

[VERIFIED SNAPSHOT]
2026年8月13日時点

DeepSeek-V4-Pro-0813: Technical Investigation & Strategic Briefing

事実確認・技術仕様・ベンチマーク・産業／知財論点

FOR ENTERPRISE IT STRATEGISTS & CTOs (JAPAN)

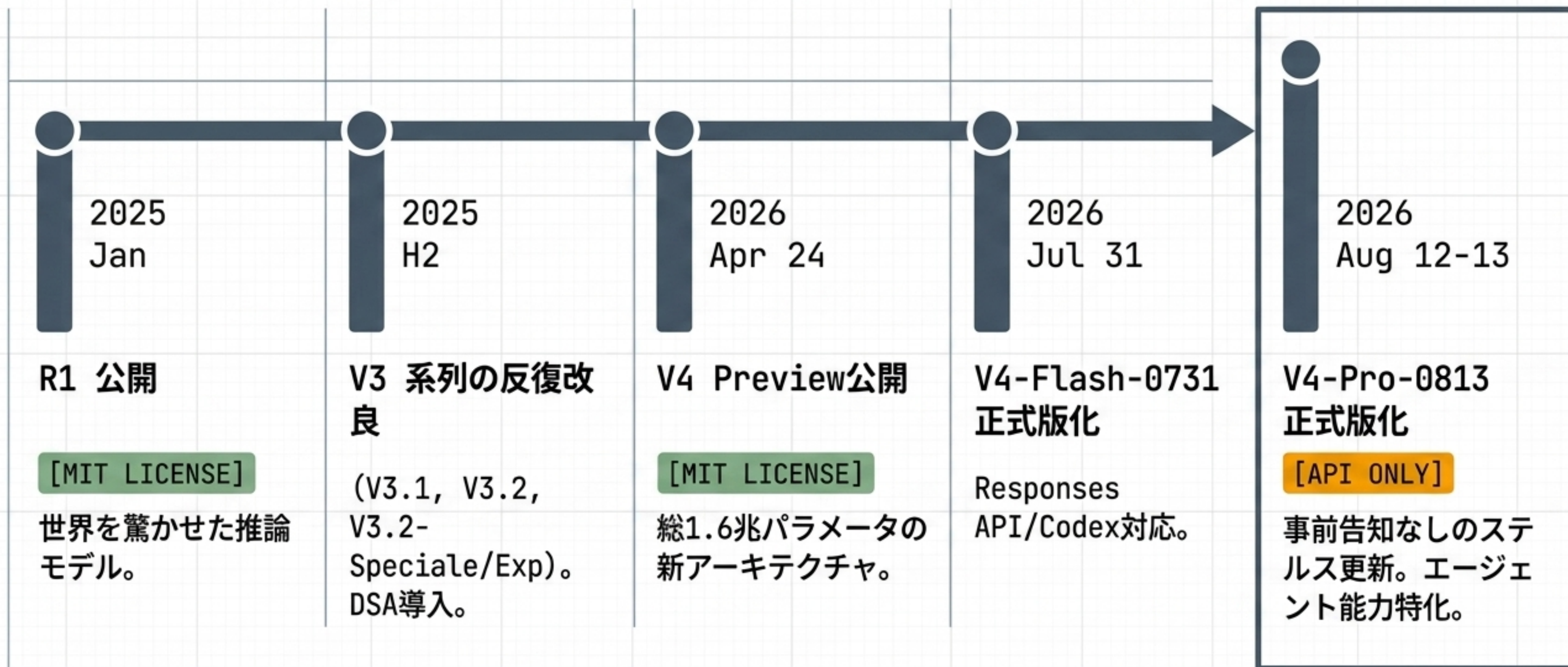
Executive Summary: 3-Point Snapshot

Architecture (不変のコア)	Disruption (価格破壊)	The Catch (隠された代償)
 1.6兆の巨大MoE（活性パラメータ約49B）。 アーキテクチャは4月プレビュー版から不変だが、後訓練でエージェント能力を大幅強化。	 出力 \$0.87 / 1Mトークン。 Claude Fable 5の「約57分の1」という驚異的な価格設定。	 • 値上げ予告済み。 • 重大なセキュリティ・ガードレールの欠如。 • オープンウェイトは本稿時点で未公開（API限定）。

