

Grok 4.7 徹底調査レポート

— SpaceXAI（旧 xAI）が 2026 年 9 月 21 日（米国時間）に発表したモデルの特徴・性能・評判と知財実務への示唆 —

作成日：2026 年 9 月 22 日

Claude Fable 5.1

要旨

- **前提はおおむね正しい。** Grok 4.7 は米国時間 2026 年 9 月 21 日（月）、日本時間 9 月 22 日（火）未明に SpaceXAI が公開した^[1,2,4]。SpaceXAI は 2026 年 2 月の SpaceX による xAI 買収後の名称（XAI LLC の DBA 名）であり、xAI と同一主体である^[2,5,6]。
- **性能は「フロンティアのやや下、価格性能比は上位」。** 独立評価である Artificial Analysis の Intelligence Index は 46 で Grok 4.6 から 2 点上昇したものの、首位の Claude Fable 5.1 および GPT-6 Astra（各 53）には 7 点差がある^[15,16]。価格は入力 2 ドル／出力 6 ドル（100 万トークンあたり）で据え置きである^[1,2]。
- **知財実務への含意は「安価な長時間エージェント／コーディング用途の実務候補」。** ただしトークン消費量が競合の 2～3 倍で実効コストは額面より高く^[15,17]、消費者向け・マルチモーダルは未成熟、法務系ベンチマークの絶対値も低い（Harvey Legal Agent Benchmark 19.6%）^[1]。機密性の観点からも、本番投入前の自社評価（PoC）が必須である。

1. 前提の真偽確認

1.1 公開日時【確認済み事実】

SpaceXAI 公式ブログ「Introducing Grok 4.7」（2026 年 9 月 21 日付）^[1]、モデルカード（Revision 2026-09-21）^[2]、および開発者ドキュメント（モデル ID 「grok-4.7」）^[3]が同日に公開された。@SpaceXAI 公式 X アカウントは 16:17 UTC に「Grok 4.7 is here」と告知し、Elon Musk 氏も 16:25 UTC に投稿した^[4]。

16:17 UTC は、米太平洋夏時間 9 月 21 日（月）9:17、米東部夏時間同日 12:17、日本時間 9 月 22 日（火）1:17 に相当する。米国基準では「9 月 21 日」、日本時間基準では「9 月 22 日」であ

り、本調査の前提「9 月 22 日公開」は日本時間換算として正しい。

1.2 公開主体【確認済み事実／報道・二次情報】

モデルカードの脚注には、SpaceXAI が XAI LLC の DBA (doing-business-as) 名であり、「xAI」と「SpaceXAI」は互換的に用いられる旨が明記されている^[2]。報道によれば、SpaceX は 2026 年 2 月に xAI を買収（xAI 評価額 2,500 億ドル、SpaceX 1 兆ドル、統合後 1 兆 2,500 億ドル）し、同年 7 月に SpaceXAI へブランド改称した^[5,6]。したがって「SpaceXAI=xAI=SpaceX 傘下の AI 部門」であり、公開主体に関する前提も正確である。

1.3 報道との整合【報道・二次情報】

ITmedia、Impress Watch、SB ビジネス+IT など国内メディアも「米国時間 9 月 21 日、SpaceXAI が Grok 4.7 を発表」と一致して報じており^[7]、一次ソースとの齟齬はない。噂・リーク段階ではなく、正式リリースである。

2. 発表内容・特徴

2.1 位置づけと技術的変更点【確認済み事実】

公式ブログは Grok 4.7 を「コーディングと知識労働に向けた同社最高性能モデル」と位置づける^[1]。Grok 4.6 より大きな新ベースモデルを採用し、強化学習（RL）の期間を延長して数時間を要する難タスクに重みを置いた。自己検証能力と長コンテキスト管理を改善し、対話エージェント「Grok Bot」のハーネスをネイティブに理解するよう訓練されている^[1,2]。

