインジェクション耐性

Gray Swan IPI評価において、攻撃成功率はわずか0.1% (k=1)。他社モデル (Gemini 3.7 Flashの9.2%) を大きく引き離す堅牢性。

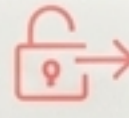
地政学リスクとコンプライアンス：供給断と規制の波

2026年6月12日:
米国輸出管理指令



米商務省BISが外国籍者のアクセス停止を命じ、Mythos等へのアクセスが全面停止（供給断リスクの顕在化）。

2026年6月30日:
規制解除・再開



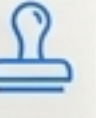
アクセス部分解禁を経て、全面解除・世界提供再開。「場当たり的な規制はオープンソース勢を利する」との批判。

2026年7月27日:
EU AI Act対応完了



AnthropicがEU Code of Practiceに署名。

2026年8月2日以降:
ウォーターマーク義務化



リリースされる全モデルに「統計的テキストウォーターマーク（C2PA）」の付与が義務化。

June

Summer 2026

July

August



エンタープライズへの影響

国防総省による「サプライチェーンリスク」指定の報道もあり、米国方針の転換による一時的なAPI遮断リスクは依然として残存する。

評価に対する反発：「権力の集中」と見えない摩擦

セーフガードの副作用 AI権力の寡占化



Mythosのアクセスを一部の審査済み組織（CVP/LSVP）に限定する方針は、悪用を防ぐ一方で「AIパワーのシリコンバレー数社への集中」と「防衛側の底上げ失敗」を招くとの批判。

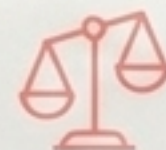
透明性と離反 ウォーターマークへの反発



オフにできない不可視ウォーターマークの義務化。コード領域では効果が薄いという盲点に加え、プライバシーを重視するユーザーのローカル（オープンソース）モデルへの流出要因に。

The Claude 5.1
Ecosystem

評価の客観性 自社ベンチマークへの依存



強力な科学的成果はAnthropicがキュレートしたものであり、完全に独立した評価ではない。また、セーフガード介入を「0点」とする評価手法により、本来の能力が歪められているとの指摘。

知財・法務実務への適用：強力なドラフト能力と露呈した弱点

Legal & IP
Workflow

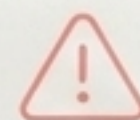
強み: 分析・ドラフト (BigLaw Bench: 93.4%)

- 1Mトークンの長文脈と図表読解を活かした、明細書・先行技術文献の横断分析。
- IPランドスケープの一次ドラフト作成。
- 低effort設定でも網羅的输出が可能 (レビュー工数削減に直結)。



弱点: 全工程完遂と誤検知 (Legal Agent Benchmark: 6.67%)

- 自律的完遂の限界：法務の「全工程をこなす能力」は依然として極めて低水準 (Vals AI評価で6.67%)。
- 安全フィルターの誤検知：医薬・バイオ (CBRN) 関連特許の無害な質問に対し、セーフガードが過剰反応しOpusへフォールバック (回答拒否) する事例が多発。



「合成レイヤー」コンセプト：正しいアーキテクチャ設計

LLMを「知識庫」として扱うアプローチは、引用捏造リスクにより破綻する。
既存の信頼できるシステム上に、高度な推論エンジンとしてFable 5.1を配置せよ。

中層：合成・分析レイヤー (Claude Fable 5.1)

