



Grok 4.5の評判・評価と今後に与える影響

概要・位置づけ

Grok 4.5は、SpaceXAI（旧xAI）が2026年7月8日（現地時間）に正式公開したコーディング・エージェントタスク特化型AIモデルです。Cursorと共同訓練された初のモデルで、パラメーター数1.5兆のV9基盤モデルにCursorのデータを追加学習させたものです。SpaceXとTeslaの社内でプライベートベータを経た後、一般公開に至りました。[\[1\]](#)[\[2\]](#)[\[3\]](#)

ベンチマーク性能：公式発表と実態のギャップ

公式が強調する指標

ベンチマーク	Grok 4.5	GPT-5.5	Claude Opus 4.8	Claude Fable 5
Terminal Bench 2.1	83.3%	83.4%	78.9%	84.3%
DeepSWE 1.0 (pass@1)	62.0%	64.3%	55.8%	66.1%
SWE Bench Pro (resolve rate)	64.7%	58.6%	69.2%	80.4%
AutomationBench-AA	51.4% (1位)	–	–	–

Grok 4.5はTerminal Bench 2.1でClaude Opus 4.8（78.9%）を上回り、GPT-5.5とほぼ同水準の83.3%を記録しています。一方、SWE Bench ProではClaude Fable 5が80.4%と圧倒的に首位であり、Grok 4.5の64.7%は15ポイント以上の差があります。独立系評価機関Artificial AnalysisのIntelligence Indexでは、Fable 5（59.9）、GPT-5.6 Sol（58.9）に次ぐ54でのランクです。[\[4\]](#)[\[5\]](#)

独立検証の評価

TryAIによる実アプリ開発比較では、Grok 4.5は**スピードとコスト効率で圧倒的な強み**を示しました。初回トークン出力が0.44秒、毎秒110トークンという速度は他モデルの約2倍で、コストは0.002セント/回答と最低水準です。ただし、3DルービックキューブのHTMLアプリ生成では**初回失敗**し再試

行が必要だったのに対し、Claude Opus 4.8とFable 5は一発で正確に生成しており、複雑タスクでの安定性に課題も見られました。^[6]

独立系エージェントベンチマークAutomationBench-AAでは51.4%で首位を獲得し、コスト1タスクあたり0.34ドルという経済性も実証されています。ただし同ベンチマークではガードレール違反が競合より多いことも指摘されています。また、一部の独立評価機関によるテストではハルシネーション率が競合よりやや高いという指摘も出ています。^{[7][8]}

価格競争力：最大の差別化要素

Grok 4.5の最も注目すべき点は、その圧倒的なコストパフォーマンスです。^[9]

モデル	入力（/1Mトークン）	出力（/1Mトークン）
Grok 4.5	\$2	\$6
Claude Opus 4.7	\$5	\$25
GPT-5.6 Sol	\$5	\$30
Claude Fable 5	\$10	\$50

さらに、SWE Bench Proでの1タスクあたりの出力トークン数がClaude Opus 4.8の67,020に対しGrok 4.5は15,954と約4.2倍のトークン効率を実現しており、実際のAPIコスト差は名目価格以上に大きくなります。マスク氏は「速度をさらに現行の2倍以上に高めることも可能」と述べており、今後の改善余地も示唆しています。^{[4][10]}

法律・知的財産業務への特筆すべき実績

Grok 4.5はHarvey’s Legal Agent Benchmarkで首位（SOTA）を獲得しています。これはマスク氏が「Grokはもともと法律タスクで非常に強い」と主張してきた路線に沿うもので、法務・IP分野のエージェント実用化において注目すべき指標です。ただし、この結果はX投稿を通じた発表であり、独立した外部機関による検証レポートは現時点で限定的です。^{[11][12]}

今後への影響：IP・特許業務の観点から

① エージェント大量展開時代の到来

Grok 4.5の低価格・高速・高トークン効率という組み合わせは、「高価なモデルを単一で使う」から「フロンティア級エージェントを大量並列展開する」へのパラダイム転換を加速します。特許調査、先行技術調査、FTO分析といった繰り返し性の高いエージェントタスクにおいて、Opus/Fableクラスの性能を1/4以下のコストで実現できる可能性は、IP部門のAI投資収益を大幅に改善させます。^{[13][14]}

② 法務エージェントの実用化競争

Harvey Legal Benchmark首位という実績は、法務・IP系エージェント（特許分類、無効化先行技術の自動スクリーニング、クレーム解釈補助など）への採用を検討する際の重要な参照点になります。PattnapsやPatentfield等の既存ツールとのAPI統合において、コスト効率面でGrok 4.5は有力な選択肢となり得ます。^[10]

③ モデル市場の構造変化：「三角形の競争」

2026年7月時点でAI市場は「最高精度コーナー（Fable 5）」「プラットフォーム深度コーナー（GPT-5.6 Sol）」「コスト効率コーナー（Grok 4.5・Muse Spark）」という三極構造に移行しつつあります。Grok 4.5は「絶対的精度の頂点」を狙うモデルではなく、エンジニアリングとエージェントタスクを速く安く確実にこなす位置取りを明確に選択しています。IP業務では、精度が最優先される明細書ドラフトや判断業務はFable 5/Sonnet系、大量並列エージェント処理はGrok 4.5、という使い分けが合理的な戦略となるでしょう。^{[15][5]}

