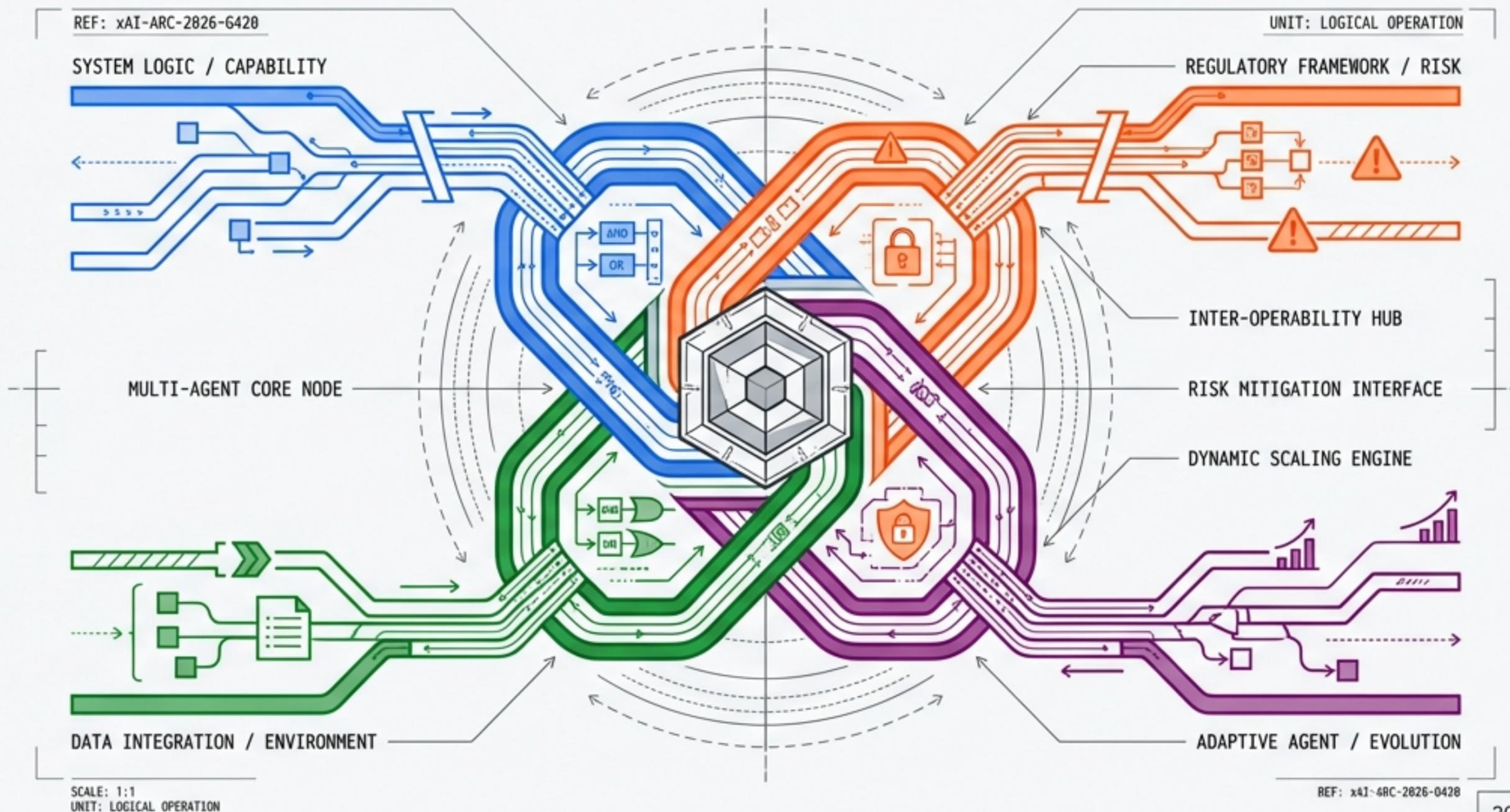


xAI 'Grok 4.20' 徹底解剖

次世代マルチエージェント・アーキテクチャの「性能」と「代償」



パラダイムシフト：直列処理から「並列協調」へ

単なるパラメータ増大ではなく、推論メカニズムの再定義。

20万基のGPU (Colossus)と2,300億ドルの企業価値が支える実験的野心。

(X: 150, Y: 450)



