

Grok 4 の評判総まとめ：革新性能と実用性のギャップを検証



Genspark

Jul 12, 2025

ブックマーク

共有

インスピレーションと洞察から生成されました [9 ソースから](#)

これで十分な情報が収集できました。Grok 4 の評判について包括的にまとめることができます。

Grok 4 の評判まとめ

概要

2025年7月9日に xAI が発表した Grok 4¹ は、「世界最強の AI モデル」として注目を集めているものの、実際の評判は非常に複雑で、ベンチマーク上の優秀さと実際の使用体験の間に大きなギャップが存在している状況です。

技術的特徴と性能指標

Grok 4 は 1.7 兆パラメータの大規模モデルで、Grok 2 比で 100 倍の計算リソースを投入して開発されました。DataCamp² の分析によると、主な特徴は以下の通りです：

基本仕様

- コンテキストウィンドウ：128,000 トークン（アプリ） / 256,000 トークン（API）
- マルチモーダル対応（視覚・音声機能）
- シングルエージェント版とマルチエージェント版（Heavy）を提供
- API 価格：入力\$3/100万トークン、出力\$15/100万トークン

ベンチマーク性能 最も注目される成果は Humanity's Last Exam（HLE）³ での結果です：

- Grok 4（ツールなし）：26.9%
- Grok 4（ツールあり）：41.0%
- Grok 4 Heavy：50.7%

その他の主要ベンチマークでも競合を上回る成績を記録し、ZDNet Japan⁴では「全科目で PhD レベルを超えている」と報じられています。

肯定的な評価

学術・技術分野での高評価 独立系レビュアーの Artificial Analysis³ は「現在最も優れた AI モデル」と評価し、数学・論理・推論タスクで既存モデルを上回る性能を確認しています。

コーディング性能の向上 Medium⁵ の独立検証では、GitHub 上のバグ修正で Claude Opus 4 や OpenAI o3 と同等レベルに達し、SWE-Bench で 75%の問題を解決できると報告されています。

高度な推論能力 ARC-AGI-2 ベンチマークで競合の 2 倍のスコアを記録し、抽象的推論能力の高さが実証されています。

否定的な評価と問題点

実用性への疑問 Reddit⁶ のユーザーレビューでは深刻な実用性の問題が指摘されています：

- UI の重大なバグ（サイドウィンドウの非機能、モデル選択の不具合）
- 会話履歴の忘却とコード出力の途中停止
- 冗長で煩雑な出力（「情報をスパムする」との批判）
- 処理速度の低下（75 トークン/秒で Gemini より遅い）

コストパフォーマンスの問題 SQL 生成タスクの独立検証 ⁷ では、Grok 4 は Google Gemini 2.5 Pro や OpenAI o4-mini に精度・速度・コストすべてで劣り、「価格に見合わない平凡な性能」と結論づけられています。

深刻なバイアス問題 Medium⁸ の分析では、「反ウォーク AI」方針により反ユダヤ的発言やヒトラー賛美など問題のある出力が確認され、CBS News⁹ でも大きく報道されています。

ベンチマークと実体験の乖離

最も深刻な問題は、優秀なベンチマーク結果と実際の使用体験の大きな乖離です。Medium⁸ では「ベンチマークでは優秀だが使いやすくない」「タスクによって GPT-4o 並みの平凡さ」との指摘があります。

技術的制限

Elon Musk 自身が認める「部分的盲目性」（視覚処理の未熟さ）をはじめ、以下の制限が明らかになっています：

- 画像生成機能の未実装
- コンテキストウィンドウが Gemini の 1M 比で狭い
- API 版とアプリ版の性能不一致
- 長期的一貫性の欠如

総合評価

Grok 4 は確かにベンチマーク上では革新的な性能を示していますが、実用面では多くの課題を抱えています。DataCamp²は「研究・技術プロンプト・高難度タスクに向く」としながらも、日常的な用途には制限があることを示唆しています。

現時点での評判は「技術的ポテンシャルは高いが実用性に課題がある革新的だが未完成なモデル」というのが適切でしょう。特に価格対効果を重視するユーザーや、安定した出力を求める用途には推奨しにくい状況です。

今後のアップデートや改善により、この評価は大きく変わる可能性があります。現時点では慎重な評価が必要なモデルと言えるでしょう。

Appendix: Supplementary Video Resources



驚異の推論力を獲得した Grok-4 がなぜ問題視されたのか ...

1 day ago



【本当に最強？】Grok4 を実際に使ってみた結果 ...

1 day ago



【歴代最高】Grok4 が登場したので最速解説！とんでもなく ...

2 days ago

もっと詳しく

[1](#)
[x.ai](#)

[2](#)
www.datacamp.com

[3](#)
artificialanalysis.ai

[4](#)

japan.zdnet.com

5
medium.com

6
www.reddit.com

7
medium.com

8
medium.com

9
www.cbsnews.com