2.2 仕様

- **パラメータ数【報道・二次情報】**：2.1 兆パラメータ（Grok 4.6 の 1.5 兆から約 40%増）と報じられている^[8,9]。ただし公式モデルカードには記載がなく、Musk 氏の発言・二次報道由来の数値である。
- **コンテキスト長【確認済み事実】**：500,000 トークン（Grok 4.6 から据え置き）^[2,10]。
- **マルチモーダル【確認済み事実】**：テキスト+画像入力・テキスト出力の「主にテキストモデル」^[2]。Musk 氏は公開前に、マルチモーダル性能には改善余地があることを認めている^[9]。

- **バリエーション【確認済み事実】**：標準版に加え、出力速度 2 倍・価格 2 倍の「Fast」版がある。Fast 版は Cursor および Grok Build 経由で提供され、公開 API には搭載されない [1,10]。

2.3 価格・提供チャネル【確認済み事実】

価格は据え置きで、20 万トークン未満のリクエストで入力 2 ドル／キャッシュ入力 0.50 ドル／出力 6 ドル（100 万トークンあたり）、20 万トークン以上で 4 ドル／1 ドル／12 ドルである。米国リージョンのエンドポイントは 10%増しとなる [1,2]。

提供チャネルは発表当日から Cursor、Grok Build、Grok API、各種サードパーティ・コーディングハーネス、モデルルーター、クラウド（OpenRouter、Vercel、Cloudflare、Snowflake、Databricks Mosaic 等）に及ぶ [1,11]。GitHub Copilot にも Pro／Pro+／Max／Business／Enterprise 向けに順次展開される（VS Code、Visual Studio、JetBrains、Xcode、Eclipse 等） [12]。Microsoft Word／PowerPoint／Excel の Grok アドインでは既定モデルとなった [1]。一方、消費者向け（Web／モバイル／X 上の Grok）は「後日追加予定」とされ、ローンチ時点では未提供である [1,13]。

2.4 学習インフラ・データ

モデルカードによれば、事前学習データのカットオフは 2026 年 6 月、補足学習には 2026 年 8 月までのデータを使用している [2]。また、Cursor の匿名化ワークフローデータでも補足学習を行った旨が脚注に記載されている [2]【確認済み事実】。SpaceX のエンジニアリングデータ（ロケット、Starlink 衛星テレメトリ、製造ログ、故障解析データ）を補足学習に投入したとの情報は報道ベースである [8,9]【報道・二次情報】。学習はメンフィスの Colossus データセンターで行われた [2,14]。

2.5 公式ベンチマーク【確認済み事実（自己申告値）】

公式ブログおよびモデルカードが公表する主要数値は表 1 のとおりである（Grok 4.7 は xHigh 設定、一部 high 設定） [1,2]。いずれも自己申告値であり、一部は評価パートナーが実施している。

表 1 Grok 4.7 の公式ベンチマーク（SpaceXAI 公表値）

ベンチマーク	Grok 4.7	Grok 4.6	比較対象（公式表記載）
CursorBench 4.0	46.3%	40.4%	GPT-5.6 Sol Max 41.7%、Claude Fable 5.1 Max 51.8%

ベンチマーク	Grok 4.7	Grok 4.6	比較対象（公式表記載）
DeepSWE v1.1	71.0% (high)	—	GPT-5.6 Sol Max 72.7%、Fable 5.1 70.0%
EEBench（電気工学）	64.0%（ブログ）／ 66.0%（モデルカード xHigh）	—	GPT-6 Astra Max 69.3%（モデルカードでは 2 位）
Terminal-Bench 4.0	38.0%	20.3%	Fable 5.1 57.9%
Harvey Legal Agent Benchmark	19.6%	—	比較 4 モデル中最高
HealthBench Professional	56.7%	—	GPT-5.6 Sol 60.5%、Fable 5.1 62.1%
GDPval（Elo）	1,695	1,605	Fable 5.1 1,735、GPT-6 Astra 1,542
AA Briefcase v1.1（Elo）	1,657（+111）	1,546	—
CADGen	44.4%	—	—