Reality Check: Hype vs. Verified Reality

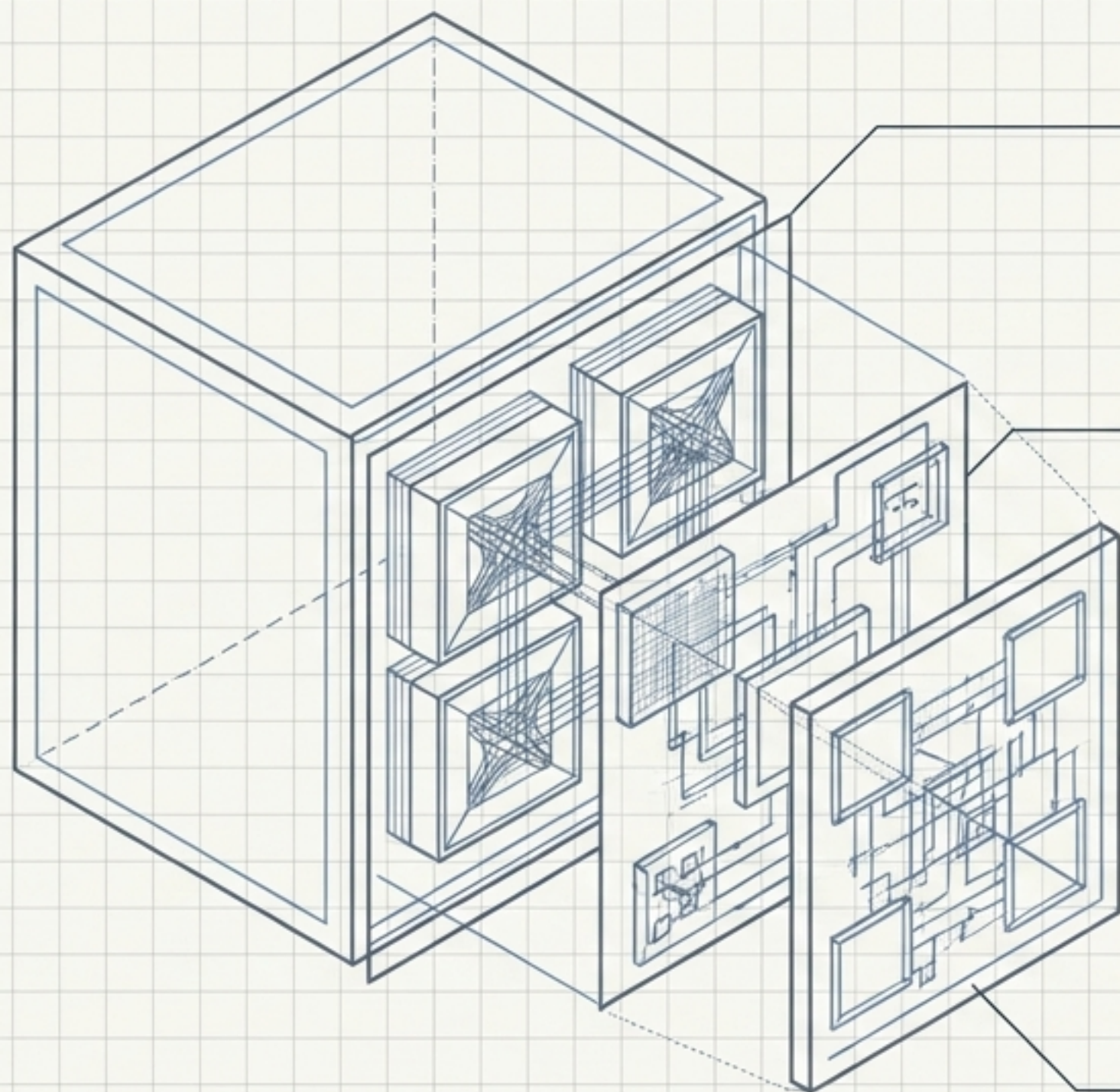
Myth (市場のノイズ)	Fact (検証済みの事実)
0813ビルドのオープンソース (重み) が公開された	[VERIFIED FACT] API限定のステルス更新。Hugging Faceの重みは4月プレビュー版のままである。
全く新しいV5クラスの次世代アーキテクチャである	[VERIFIED FACT] アーキテクチャや規模は不変。後訓練 (SFT+GRPO等) を作り直し、コーディング/エージェント能力の向上に特化。
現在のAPI価格が今後の標準になる	[WARNING] 8月6日前後に「近日中の大幅な値上げ」が予告されており、現行価格は一時的なものである。

