

Grok 4.7 徹底評価：知財・法務 部門のための導入現実と戦略

SpaceXAIの最新モデルに対する「期待」と
「実効ROI」の乖離を解き明かす

EXECUTIVE BRIEFING

IP & LEGAL STRATEGY

2026.09

宇宙航空技術概略図と精密診断報告

Aerospace Technical Schematic and Precision Diagnostic Report

モデル仕様・プロファイル



Target: [grok-4.7] (Released: 2026-09-21/22)

Architecture: [2.1T Parameters] (Estimated)

Context Window: [500,000 Tokens]

Modality: [Text + Image Input / Text Output]

Training: [Memphis Colossus] (Cutoff: Aug 2026)

経営陣向けエグゼクティブ・サマリー

【ポジショニング】

フロンティアのやや下、価格性能比は上位。
コーディングと長文処理に特化。

【実世界のコスト構造】

額面価格は安価だが、トークン消費量が膨大なため「実効コスト」は競合を大きく上回る。

【知財・法務での推奨用途】

法務特化の精度は低く（Harveyスコア19.6%）、
機密性確保のためAPI経由のPoC利用に限定すべき。

	Claude Fable 5.1	GPT-6 Astra	Grok 4.7 (SpaceXAI)
Intelligence Index (AA)	53 (首位)	53 (首位)	46 (ミッドパック)
Harvey Legal Benchmark	N/A	N/A	19.6% (低水準)
実世界のタスク完了コスト	最適化	最適化	高 (膨大なトークン消費)
最適な知財・法務ユースケース	特許明細書の執筆、高度な法的推論	複雑なポートフォリオ分析、大規模データ処理	電気工学CADデータの解析、長コンテキストの一次処理

【戦略的結論】 Grok 4.7は「純粋な知能」で首位を争うモデルではなく、特定のエンジニアリングタスクと長文処理に特化したTier 2モデルである。

公式発表値（自己申告）



Terminal-Bench 4.0:

38.0%

(自社「Grok Build」ハーネス使用)

EEBench: **66.0%**

(モデルカード xHigh設定)

独立評価 (Artificial Analysis)



Terminal-Bench 4.0:

26.0%

(標準化ハーネスでの実力値)