④ 毎月新モデルリリースという圧力

マスク氏は「今年はゼロから学習した新規モデルをSpaceXから毎月リリースする」と明言しており、Grok 5（6兆パラメータ級MoEが予定）への橋渡しとしてGrok 4.5は位置づけられています。このペースでの性能向上が実現すれば、2026年後半にはGrokが純粹性能面でも最上位層に食い込む可能性があります。ただしその前提として、ハルシネーション問題の改善と、EU圏での利用制限解消（2026年7月中旬以降予定）が課題として残ります。^{[3][16][9]}

⑤ 日本企業・IP担当者への示唆

Cursor連携によりコーディング系エージェントとのシームレスな統合が可能であり、特許マップ作成の自動化やIPランドスケープ分析パイプライン構築においてコスト面での新しい選択肢が生まれます。一方でEU以外でも規制対応や情報セキュリティの観点からSpaceXAIのデータハンドリングポリシーの確認は必須です。既存ツール（Patsnap Eureka、Patentfield AIR、Summaria等）との補完的使い分けにおいて、Grok 4.5は「高速・低コスト中間処理モデル」として組み込む価値が高い一方、最終精度が問われるクリティカルな判断はFable 5やClaude Sonnet系を維持する二層構造が当面のベストプラクティスといえます。^{[14][5]}

★
★★

1. <https://gigazine.net/news/20260709-grok-4-5-ai/>
2. <https://hashout.jp/ai/2598/>
3. <https://gigazine.net/news/20260630-grok-4-5-cursor/>
4. <https://biz.chosun.com/jp/jp-it/2026/07/09/FW07FMMQE5E5ZOURCKKHPI6E5U/?outputType=amp>
5. <https://kenhuangus.substack.com/p/the-july-2026-frontier-muse-spark>
6. https://gigazine.net/gsc_news/en/20260709-grok-gpt-claude-build-apps
7. <https://www.binance.com/en/square/post/342799848209361>
8. https://x.com/ai_jitan/status/2075728007323603191
9. <https://www.techno-edge.net/article/2026/07/10/5277.html>
10. <https://www.marktechpost.com/2026/07/08/spacexai-releases-grok-4-5/>
11. <https://x.com/ValsAI/status/2074954965043519600>
12. <https://digg.com/tech/i9w2s7cd>
13. <https://innovatopia.jp/ai/ai-news/112897/>
14. <https://www.youtube.com/watch?v=CNm30PA30ns>
15. https://note.com/kazu_t/n/n9eac6270b41f

16. https://note.com/yasuda_forceai/n/n7d74004e94d5
17. <https://note.com/hirokimiyano/n/n8d4a408044d2>
18. https://gigazine.net/gsc_news/en/20260709-grok-4-5-ai/
19. <https://jls42.org/en/news/ia-actualites-8-jul-2026>
20. <https://www.emergingtechdaily.com/post/smartest-ai-2026-grok-4-vs-claude-4-5-vs-gemini-3-vs-gpt-5-2-tests>
21. <https://note.com/zephel01/n/n2783d84b389f?hl=en>
22. <https://aifriends.jp/xai-grok-4-5-private-beta-tesla-spacex/>
23. <https://vallettasoftware.com/blog/post/grok-4-5-vs-claude-vs-gpt>
24. <https://biz.chosun.com/en/en-it/2026/07/09/R4JZQ5HN2VFNTDJFPHSLP7XFM4/>
25. <https://temando.com.au/technology/the-ai-arena-gpt-gemini-claude-and-grok-battle-for-supremacy-in-2026/>
26. <https://uravation.com/media/grok-4-5-latest-model-guide-2026/>
27. <https://ai-revolution.co.jp/media/what-is-grok-4-5/>
28. <https://devgent.org/grok-4-5-private-beta-at-spacex-tesla-cursor-data-training-reaches-opus-ja/>
29. <https://felo.ai/ja/tools/grok-4-5>
30. <https://gigazine.net/news/20260709-grok-gpt-claude-build-apps/>
31. https://tenbin.ai/media/ai_news/grok-4-5-coding-guide
32. <https://jp.beincrypto.com/grok-4-5-automationbench-agent-benchmark/>
33. <https://uruchikara.jp/grok-4-5-v9-ec-katsuyou-2026/>
34. <https://x.com/KagawaTomo/article/2075331032648405148>
35. <https://handyai.substack.com/p/model-drop-grok-45>
36. https://dev.to/bean_bean/fable-5-vs-grok-45-for-coding-7-reports-aggregated-july-2026-4emg
37. <https://news.livedoor.com/topics/detail/31785645/>

38. <https://b.hatena.ne.jp/entry/s/gigazine.net/news/20260709-grok-gpt-claude-build-apps/>
39. <https://news.yahoo.co.jp/articles/31dfe11f2977106f9e89586d5dd1e3f3be22b451>
40. <https://tech.algomatic.jp/entry/2026/07/09/204004>
41. <https://cursor.com/ja/blog/grok-4-5>