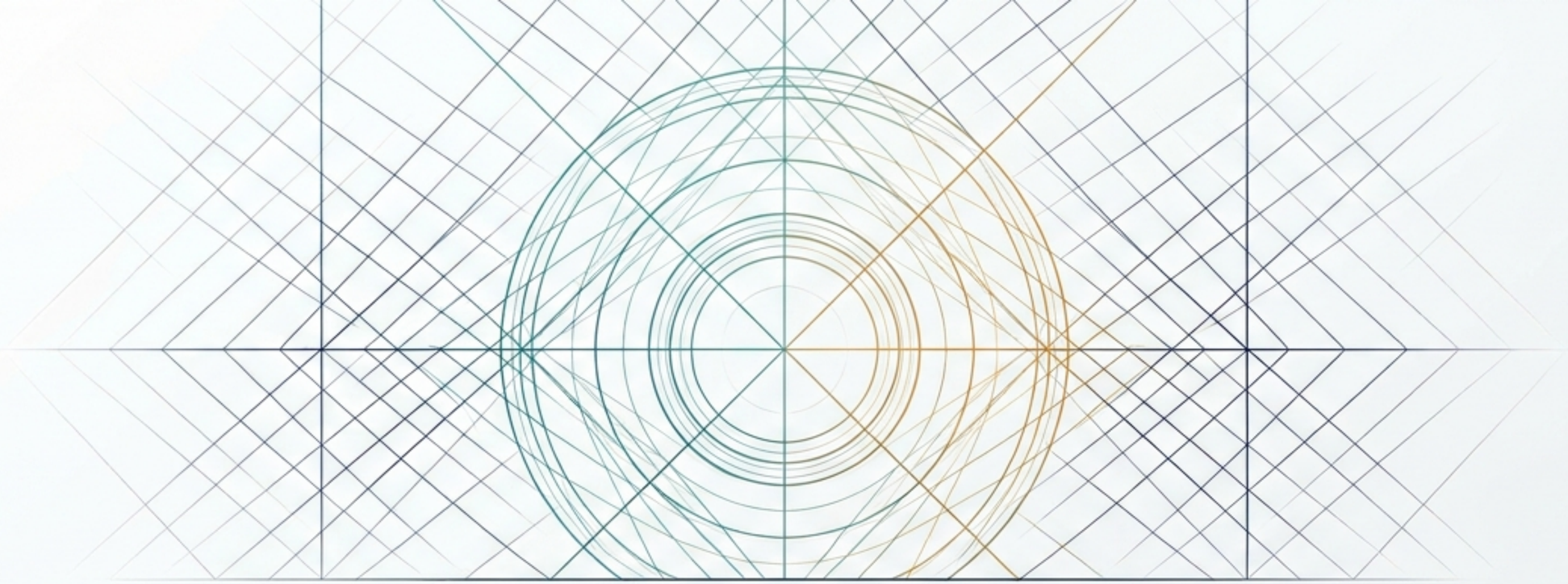


Qwen4-Max 導入評価ブリーフィング：期待と現実の境界線

企業利用における性能・コスト・法的リスクの総合診断
（2026年11月16日 正式リリース対応版）

IT戦略・AI導入・知財法務エグゼクティブ向け資料

Gemini Notebook

Qwen4-Maxは確実な進化を遂げたが、市場全体を凌駕する「万能薬」ではない



性能の進化（Evolution）

タスク当たり費用 約67%削減（前世代比）

2.4兆パラメータの巨大モデルでありながら、推論効率の最適化により長時間の専門業務における実用性が大幅に向上。



市場ポジション（Positioning）

Intelligence Index 43 / タスク費用 $3.27

同価格帯・同性能帯にGPT-6.1 Sol等の強力な競合が存在。全てのベンチマークで最先端の米国製モデルを上回ったわけではない。



導入リスク（Governance）

デフォルトでの30日間データ保持

Zero Data Retention（ZDR）は自動適用されず、東京リージョン選択時も中国本土ノードへルーティングされる構造的リスクが存在する。

Gemini Notebook