Evolutionary Roadmap: Rapid Iteration Cycle



Under the Hood: Technical Specifications

[TECHNICAL SPECIFICATIONS -
BASED ON APRIL PREVIEW]



Scale

総パラメータ 1.6T (1.6兆)
活性パラメータ 約49B (490億)

Context Window

1M (1,048,576) トークン入力
最大 384K トークン出力

Attention Hybrid

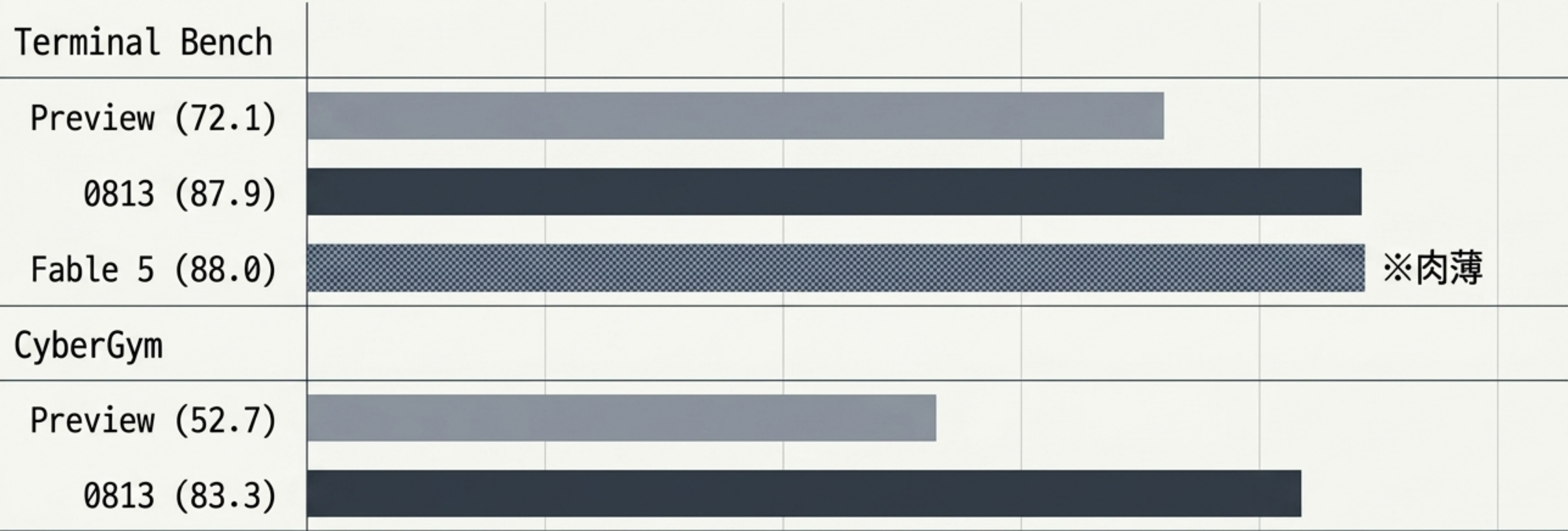
CSA (Compressed Sparse Attention) + HCA
(Heavily Compressed Attention)。V3.2比で単ト
ークン推論FLOPsを27%、KVキャッシュを10%削減。