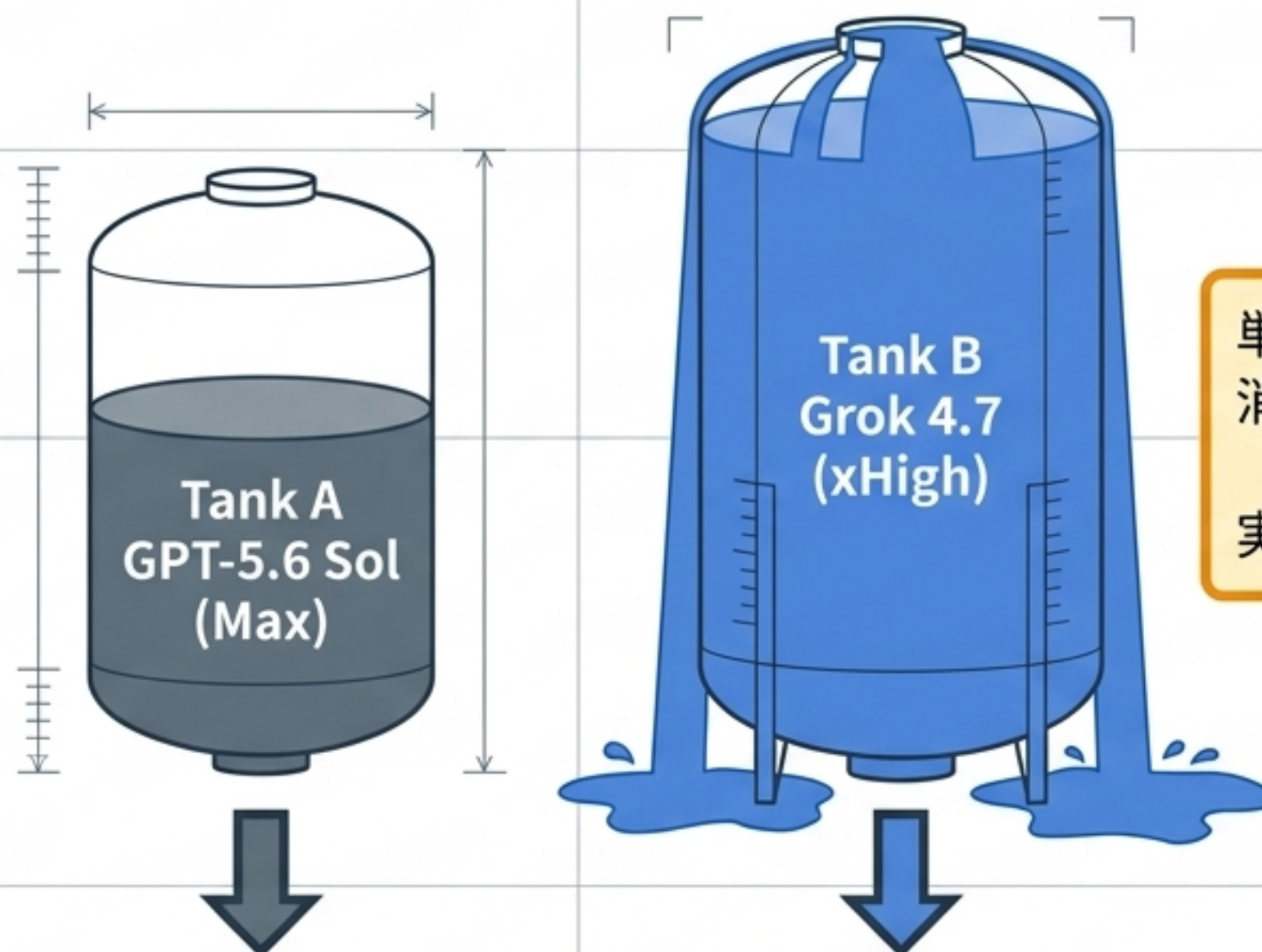
Note: GPT-6 Astra (60%)、Claude Fable 5.1 (55%)と大きな差が存在。

【ベンチマークの現実】

自社環境に依存した結果と、独立機関の標準環境での結果に大きな乖離（38% vs 26%）が存在する。マーケティング上の数値をそのまま実務評価に適用するのは危険。

[安価なトークン単価] × [膨大なトークン消費量] = [高額なタスク完了コスト]

単価: 比較的高額
消費量: 1タスクあたり
27,000 トークン
実効コスト: \$1.99 / タスク



単価: 安価 (\$2/\$6 per 1M)
消費量: 1タスクあたり
81,000 トークン
実効コスト: \$3.74 / タスク

【ROIの浸食】 Grok 4.7のトークン単価は安いですが、タスク完了に必要な出力トークンが競合の約3倍に達するため、大量の特許明細書生成などの業務では実効ROIが大幅に悪化する。

処理速度 (Speed)