注：Terminal-Bench 4.0 の公式値は自社「Grok Build」ハーネスによる測定値であり、Artificial Analysis の標準化ハーネスでは約 26%と大きく低い（6 章参照）^[15]。

2.6 安全性・アラインメント【確認済み事実（自己申告値）】

「全面刷新した新セーフガードスタック」を搭載したとされる^[1]。モデルカードによれば、LatchBio バイオ安全ベンチマークで 62.4%（首位）、HackerBench v0.3 で危険なデュアルユースプロンプトの通過率 3.3%、CBRN 拒否（Bio refusal recall 100%、Chem 99.9%）、児童安全コンプライアンス違反 0.0%、MASK-Rectified（誠実性）0.00%、Sycophancy 0.03%である^[2]。生物・化学のデュアルユース能力は xAI の Frontier AI Framework（FAIF）の安全閾値を下回るとし、危険タスクではむしろ過去モデルより低スコアと説明している^[2]。一部サイバーセキュリティ・パートナーには、レッドチーム機能を招待制で提供開始した^[1]。

3. 評判・反響

3.1 独立評価（Artificial Analysis）【第三者評価】

Artificial Analysis の Intelligence Index（v4.3.2）で Grok 4.7 は 46（Grok 4.6 から+2）を記録し、同社は「SpaceXAI をトップ 4 ラボに押し上げた」と評価した^[15]。ただし順位としては、Claude Fable 5.1 および GPT-6 Astra（各 53）、Claude Opus 5（51）、Claude Fable 5（50）、Muse Spark 1.3（48）、GPT-5.6 Sol（47）に続く位置であり、首位 2 モデルとは 7 点差の「ミッドパック」である^[15,16]。Coding Agent Index（Grok Build 併用）は 56 で GPT-5.6 Sol を抜き 4 位。

長時間エージェント型知識労働（AA-Briefcase、GDPval）での改善が顕著とされる^[15]。

表 2 Artificial Analysis Intelligence Index（v4.3.2）における主要モデルの位置

モデル	Index	備考
Claude Fable 5.1	53	首位（同点）
GPT-6 Astra	53	首位（同点）
Claude Opus 5	51	
Claude Fable 5	50	
Muse Spark 1.3	48	
GPT-5.6 Sol	47	
Grok 4.7	46	Grok 4.6（44）から+2

出典：Artificial Analysis^[15]および OfficeChai^[16]の整理に基づく。

3.2 トークン消費・速度・ハルシネーション【第三者評価】

最大の批判点はトークン消費量である。Artificial Analysis によれば、Grok 4.7（xhigh）は Intelligence Index の 1 タスクあたり約 81,000 出力トークンを消費し、Grok 4.6（xhigh）の約 38,000（high 設定では約 36,000）の 2 倍超、Muse Spark 1.3（max）の約 60,000、GPT-6 Astra（max）の約 27,000 を大きく上回る^[15]。VentureBeat は、額面のトークン単価が安くても完了タスクあたりの実効コストは高くなりうると指摘し、Artificial Analysis の換算では Grok 4.7 xHigh は 1 タスク約 3.74 ドル、GPT-5.6 Sol Max は約 1.99 ドルとされる^[17]。

速度は長文プロンプトで約 188 トークン／秒、Intelligence Index の 1 タスク平均約 7.1 分で、同社のモデルページは「かなり低速かつ非常に冗長」と評している^[15]。一方、ハルシネーションは改善しており、AA-Omniscience ハルシネーション率は Grok 4.6（high）の 34%から 29%へ低下し、正答率は 48%→47%とほぼ横ばいである^[15]。