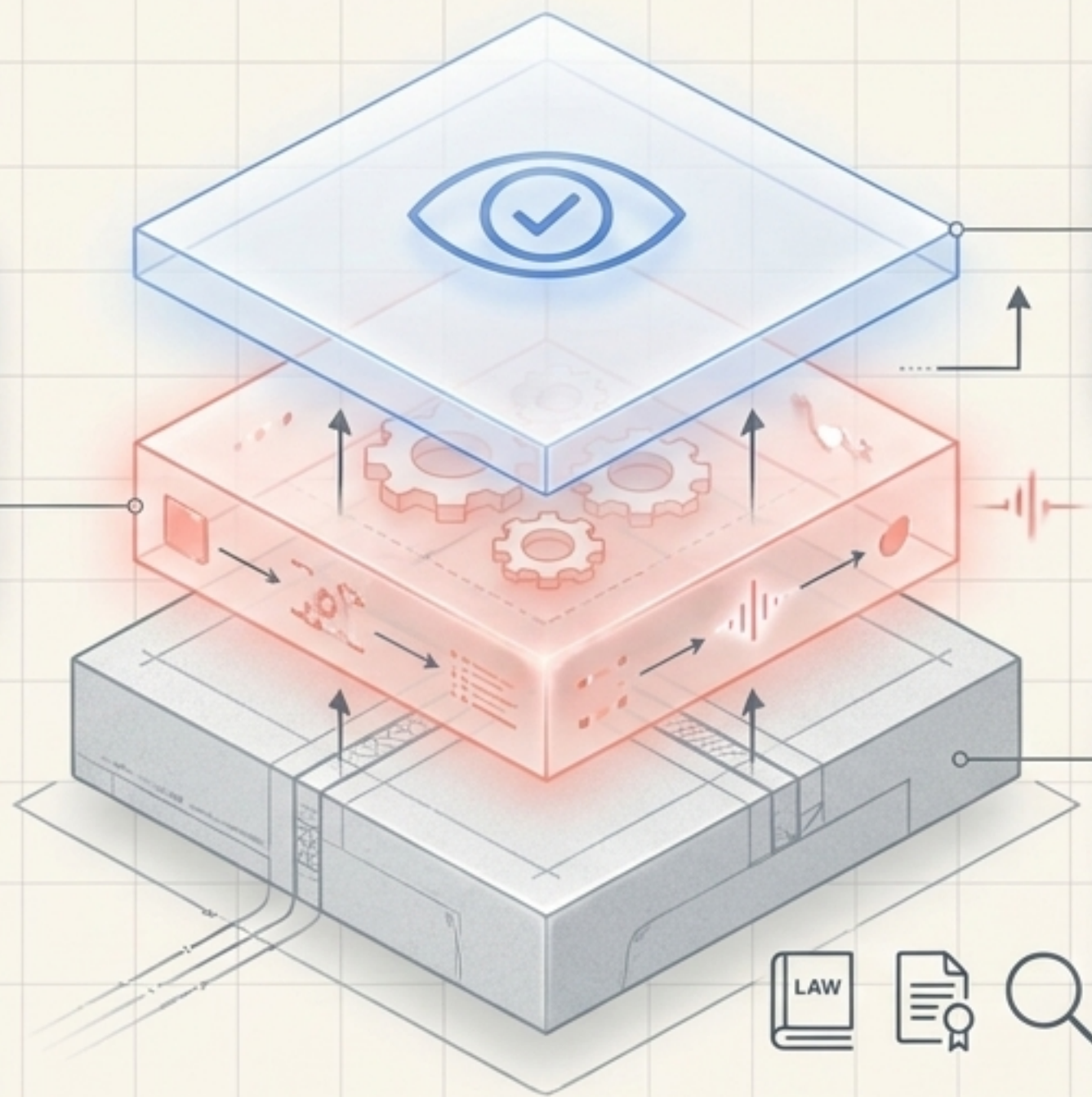
高次元の推論・ドラフト生成。ライブの法令や特許の実在性確認は行わず、下層から提供された事実のみを処理する。

上層：検証レイヤー (Human-in-the-Loop)

人間の弁理士・弁護士による最終判断と出典検証。現時点では完全自動化を避け、補助ツールとして設計する。

下層：事実データベース (Trusted DBs)

Westlaw, Espacenet, Google Patents等の商用・公的データベース。事実情報の唯一の情報源 (Single Source of Truth)。



エンタープライズ導入の指針と監視すべき閾値

導入ロードマップ

【短期: Fable 5.1への移行検証】

- キャッシュ依存度が高いエージェント作業、大規模リファクタリングを移行（コスト最大45%減）。
- 移行時の必須対応：強制ツール使用の除去、履歴追記型運用への変更、effortの再チューニング。

※Opus 5で到達可能なタスクは、安価なOpus 5を継続。



【中期: 本番環境の構築（2026年秋～）】

- EFS (Enterprise Frontier Safeguards) 展開を待ち、ゼロデータ保持環境での本番投入。
- 機密情報に対するウォーターマークの影響評価。

再評価すべき4つの監視トリガー

- ☐ 1. 独立系スコアの逆転：Artificial Analysis等の評価で競合モデルが上回った場合。
- ☐ 2. [Mythos](#)アクセスの緩和：LSVP/CVPプログラムが国際的・広域に開放された場合。
- ☐ 3. ウォーターマーク規制の動向：EU AI Actの検知API展開と規制当局の判断。
- ☐ 4. 米国輸出管理のリスク：供給断（API遮断）につながる地政学的措置の再発。