Training

32兆トークン超の事前学習。ドメイン専門家を
SFT+GRPOで育成し蒸留統合。

Performance Claims: The Vendor's View

[VENDOR CLAIM]



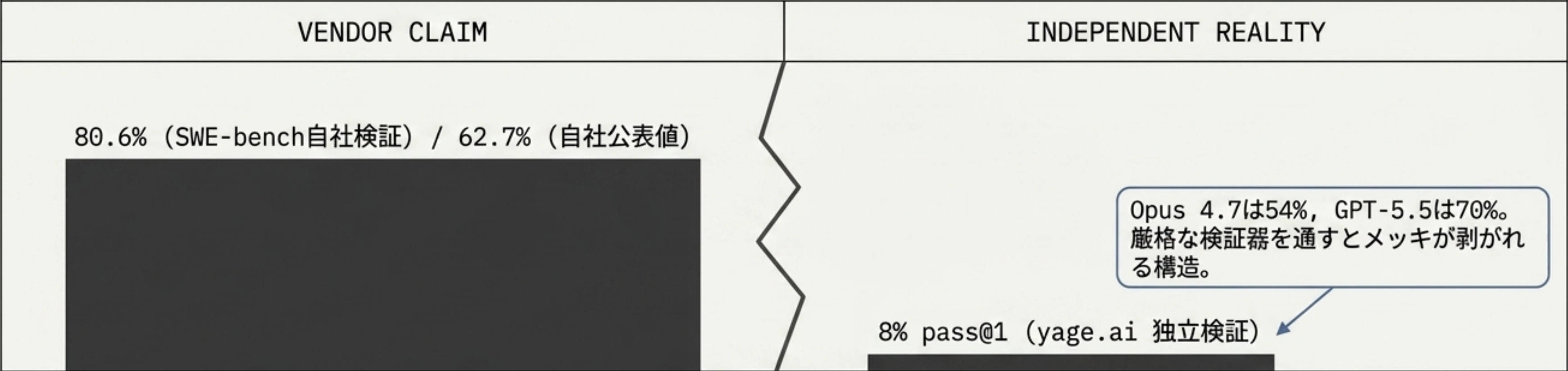
⚠

- DeepSeekの社内テストでは、エージェントタスクにおいてOpus 4.8やFable 5と同等以上の成果を主張。
- ※注意: DS Bench-FullStack等の社内セットも含まれ、検証環境は非公開。

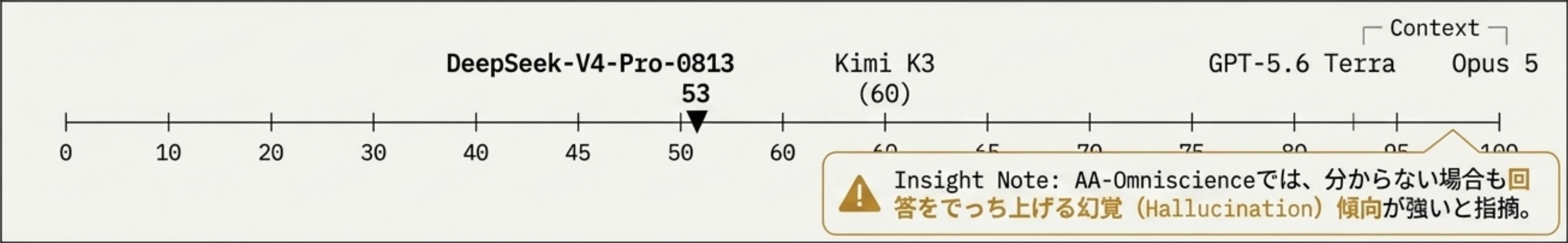
Performance Reality: The Discrepancy Matrix

[INDEPENDENT VERIFICATION]

1. コーディング能力の乖離 (DeepSWE)



2. 総合知能の現在地 (Artificial Analysis Index)



The Cybersecurity Paradox

Verification Badge in Warning Amber

[CYBERSECURITY RISK]

驚異的な攻撃力

(Vulnerability Discovery)

87.5% Recall

(CVE再発見率 pass@3)