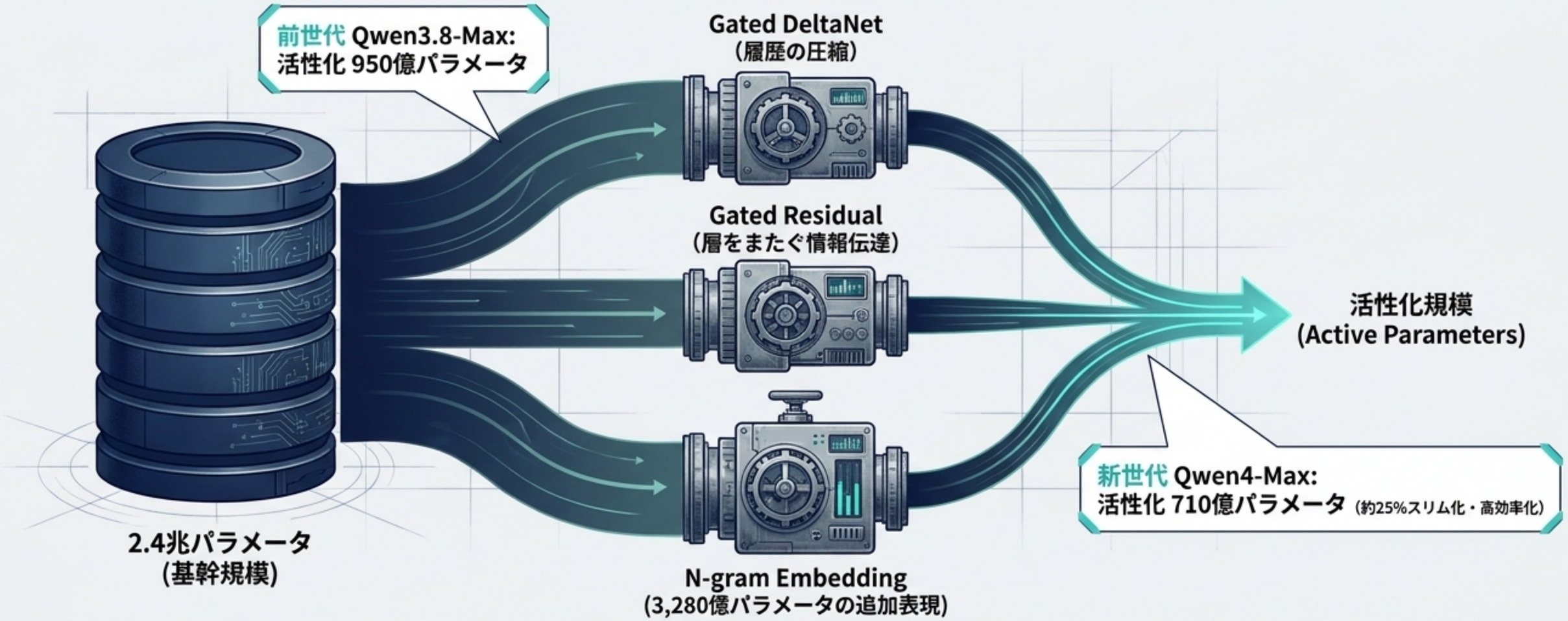
プレビュー版での検証結果と、正式版（11月16日公開）の仕様は厳密に区別される

項目（Parameter）	qwen4-max-preview（11/11公開）	qwen4-max 正式版（11/16公開）
API課金形態	Token Plan限定	従量課金API ＋ Token Plan
推論の標準設定	xhigh	high ⚠
提供される推論強度	low / medium / xhigh	low / high / xhigh
思考と回答の合計上限	393,216 トークン （QwenCloud表示）	262,144 トークン （詳細仕様書ベース）

プレビュー版の実使用レビュー（17分でのLP制作など）は、標準設定『xhigh』での結果を含むため、正式版（標準『high』の評価としてそのまま採用することはできない。

Gemini Notebook

計算リソースを「絞り込む」ことで、同規模の基幹モデルよりも高い推論効率を実現



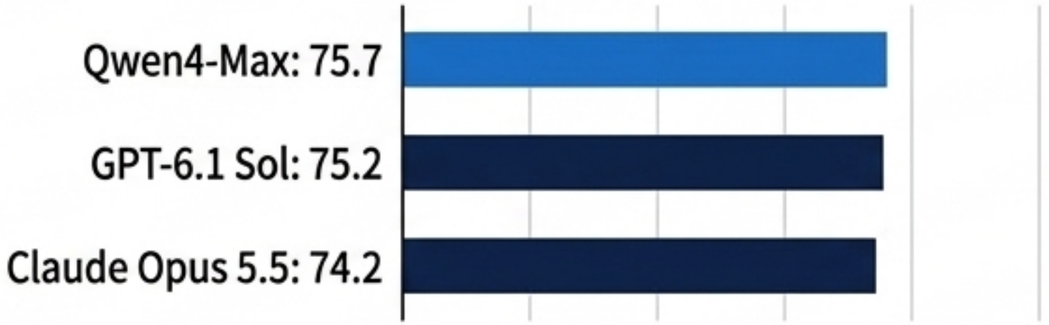
総パラメータ数を増やすのではなく、1トークン処理時に稼働する『アクティブな歯車』を25%削減しつつ精度を向上させた。