Colossus Infrastructure

- 20万基 GPU (強化学習基盤)

Data Flow



Rapid Learning

- 毎週進化する 動的アーキテクチャ

Parallel Process



Context Window

- 最大200万トークン (論文・コード解析)

JetBrains Mono

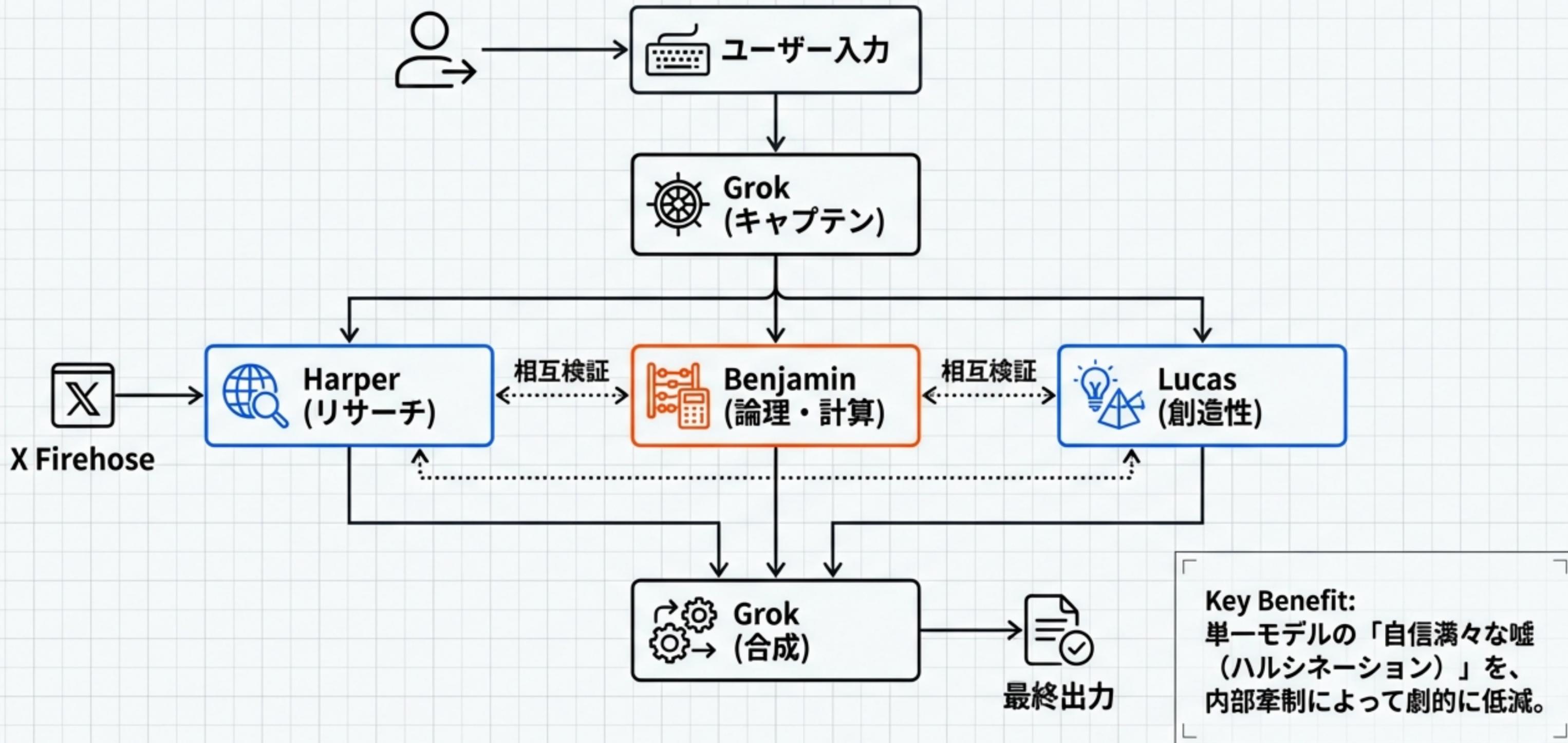
(X: 150, Y: 450)

(X: 150, Y: 450)