3.3 メディア論調【報道・二次情報】

海外専門メディアの論調は「着実な進歩だがフロンティアには届かず」で一致する。Decrypt は「大きくなったが、AI フロンティアのパーティーには遅刻」^[8]、The Decoder は「格安だが Claude・GPT-6 との差は大きい」^[18]、VentureBeat は「コーディング改善と据え置き価格だが、高トークン消費が実世界 ROI を脅かす」^[17]と評した。国内の ITmedia、Impress Watch、SB ビジネス+IT は「同社最高性能・価格据え置き・法務／電気設計で改善」と比較的中立的に報じて

いる^[7]。

なお、主要通信社では Reuters が短い速報を配信したのみで、Bloomberg、TechCrunch、The Verge、The Information による個別詳細は本調査時点で確認できなかった（未検証）。報道の中心は VentureBeat、The Decoder、SiliconANGLE 等の専門メディアである。

3.4 開発者反応 【報道・二次情報】

賛否両論である。Terminal-Bench での成績を「ひどい (horrendous)」と酷評する開発者がいる一方、CursorBench の価格性能比を「バカげているレベル」と称賛する声もある^[17]。Hacker News では、過去の Grok モデルが他社モデルの解けなかった設定問題を解いた経験を挙げる肯定的な声や、「フロントエンド Web 開発は優秀だがバックエンドは Claude が上」といった使い分けの評価など、ベンチマーク不信と実用評価が混在している^[19]。

3.5 競合比較 【第三者評価／分析・推測】

2026 年 9 月時点のフロンティアは GPT-6 Astra（9 月上旬発表）と Claude Fable 5.1 の二強（Intelligence Index 各 53）であり、Grok 4.7 はその下位グループで DeepSeek V4.1、Qwen3.8-Max、Gemini 3.8 Flash、Meta Muse Spark 1.3 等と価格性能比を競う位置にある^[15,16]。Musk 氏自身も公開前に「Grok 4.7 は Opus 5.0 相当であり 5.1 ではない」と予防線を張り、次期モデル（4.8＝顕著な改善、4.9＝Astra／Fable クラス、5＝AGI 級）を予告している^[9] 【報道・二次情報】。

4. 今後の影響・示唆

4.1 xAI／SpaceX の戦略 【報道・二次情報／分析・推測】

SpaceX は 2026 年 6 月 12 日に史上最大規模の IPO（1 株 135 ドル、約 750 億ドル調達、時価総額約 1 兆 7,700 億ドル）を実施した^[20,21]。Grok 4.7 は IPO 後の AI 部門の実力誇示という文脈を持つ。戦略の核は、ロケット・Starlink・データ・AI・X の垂直統合と、Musk 氏が掲げる「宇宙データセンター（orbital data centers）」構想である^[6,21]。学習基盤の Colossus（メンフィス、報道では 22 万 GPU／300MW）については、Anthropic が同施設のコンピュート契約を結んだとも報じられている^[14]。一方、AI 部門は年間約 64 億ドルの営業損失を抱えるとされ、IPO 時には「AI 事業が SpaceX の評価に影を落とす」との専門家見解も出ている^[20,21]。

4.2 AI 業界での位置づけ 【分析・推測】

「最高性能を追わず、価格性能比と広範な提供チャンネル（X／Tesla／Cursor／Copilot）で普及を狙う」戦略が鮮明である^[22]。xAI・OpenAI・Anthropicの三つ巴の価格・能力競争がCursor等の開発ツール内で展開されており、企業ユーザーは「ドロップダウンでモデルを切り替える」交渉力を得つつある。

4.3 知財実務への含意 【分析・推測】

(a) ドラフティング／長時間タスク Grok 4.7 は長時間・多段の作業継続を設計目標としており、明細書ドラフト、先行技術調査の一次生成、包袋管理等の長コンテキスト業務に理論上は適する。ただし法務系ベンチマークの絶対値（Harvey Legal Agent Benchmark 19.6%^[1]）は依然低く、人手レビュー前提の一次ドラフト用途に限定すべきである。