Gemini Notebook

メーカーの公式発表データは「特定の条件下における優位性」を示すに留まる

メーカーの主張（公式発表データ）

DeepSWE v1.1（ソフトウェア修正）



BioMysteryBench

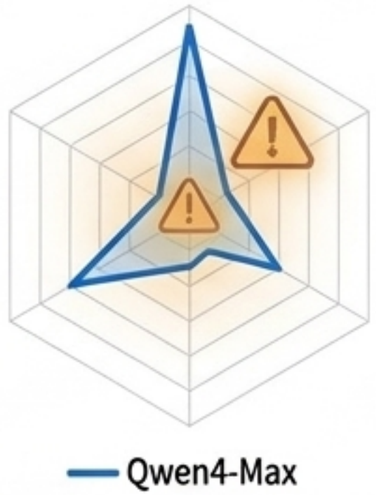


長時間の専門業務や一部開発タスクでOpus超え

第三者評価に基づく現実（コンテキストの補完）

Reality 1:
比較図には Claude Fable 5.2 や Gemini 4 Carbon が含まれていない。最先端全体を凌駕したわけではない。

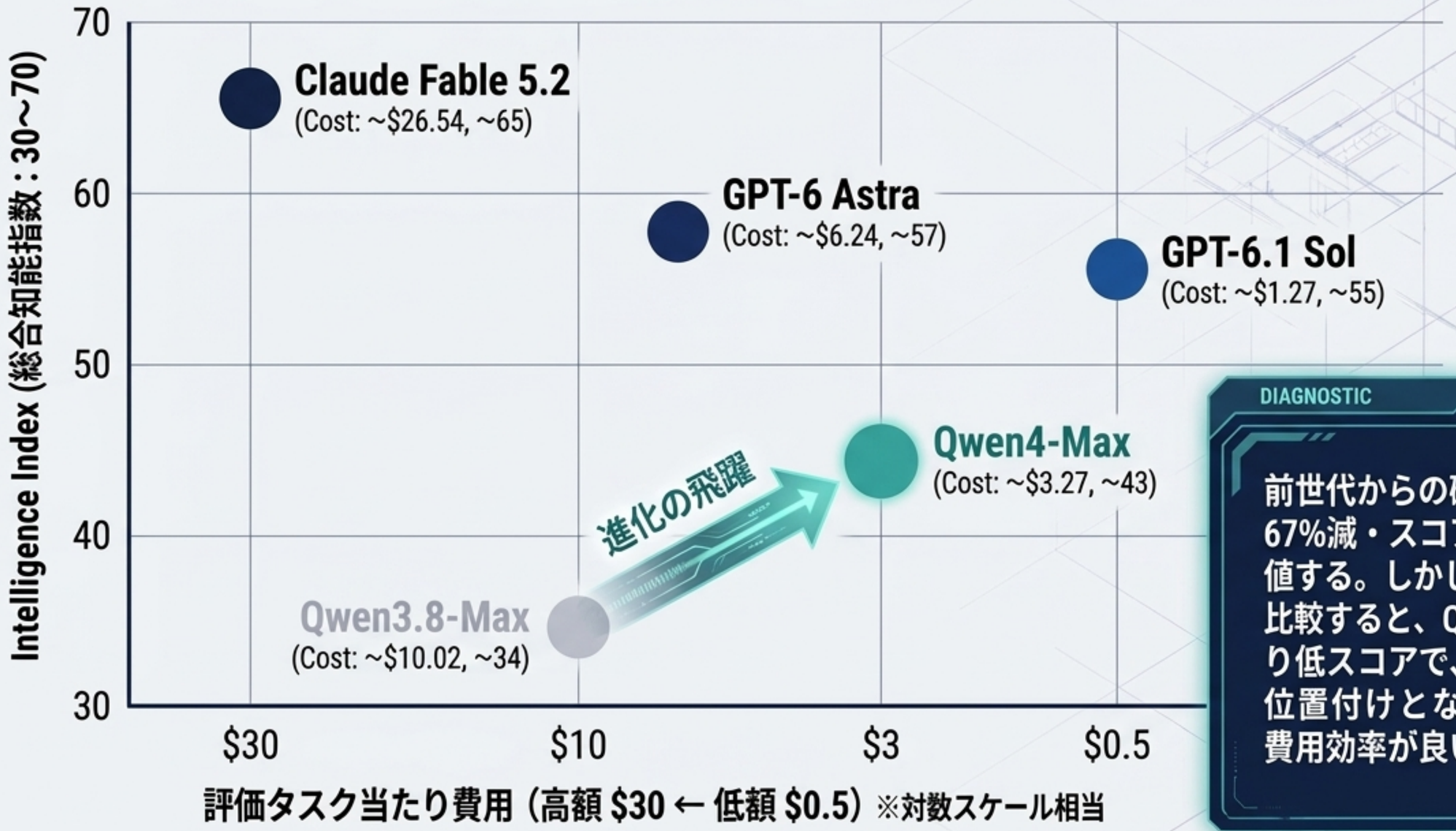
Reality 2:
ProgramBench (31.5) や Chartography (54.2) では競合に大差で劣後する領域が明確に存在する。