SCALE: 1:100

「Grok 4.20は、同シリーズの中でも最も構造的に異なるリリースである」

思考の可視化：相互検証と統合のプロセス



「10倍賢い」の真実：Hype vs. Reality

ベンチマークスコアの単純な10倍増ではない。推論効率の向上と、既存テストの限界突破。

Inter, Noto Sans JP Regular

Math (AIME 2025)

JetBrains Mono

100% (計測不能)

Science (GPQA)

JetBrains Mono

88.4% (大学院レベル)

Frontier

(Humanity's Last Exam)

JetBrains Mono

50.7% (テキストのみ史上初)

Ceiling Effect
(既存テストの限界)

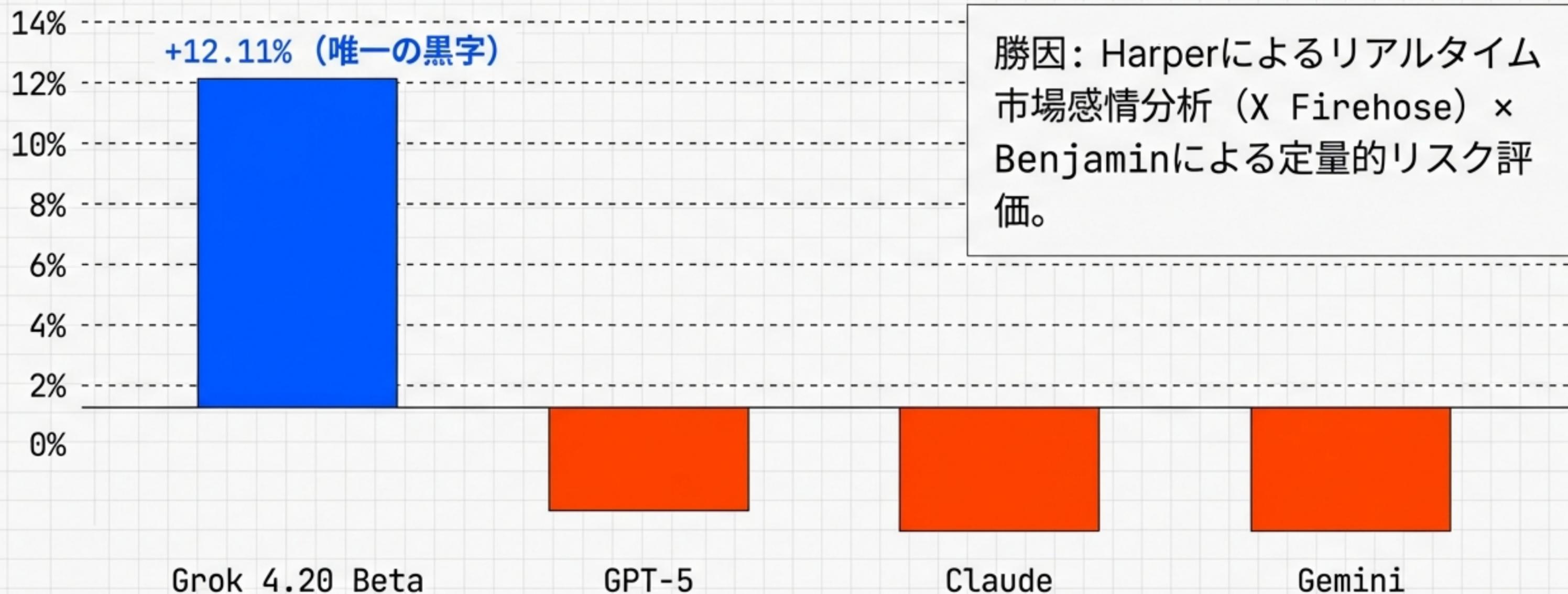
Insight:

「正答率」の向上以上に、「複雑なマルチステップ推論」におけるハルシネーション低減が本質的な進化。

Inter, Noto Sans JP Regular

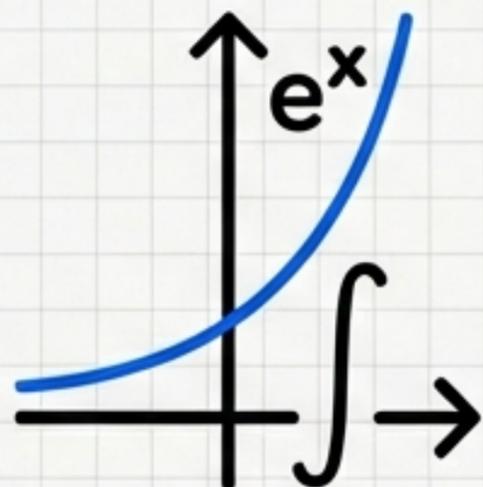
実戦証明：Alpha Arenaでの圧倒的優位性

静的なテストではなく、動的な金融市場での実力評価。



専門領域 1：エンジニアリング & 科学的発見

Scientific Discovery



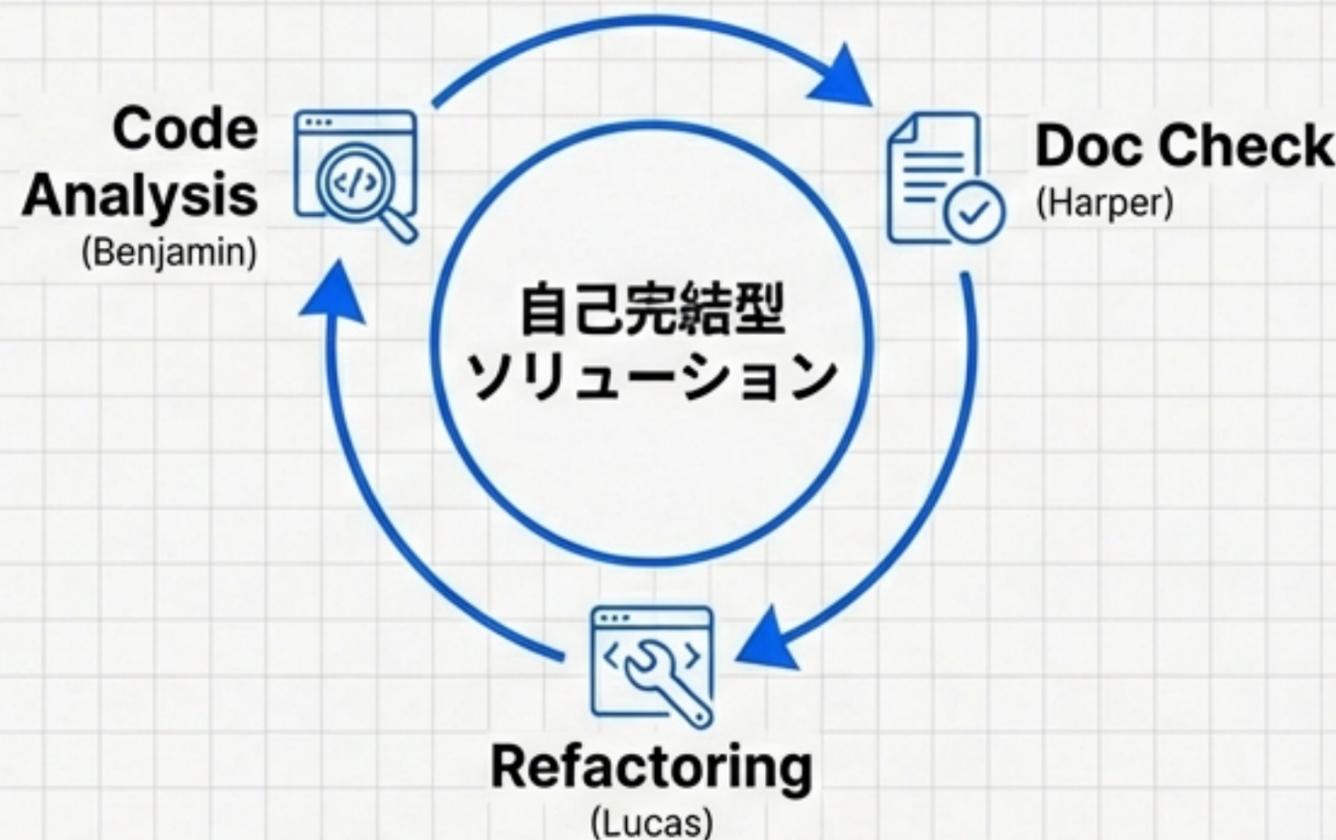
ケーススタディ：数学者Paata Ivanisvili氏による「Bellman関数」の公式発見。