計測値:	188 tokens/sec
所要時間:	平均 7.1分 / タスク
ステータス:	低速・冗長。長文プロンプトでは回答完了までにかなりの時間を要する。

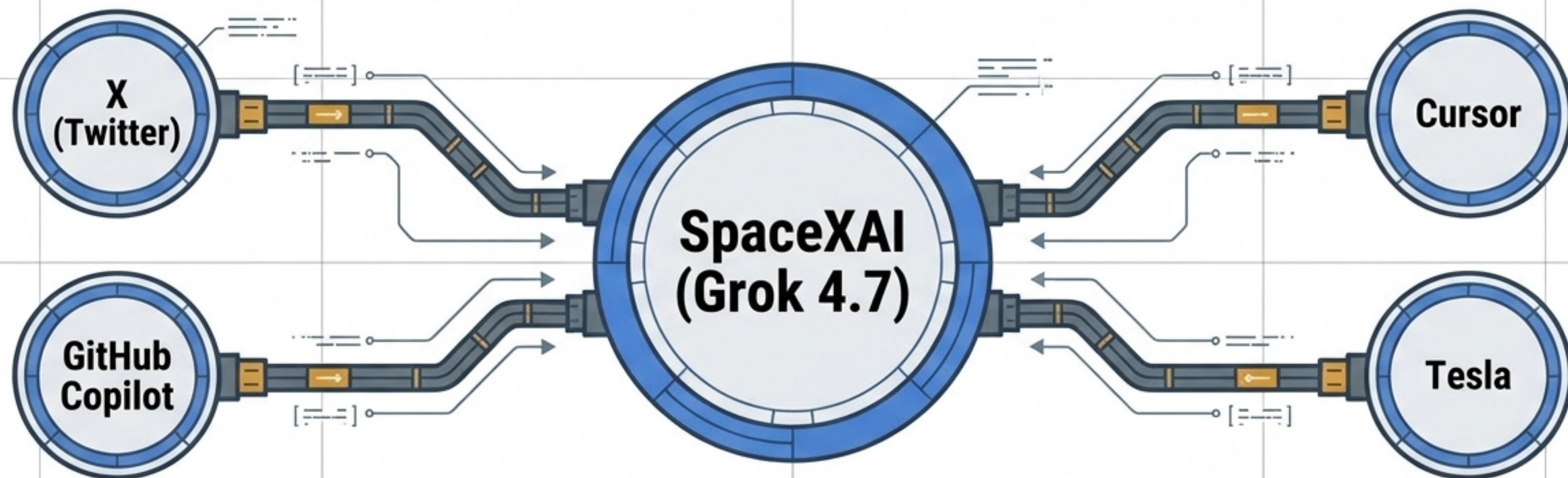
ハルシネーション (Accuracy)



虚偽率:	29% (AA-Omniscience)
正答率:	47% (横ばい)
ステータス:	Grok 4.6 (34%) から改善は見られるが、依然として実用上のハードルは残る。

【評価】 長時間の作業を要する自律エージェント型では改善が見られるが、即時性と正確性が求められる対話型の高度な法務タスクには不向きである。

「Grok 4.7はClaudeやGPTを純粋な知能で倒すためのモデルではない。コモディティ化とエコシステム統合のための武器である。」



SpaceX IPOの背景

史上最大規模のIPO（時価総額約1兆7,700億ドル）。AI部門の実力誇示と、メンフィスの「Colossus（宇宙データセンター）」インフラストラクチャへの投資を正当化する文脈を持つ。

エコシステム・ロックイン

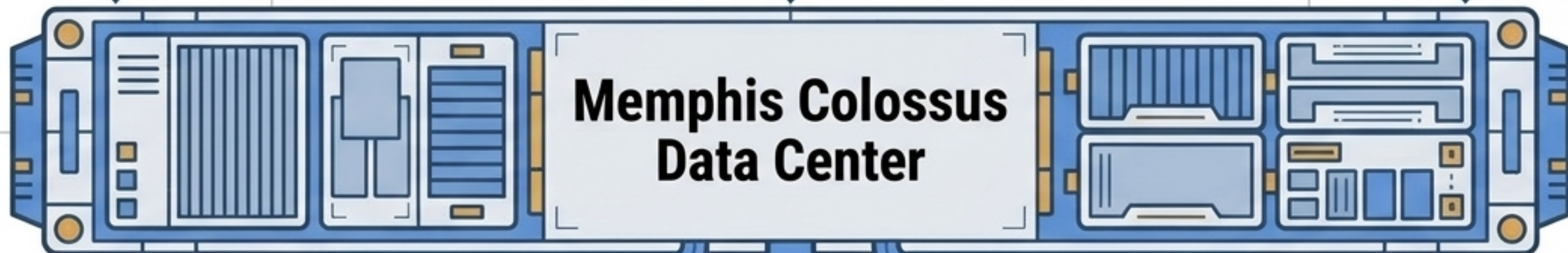
最高性能を追わず、据え置き価格設定と広範なチャネル（開発ツールやMS Office統合）を通じて、開発者や企業ユーザーのシェアを強引に獲得する戦略。