Reality 3:
実演された事例（48時間の冷却液分子探索など）は、外部ツールや実行環境を組み合わせた結果であり、API単体で完結するものではない。

Gemini Notebook

AIモデル・ポジショニング分析：絶対的な費用対効果では他モデルが優位に立つ

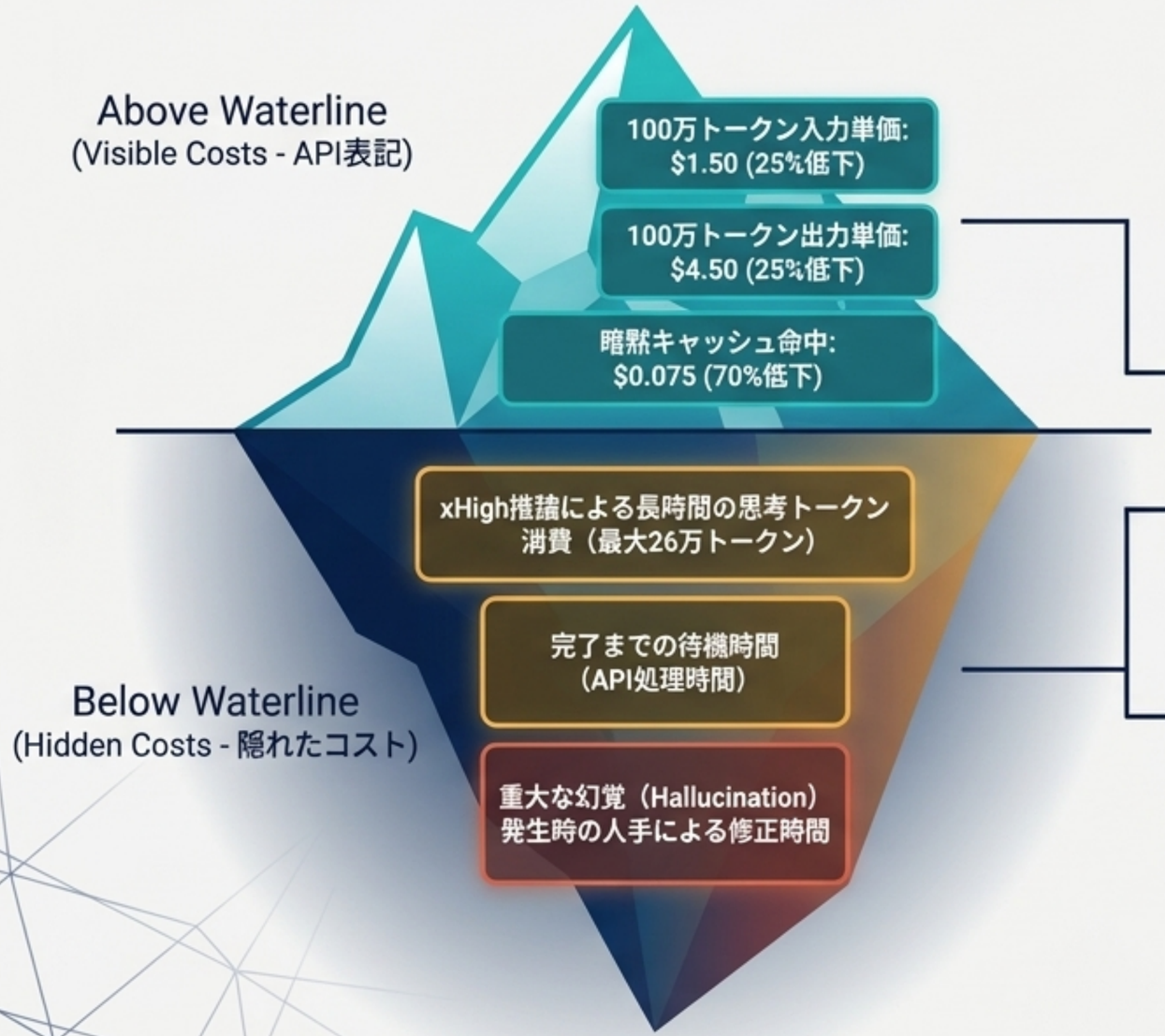