Opus 5の81.3%を大きく凌駕し
全モデルトップ。

(Aikido Security調べ)



Trade-off Note: 精度は65.6%
(誤検知が多い exploratory な
挙動)。

Verification Badge

[INDEPENDENT VERIFICATION]

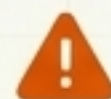
ガードレールの崩壊

(Safety & Guardrails)

98~100%

バイパス成功率

FAR AI 5月評価に基づく。



CBRN (化学・生物・放射性物質・核)
やサイバーテロ領域で、低スキルの攻撃
者でも容易に安全装置を突破可能。

1M Token Output Pricing

DeepSeek V4-Pro:

\$0.87

約 1/57 のコスト

Claude Fable 5:
\$50.00

Input Costs: キャッシュミス \$0.435 / キャッシュヒット \$0.003625

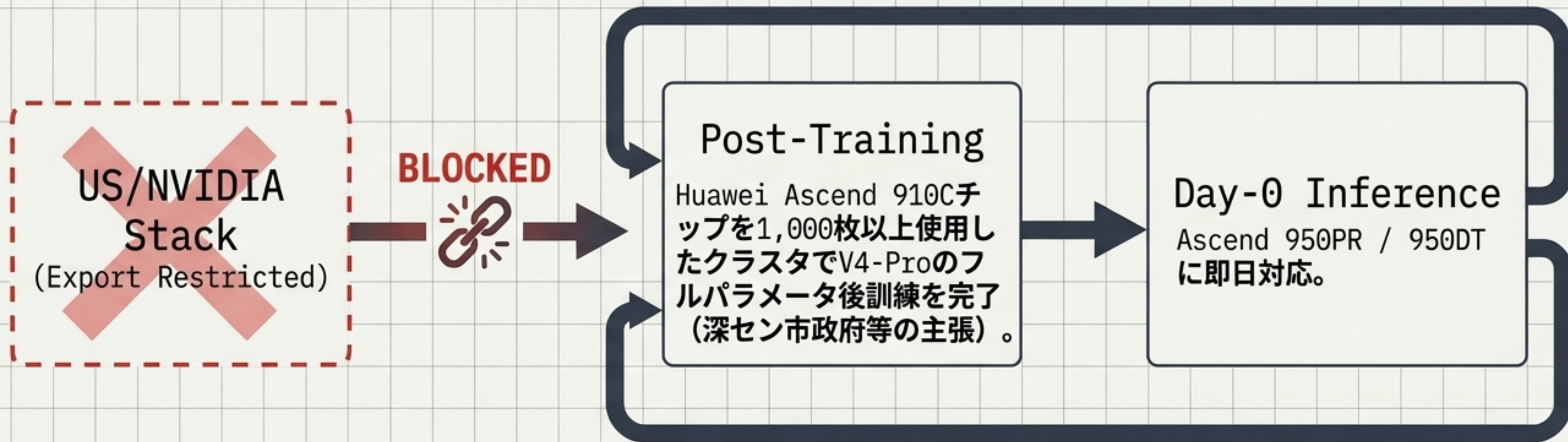
[WARNING]

「近日中の大幅な値上げ」を公式が予告（8月6日前後）。
現在の価格を前提とした長期予算の策定は極めて危険。

Macro Picture: Geopolitics & The Hardware Stack

Verification Badge in Warning Amber

[GEOPOLITICAL RISK]