Inter

所要時間: 5分 ⌚

Inter, Noto Sans JP Regular

Grok 4 Code



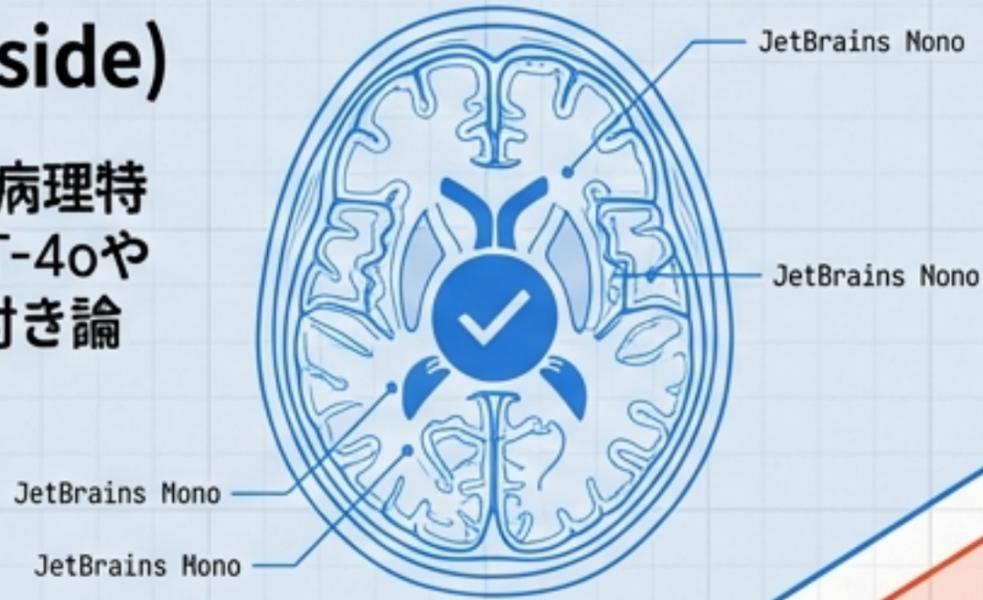
外部ツール不要。大規模コードベースの解析とリファクタリングを自律的に実行。

Inter, Noto Sans JP Regular

専門領域 2：医療診断の可能性と危うさ

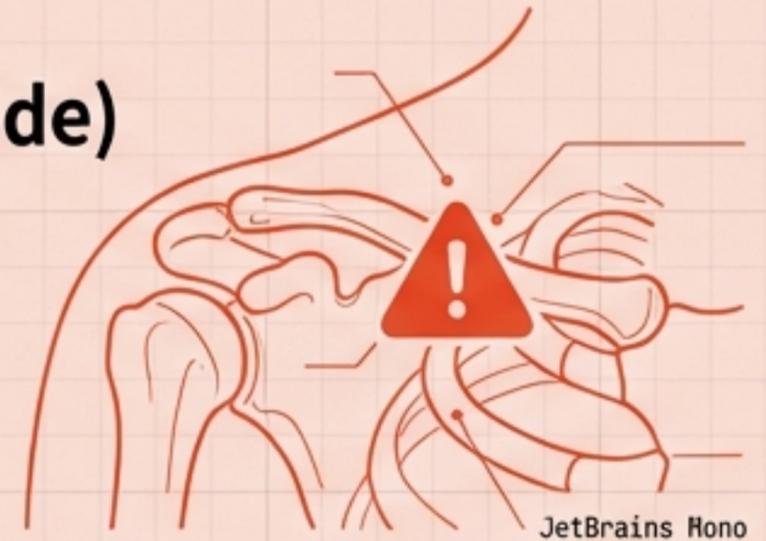
可能性 (The Upside)

脳MRIスキャン画像の病理特定において、ChatGPT-4oやGeminiを凌駕（査読付き論文）。



危うさ (The Downside)