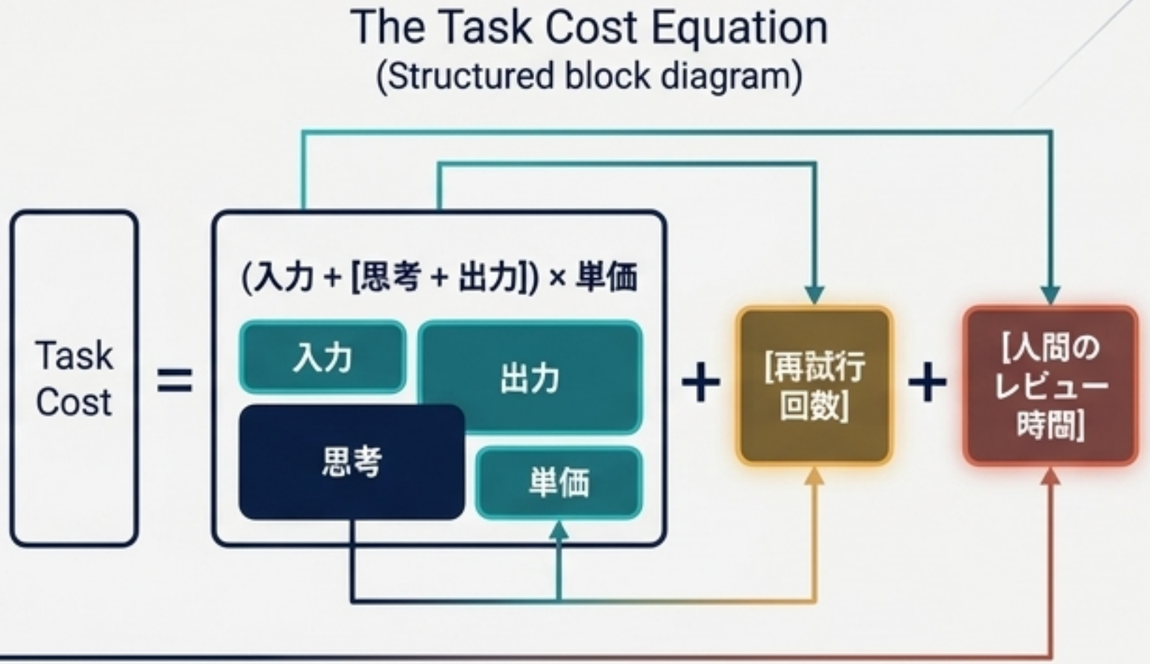
DIAGNOSTIC

前世代からの確実な飛躍（費用67%減・スコア向上）は評価に値する。しかし、GPT-6.1 Solと比較すると、Qwen4-Maxは『より低スコアで、り高コスト』な位置付けとなる。市場で最も費用効率が良いわけではない。