【独自の学習ソース (SpaceX Engineering Data)】



- ロケット・Starlink衛星のテレメトリ
- 製造ログ、故障解析データ
- Cursorワークフローの匿名化データ

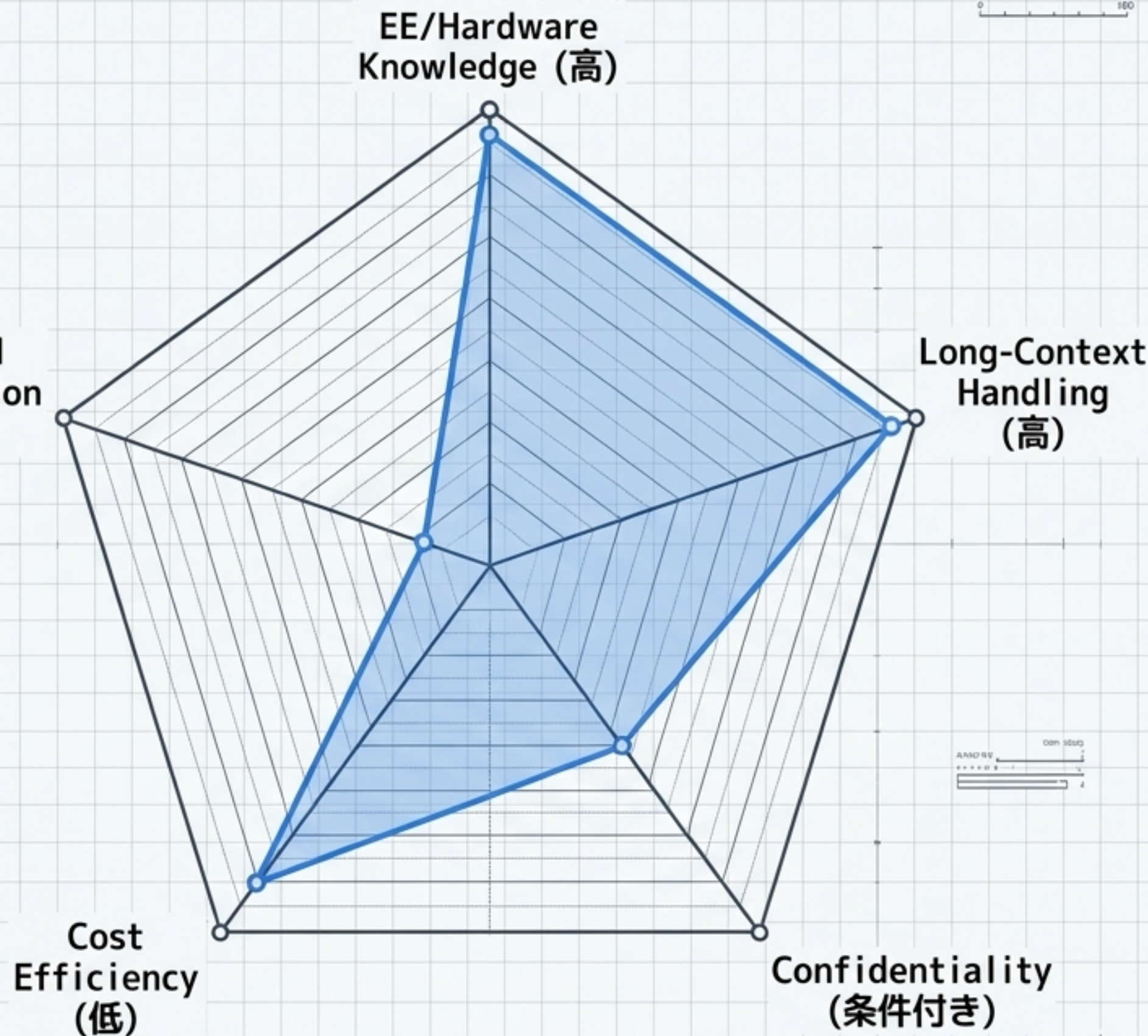


EEBench (電気工学): 64.0% - 66.0%
(高いハードウェア・電子工学理解)