制度面では、USPTOの「Revised Inventorship Guidance for AI-Assisted Inventions」（2025 年 11 月 28 日発効、Federal Register 通知 2025-21457）^[23]が 2024 年 2 月版ガイダンスを全面撤回し、AIを一般の研究ツールと同等に扱いつつ、発明の着想（conception）は人間が担う必要があるとの整理になったと解説されている^[24]。AI活用成果物の権利化には、人間の着想への実質的関与を記録する運用が不可欠である。

(b) コスト 額面 2 ドル／6 ドルは安価だが、トークン消費が競合の 2～3 倍のため^[15,17]、大量の明細書生成や大規模ポートフォリオ分析では GPT-6 Astra 等の方が実効コストで有利となる可能性がある。自社の代表的業務でタスク完了あたりコストを実測すべきである。

(c) 機密性 特許業務は未公開の発明情報を扱うため、学習データ利用・データ保持ポリシーの精査が必須である。Grok 4.7 はエンタープライズ向けに BAA／DPA 等の枠組みを用意しているが^[2]、消費者向けサーフェスとは扱いが異なる。API／エンタープライズ契約での利用を推奨する。

(d) エンジニアリング分野の強み SpaceX データによる電気工学（EEBench 64～66%）・CADGen（44.4%）の相対的強さは^[1,2]、ハードウェア／電子／機械分野で技術理解・図面理解を要する知財業務において一定の優位となりうる。ただし自社ハーネス依存の数値であり、独立検証は限定的である。

5. 提言 【分析・推測】

- (1) **段階的 PoC（即時）**：消費者アプリではなく Grok API／Grok Build／Cursor 経由で、代表的知財タスク（先行技術サマリ、明細書一次ドラフト、OA 応答ドラフト、クレームチャート）を固定セットとして用意し、Grok 4.7・GPT-6 Astra・Claude Fable 5.1・Gemini 系を「タスク完了あたりコスト・所要時間・再試行回数・人手介入量」で横比較する。
- (2) **コスト実測の最優先**：額面単価ではなく「成功完了 1 件あたりの総コスト」で判断する。Grok 4.7 のトークン消費は競合の 2～3 倍であり、額面の安さが逆転しうる。
- (3) **機密ワークフローの隔離**：未公開発明情報はエンタープライズ契約（ZDR／BAA／DPA 相当）の下でのみ使用し、消費者向け Grok（X 上の Grok 等）は知財機密には使わない。
- (4) **用途の切り分け**：長時間エージェント／コーディング／電気・CAD 系は候補として評価し、法務系の高精度要求タスクは当面 Fable 5.1／Astra を主軸とする。
- (5) **再評価トリガー**：次期 Grok 4.8／4.9（Musk 氏は 4.9 を「Astra／Fable クラス」と予告）が Intelligence Index 50 超、Terminal-Bench 独立測定 40%超、トークン消費の競合並みへの改善を同時達成した時点で、主軸候補として再評価する。

6. 留意点（不確実性・未検証事項）

- **パラメータ数**（2.1 兆）および「40%増」は Musk 氏の発言・二次報道由来で、公式モデルカードに記載がない^[8,9]。
- **ベンチマークの二重性**：公式の Terminal-Bench 4.0=38.0%は自社「Grok Build」ハーネスによる測定値である。Artificial Analysis の標準化ハーネスでは同ベンチマークが約 26%（GPT-6 Astra 60%、Fable 5.1 55%）と大きく低く、ネイティブハーネスでも 33%であった^[15]。同社は Grok Build 併用の結果を Intelligence Index とは別扱いとして明記している。EEBench もブログ 64.0%／モデルカード 66.0%と食い違う^[1,2]。
- **自己申告値**：公式ベンチマークはすべて自己申告値である^[25]。独立検証は Artificial Analysis の Intelligence Index／Coding Agent Index が中心で、ARC-AGI、SWE-bench、HLE、GPQA、AIME 等の標準学術ベンチマークの独立スコアは本調査時点で確認できなかった（未検証）。
- **消費者向け・マルチモーダル**は未成熟／未提供である。ローンチ時点で Web／モバイル／X 上の Grok には未展開^[1,2]。