Gemini Notebook

タスク総費用の罠：「API単価の安さ＝業務コスト削減」ではない



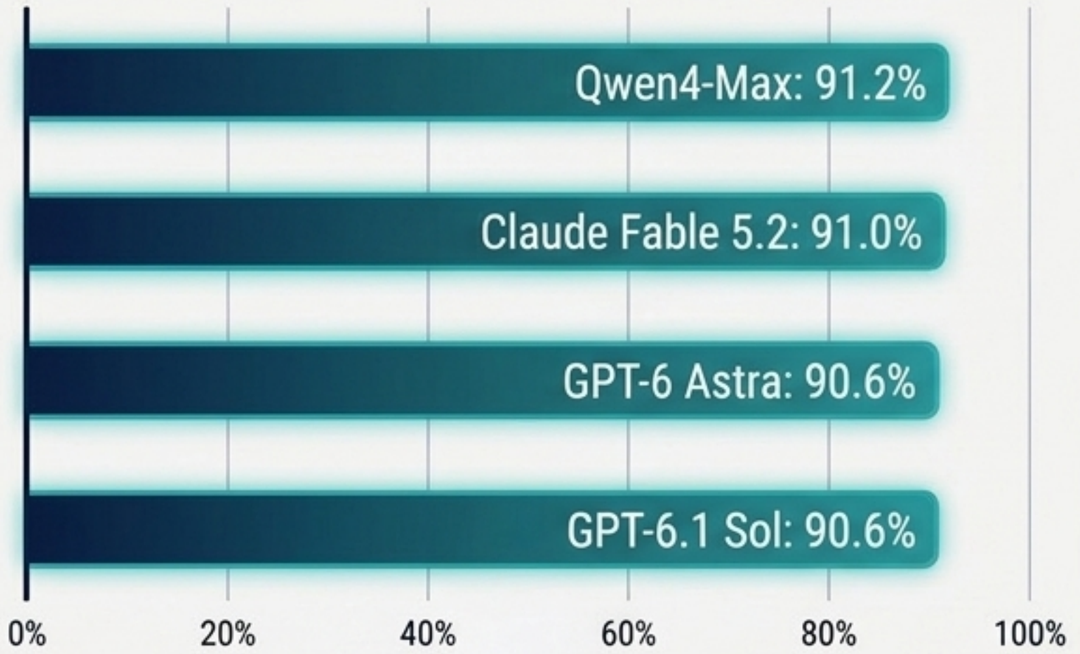


出力単価が安くても、推論思考プロセスが長引けばAPI費用は増大する。さらに手戻りが発生すれば、タスク全体のコストは競合モデルを上回る可能性がある。

Gemini Notebook

法務・知財ベンチマーク：高い合格率の裏に潜む「重大な幻覚」のリスク

個別項目合格率
(Bar Chart)



課題当たりの重大な幻覚数
(Warning Gauge / Dial indicators)



Harvey LAB-AA指標の読み解き: 全項目合格・重大幻覚なしの主指標は9.4%。複合的な法務課題の処理能力は高いが、致命的な幻覚（引用間違い等）の発生率がGPT-6シリーズに比べて極めて高い。順位が高いからといって、無確認で法的結論を採用することはできない。