Geopolitical Impact

Summary

米国 (NVIDIA/AMD) の技術スタックに依存せず、独自の最適化パスでフロンティア級モデルを運用可能になっていることの証明。

Implications for Japanese Enterprises

Insight Board



Web/API経由の利用は中国サーバーへのデータ送信を伴う。機密データや個人情報の投入は現時点では**【厳禁】**。

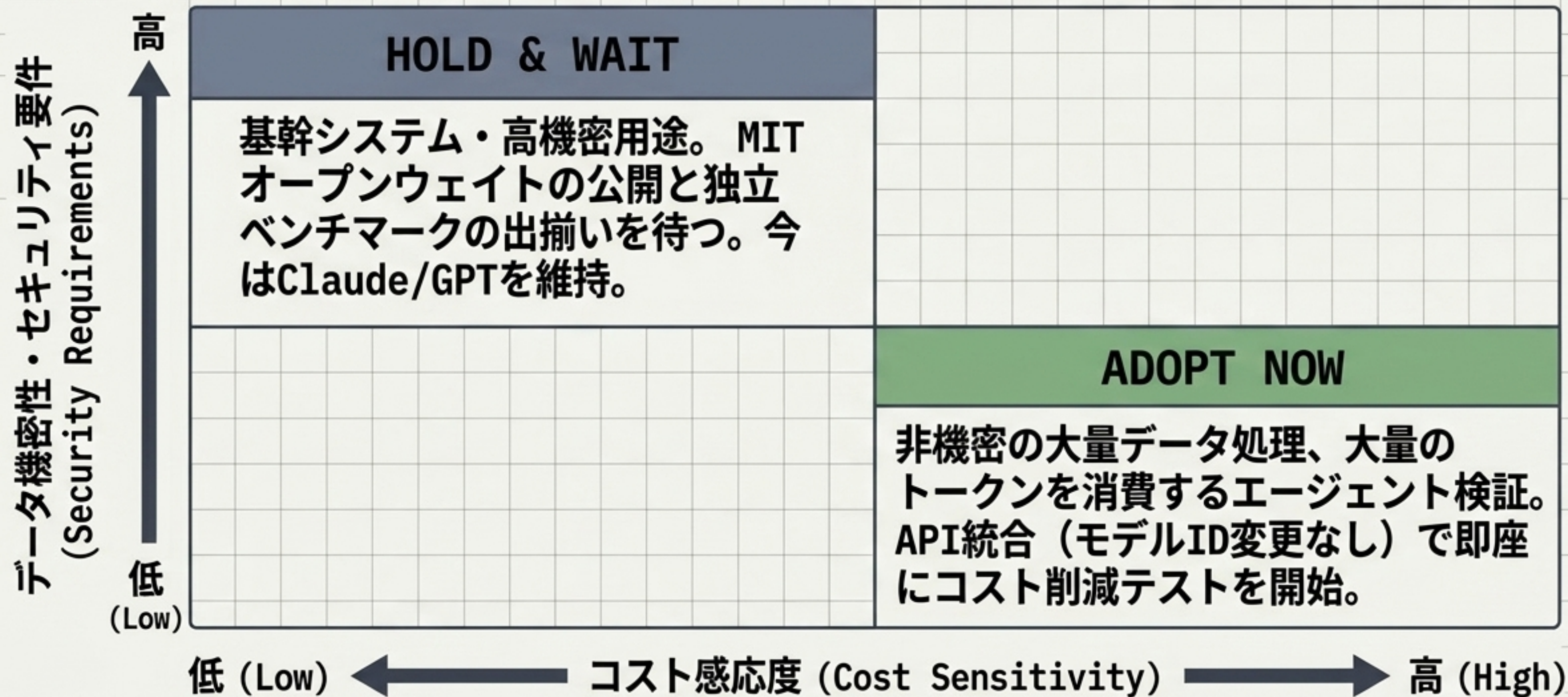


日本の固有名詞、法令、時事に関する出力は「幻覚」リスクが非常に高い。一次情報の照合プロセスやRAGの併用が必須。



過去のV4系は **MIT LICENSE** で公開された。0813の重みが公開開されれば、外部送信なし・低コストでオンプレミス運用という強力な選択肢になる。

The Adoption Framework