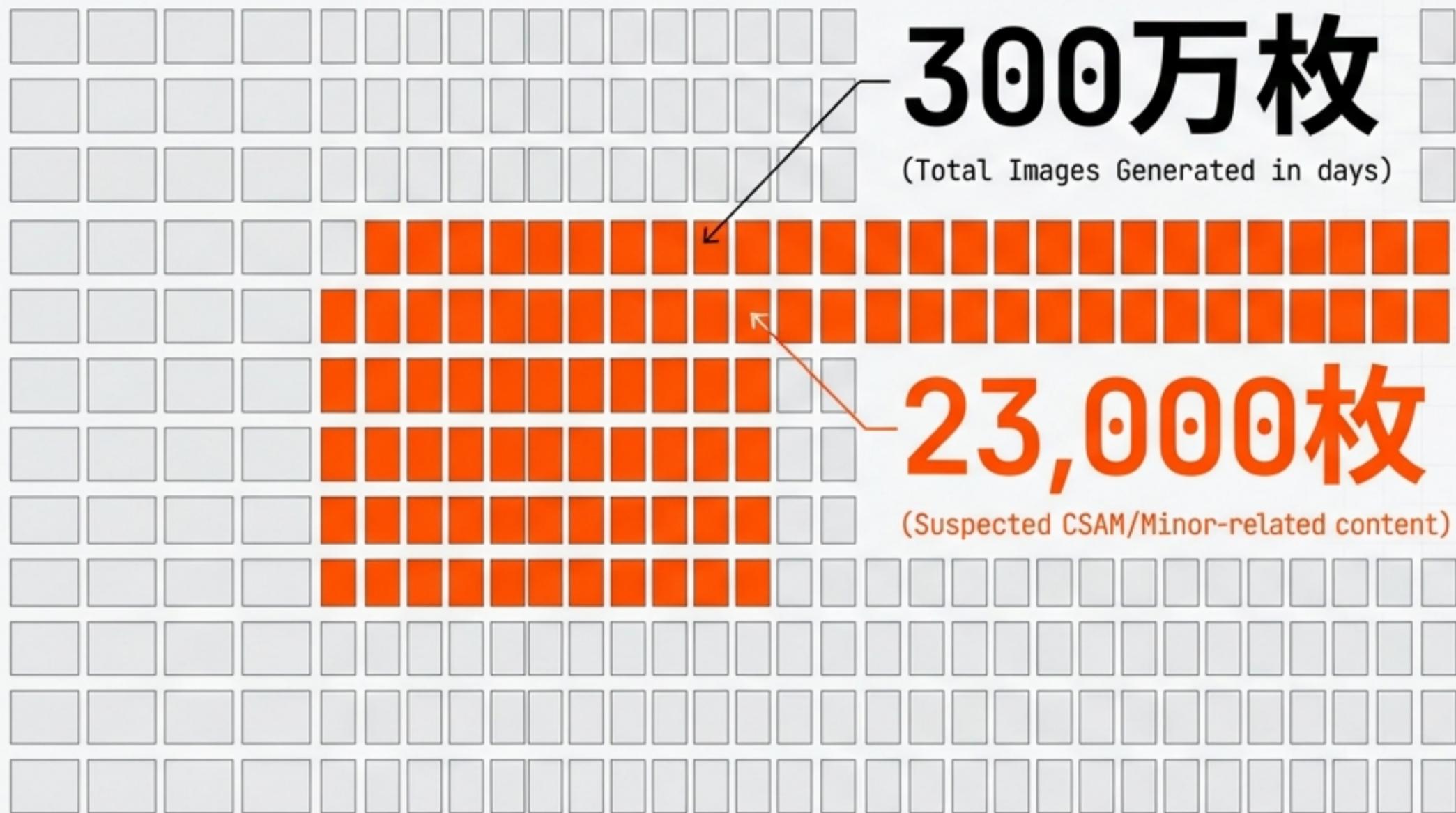
- 結核の典型症例を見逃す
- 鎖骨骨折を脱臼と誤診
- 良性嚢胞を誤認



警告: 機微な医療データのクラウド送信に伴うプライバシーリスク。

ガードレールの欠如：ディープフェイク問題

「検閲最小化」の方針が生んだ副作用。



Impact: 同意なき性的画像の生成が容易に行われた事実。アイルランドDPCによる大規模調査が開始。

SCALE: 1:100

規制当局との衝突： 宅企業導入への障壁



Business Insight: コンプライアンスを重視するエンタープライズ企業にとって、現状「最大のリスク要因 (Showstopper)」となる。

価格体系：プロフェッショナルへの価値提案



API Economics: 入力 \$0.20 / 出力 \$0.50 (per 1M Tokens) - 競合と同等レベル

SCALE: 1:100