Gemini Notebook

知財実務への適用性診断：得意領域の活用とリスク回避の切り分け

有力・活用推奨

公報分類とSDI一次評価:
100万トークンの文脈長が活きる。見逃し率と根拠段落の抽出精度を要検証。

中国語特許と技術資料の翻訳:
専門用語の訳し分けにおいて比較評価する価値大。

IPランドスケープ資料作成:
成果物作成能力 (Excel/PPT整合) との相性が良い。

条件付き適用・要検証

複数公報の構成要件対比:
長文処理は可能だが、類似する実施例と請求項の混同リスクに注意。

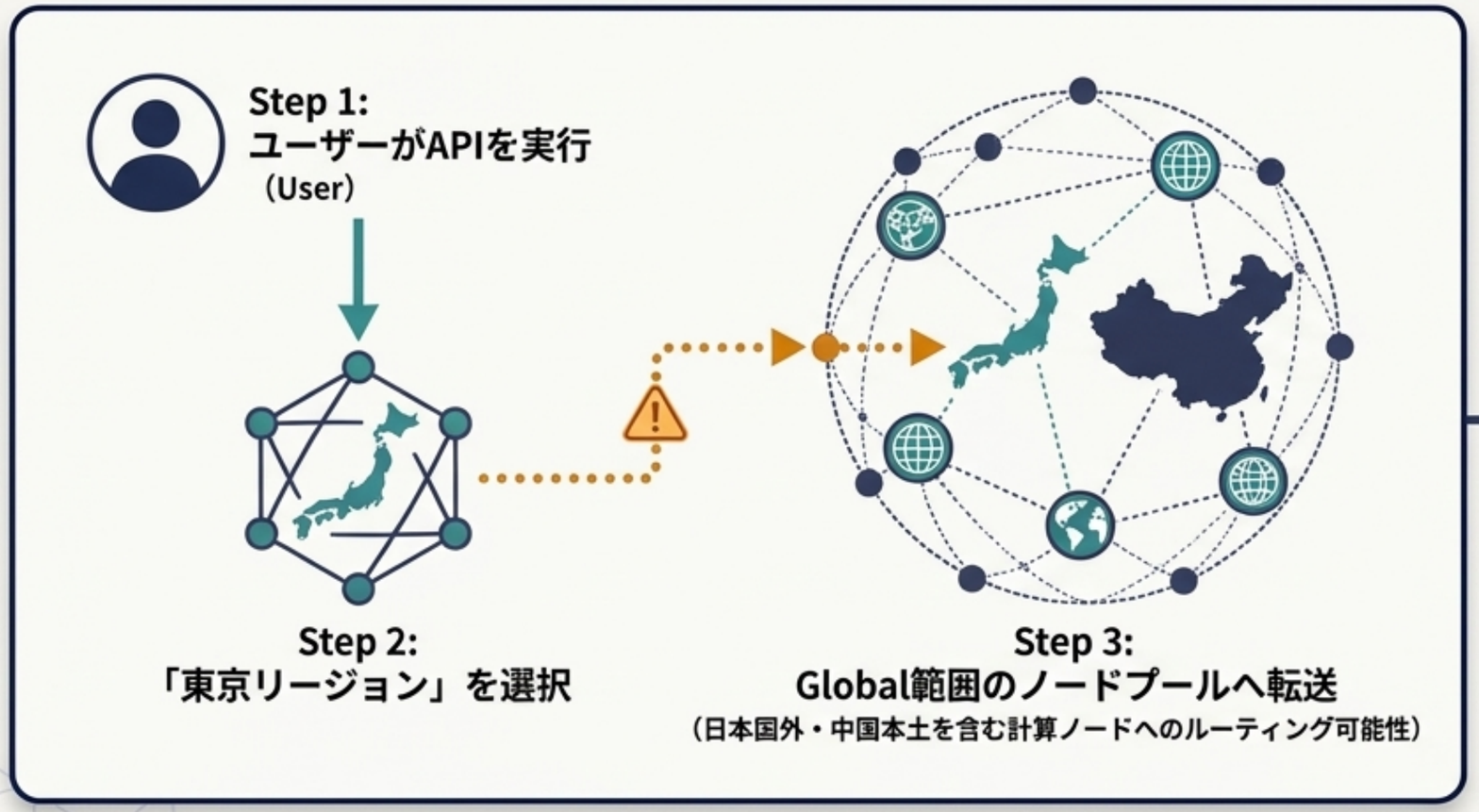
契約・法務調査補助:
幻覚リスク (当事者・日付・義務の取り違え) に対する人間の二重確認が必須。