CADGen: 44.4%
(図面・CADデータの空間理解力)

【知財実務への直結】

機械・電子・ハードウェア分野における特許調査や、複雑な技術仕様書の解析において、他社モデルにはない独自の強みを発揮する可能性がある。



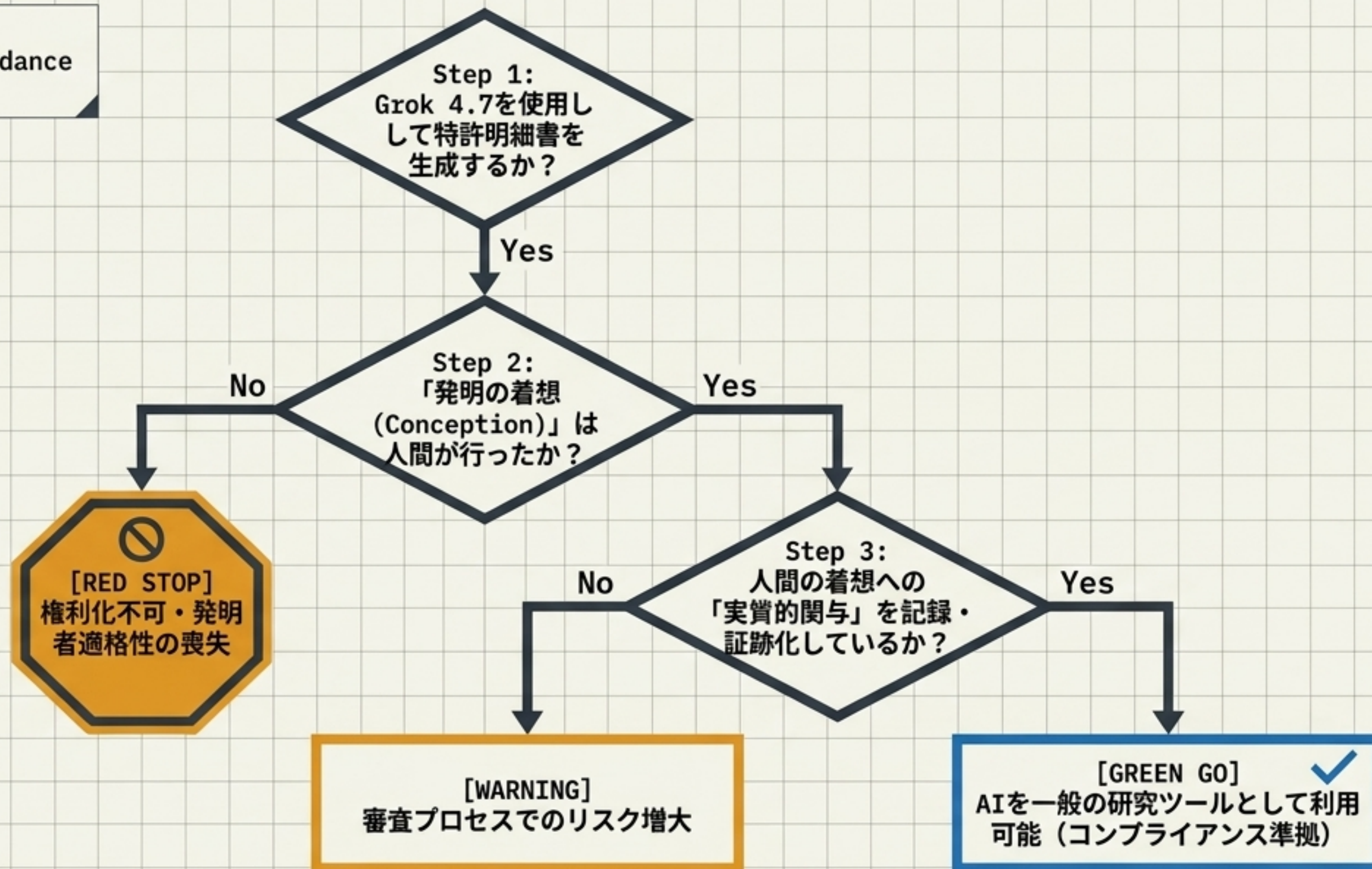
【特化型ツールとしての適性】

- EE/Hardware Knowledge: 宇宙・エンジニアリングデータ学習による優位性。
- Long-Context Handling: 50万トークンの入力枠と、複数ステップの処理能力。
- Legal Precision: 法的タスクのベンチマーク (Harvey 19.6%) は極めて不十分。
- Cost Efficiency: 冗長な出力によるトークン浪費でROIが悪化。
- Confidentiality: APIやEnterprise契約下でのみ安全が担保される。

結論：形状は「技術理解」と「長文処理」に鋭く突出し、「法務精度」と「コスト効率」で陥没している。万能な法務アシスタントではなく、ニッチな技術解析ツールである。

AEROSPACE SAFETY

Context:
USPTO Revised Inventorship Guidance
(2025年11月28日発効)



【鉄則】 Grok 4.7は「着想」を生み出す発明者としてではなく、技術仕様をパースする「翻訳機」として用いること。

Consumer Surface (X Platform Grok)



【RED / DO NOT USE】

- 未公開発明情報（機密）の入力は厳禁。
- 学習データとしての利用やデータ保持ポリシーに懸念あり。

CORPORATE FIREWALL /
ISOLATION BARRIER

Enterprise Surface (Grok API / Enterprise Copilot)



【GREEN / APPROVED FOR POC】

- ZDR (Zero Data Retention) / BAA / DPA等の枠組みを通じた利用。