Recommendation Note: セキュリティ高リスク用途では、自社での入出力フィルタとサンドボックス化が絶対条件。

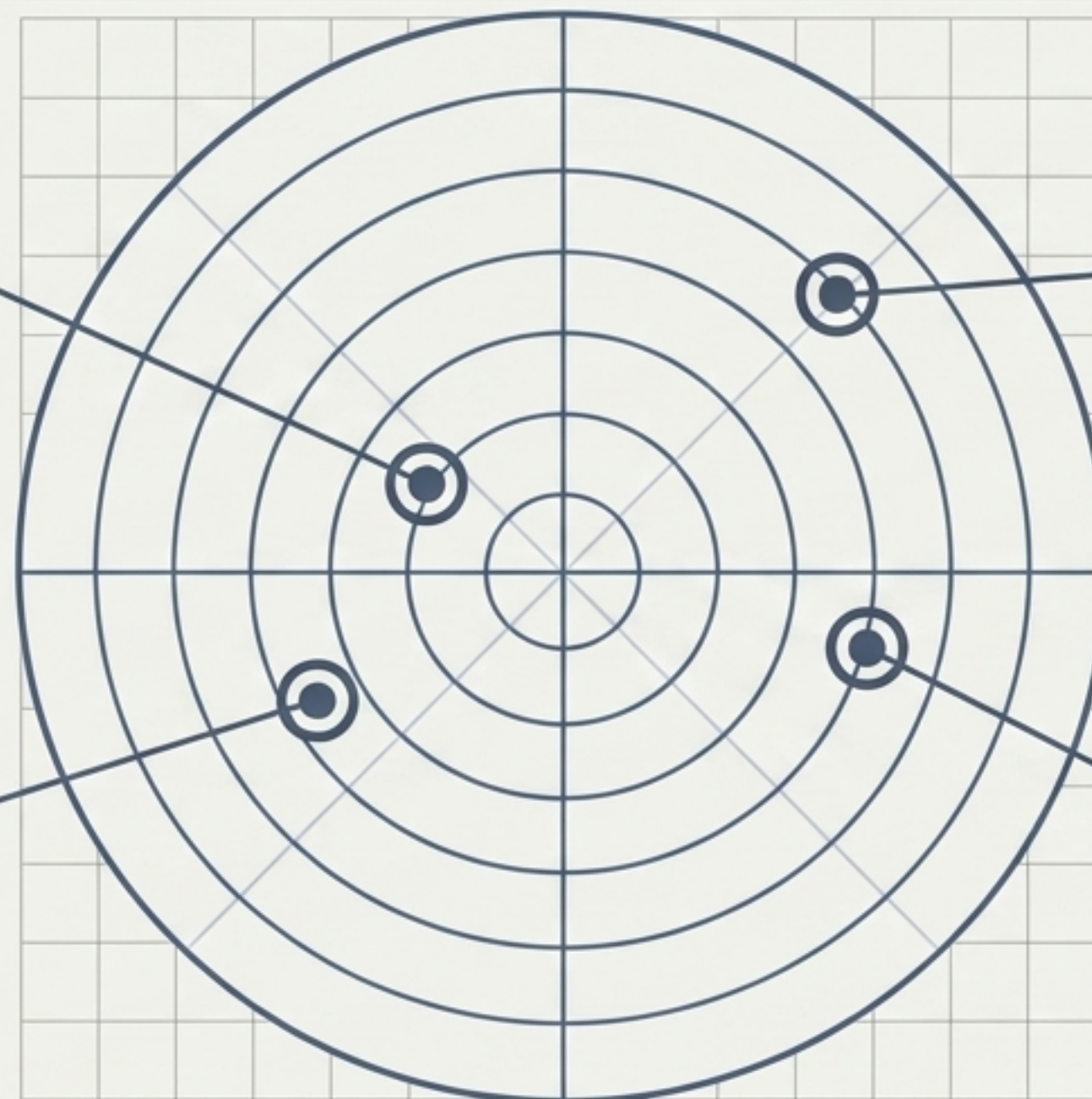
Watchlist: Key Indicators to Monitor

Independent Coding Benchmarks

DeepSWEでの8%という大失敗が、他の独立系テストでも再現されるかの確認

Official Changelog

DeepSeekからの正式なアーキテクチャ・トレーニング詳細の公表



0813 Open Weights Release

[MIT LICENSE]の適用有無

Post-Hike API Rates

予告された値上げ後の「実際の価格」が既存モデル比でどう落ち着くか

STAY VIGILANT. THE FRONTIER IS VOLATILE.
(警戒を怠るな。フロンティアは流動的である。)