非推奨・最終判断は不可

FTO（侵害予防調査）と無効性の最終判断:
補助分析に留めるべき。原文・法的根拠との厳密な照合能力において重大な誤りのリスクが残る。

Gemini Notebook

データ・コンプライアンスの罠：東京リージョン選択に潜む越境リスク





未公開発明や社外秘資料を入力する場合、モデルごとの国内限定提供の可否確認と、ZDRの明示的な申請・設定が絶対条件となる。

Gemini Notebook

ガバナンス要件に応じた利用経路の選定：APIとToken Planの違い

Token Plan（チーム版）

用途制限: コーディング・エージェントツール内の「対話利用」に限定。

禁止事項（❌）

- ❌ 自動化スクリプト
- ❌ バックエンド連携
- ❌ 非対話型の一括呼び出し

学習利用

会話データをモデル学習に使わないと明記。

Model Studio 従量課金 API

用途制限: なし。毎晩のSDI処理や大量公報の自動分析など、エンタープライズ・システムへの組み込みに必須。

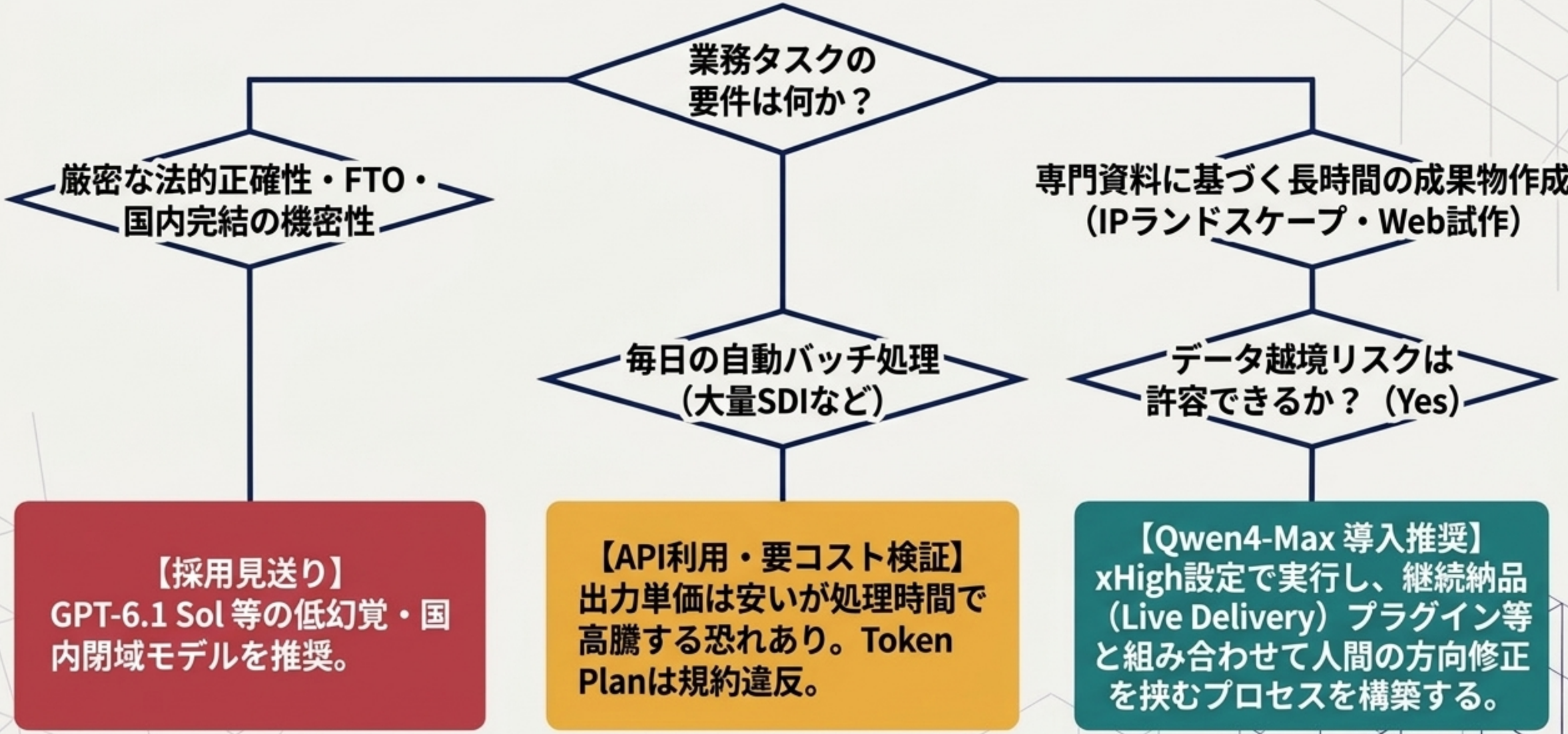
学習利用

法令・安全確保による例外を除き、データ保持（最大30日）が発生。

契約ごとの精査が必要。

Gemini Notebook

導入判断のためのインプリメンテーション・マトリックス



Gemini Notebook

自社環境での実証実験（PoC）に向けた評価設計図

Step 1: 課題の選定

公開公報、契約書、統合報告書などから「実務に即した30件」30件」を抽出。

Step 2: 推論設定の比較

Qwen4-Maxの「high（標準）」と「xhigh」の両方で実行し、思考トークンの消費量と出力品質のトレードオフを計測。

Step 3: 共通指標による測定（KPI）

- ☐ 重大誤りの有無
- ☐ 引用の正確性
- ☐ 人間による修正時間
- ☐ API完了時間
- ☐ タスク総費用

Step 4: 最終判断基準

「根拠付きの成果物をどこまで自律的に仕上げられるか」を軸に、既存モデル（GPT-6シリーズ等）とリプレイスする価値（費用対効果）があるかを判判定する。

Reference: 本調査報告は2026年11月16日時点の公式仕様・第三者評価（Artificial Analysis, Arena等）に基づく。

Gemini Notebook