- 自社環境でのAPIキー管理とセキュアなエンドポイント経由でのみ、特許業務への試験投入が可能。

【アイソレーションの徹底】 同じ「Grok 4.7」でも提供チャネルによって機密性の扱いが全く異なる。
知財部門はAPI経由の利用に隔離を徹底せよ。

ハイブリッド型 知財・法務ワークフローの提案

Phase 1: 複雑な技術データの解析

[Use: Grok 4.7]

Task: 電気回路図 (CADデータ)、ハードウェア仕様書、センサーテレメトリのパースと要約。

Why: 宇宙・エンジニアリングデータに裏打ちされた空間把握能力と、50万トークンの巨大な処理枠を活用。



Phase 2: 法的文書への落とし込み

[Use: Claude Fable 5.1 / GPT-6 Astra]

Task: 抽出された技術要件を基にした特許請求の範囲 (クレーム) の作成、明細書のドラフティング。

Why: 圧倒的な法的推論力、コンテキストに沿った簡潔な出力、そして実効コストの低さ。

**【最適解】 「Grokで技術を読み解き、Claudeで特許を書く」
適材適所のマルチモデル戦略が現在のエンタープライズ最適解である。**

即時アクション (Immediate PoC Focus)

☐ 隔離環境の構築

消費者向けアプリを禁止し、CursorまたはGrok API経由での限定テスト環境を整備する。

☐ コストと所要時間の実測

先行技術サマリ等の固定タスクを用い、「1完了タスクあたりの総コスト」と「所要時間」をGPT-6 Astra等と実測・横比較する。

再評価トリガー (When to upgrade)

次期モデルが以下の条件を同時達成した時点で主軸候補として再評価する：

- ☐ 1. Intelligence Indexが50を突破 (Fable/Astraクラス)。
- ☐ 2. Terminal-Benchの独立測定値が40%を超過。
- ☐ 3. 1タスクあたりのトークン消費量が競合並みに適正化。

【最終勧告】 現時点での全面採用は見送り。エンジニアリング分野の「特化型ツール」として試験運用しつつ、次世代モデルの飛躍に備えよ。