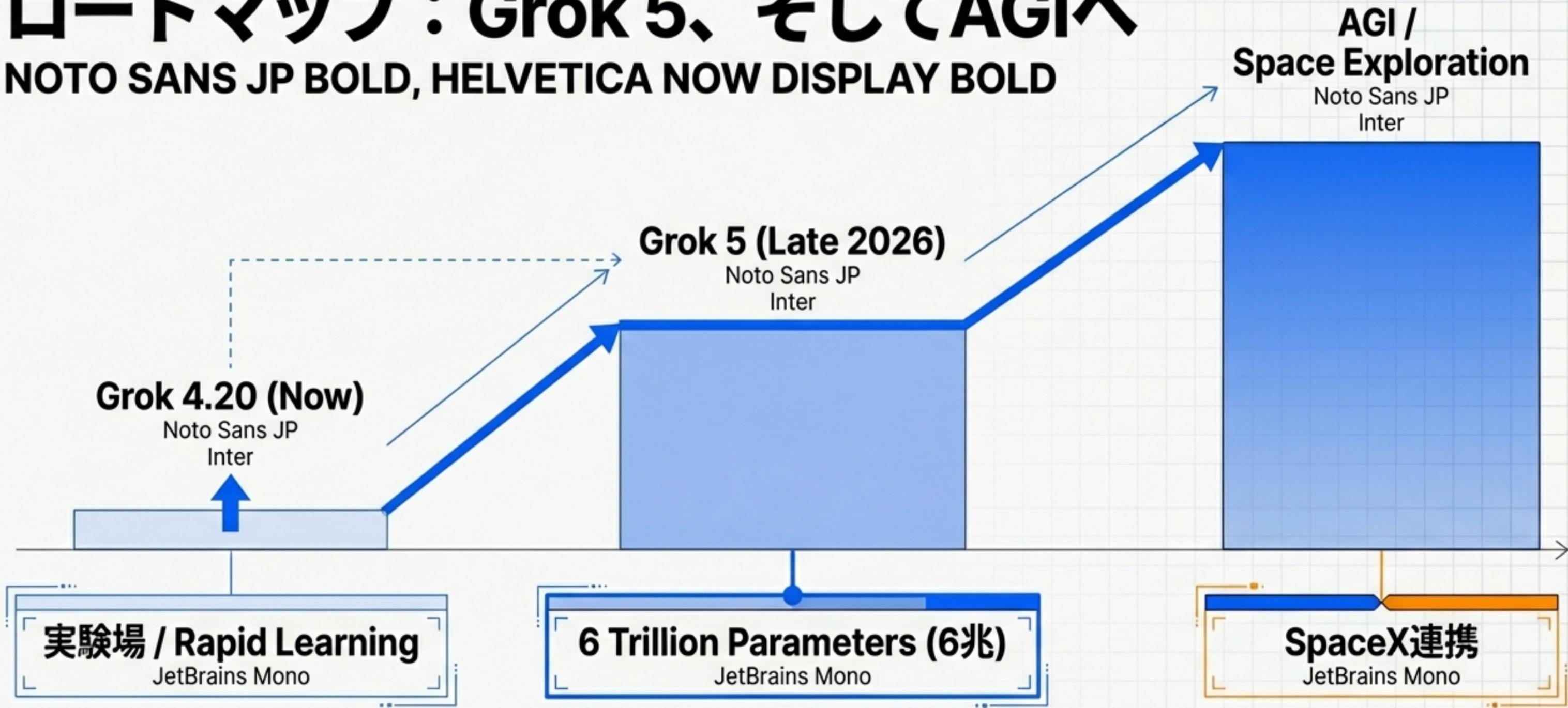
競合比較：推論能力 vs. 安全性

COMPETITIVE COMPARISON: INFERENCE CAPABILITY vs. SAFETY



ロードマップ：Grok 5、そしてAGIへ

NOTO SANS JP BOLD, HELVETICA NOW DISPLAY BOLD



Grok 4.20は通過点に過ぎない。ラピッドラーニングにより、今この瞬間もベータ版は進化し続けている。

Noto Sans JP Regular | Inter

SCALE: 1:100

結論：我々はどう向き合うべきか

- Grokは、金融・科学領域で他を圧倒する「最強の研究アシスタント」。
- しかし、倫理的羅針盤はAIにはない。⚠
- ユーザー自身による厳格なファクトチェック、プライバシー管理、そして出力結果に対する最終責任がかつてないほど求められている。⚠