- **大手通信社の扱い**は限定的で、Bloomberg・TechCrunch・The Verge・The Information の個別詳細は確認できなかった（未検証）。
- **安全性数値**も自己申告であり、広範な第三者検証は途上である^[17]。
- **政治的バイアス・誤情報リスク**：Grok は過去世代（Grok 3/4）で反ユダヤ的発言や政治的バイアスの問題を起こした経緯がある。Promptfoo による第三者評価（Grok 4 対象、2,500 問・7 段階リッカート）では、Grok は他の主要 AI より右寄りだが依然として中道左派に位置し、Musk 氏自身の企業に対してむしろ厳しいと分析されている^[26]。Grok 4.7 世代での独立バイアス検証は本調査時点で未確認である（未検証）。

参考文献

-
- [1] SpaceXAI, "Introducing Grok 4.7"（公式ブログ）, 2026 年 9 月 21 日. <https://x.ai/news/grok-4-7>
 - [2] SpaceXAI, "Grok 4.7 Model Card", Revision 2026-09-21. <https://media.x.ai/v1/website/4p7card-5eccc980.pdf>
 - [3] SpaceXAI, 開発者ドキュメント（docs.x.ai、モデル ID: grok-4.7）. ※具体ページの URL は本調査で未検証
 - [4] BASENOR, "Grok 4.7, Explained in 5 Points". <https://www.basenor.com/blogs/news/grok-4-7-explained-in-5-points>
 - [5] Wikipedia（英語版）, "SpaceXAI". ※URL 未検証
 - [6] Yahoo Finance, "Elon Musk confirms SpaceX merger with xAI ahead of IPO".
<https://finance.yahoo.com/news/elon-musk-confirms-spacex-merger-with-xai-ahead-of-ipo-220616795.html>
 - [7] 国内報道：ITmedia NEWS、Impress Watch、SB ビジネス+IT（Grok 4.7 発表に関する報道、2026 年 9 月）.
※各記事の URL は本調査で未検証
 - [8] Decrypt, "xAI Launches Grok 4.7. It's Bigger, But Late to the AI Frontier Party".
<https://decrypt.co/378824/xai-launches-grok-4-7>（Yahoo Tech 配信版：
<https://tech.yahoo.com/ai/gemini/articles/xai-launches-grok-4-7-171603280.html>）
 - [9] CellCog, "Grok 4.7 Released: Specs, Price, Benchmarks, Rumors Graded". <https://cellcog.ai/blog/grok-4-7-release-date/>
 - [10] Kingy AI, "Grok 4.7 Is Out: New Features, Pricing and How to Try It". <https://kingy.ai/blog/grok-4-7-release-features-pricing-access/>
 - [11] LLM Stats, "Grok 4.7 Release: Same Price, Longer Horizons". <https://llm-stats.com/blog/research/grok-4-7-launch>
 - [12] Android Headlines, "Grok 4.7 Brings Big Coding Upgrades to Challenge Claude AI at Unchanged Pricing",

- 2026 年 9 月. <https://www.androidheadlines.com/2026/09/grok-4-7-ai-launch-coding-upgrades-pricing.html>
- [13] SQ Magazine, "xAI Launches Grok 4.7, Its Most Capable Coding Model Yet". <https://sqmagazine.co.uk/xai-launches-grok-4-7-coding-model/>
- [14] BitMEX Blog, "SpaceX IPO Guide: S-1 Breakdown, Valuation & Trading Strategy". <https://www.bitmex.com/blog/spacex-ipo-guide>
- [15] Artificial Analysis, "Benchmarking Grok 4.7". <https://artificialanalysis.ai/articles/benchmarking-grok-4-7>
- [16] OfficeChai, Grok 4.7 の Artificial Analysis Intelligence Index 順位に関する記事, 2026 年 9 月 21 日. ※URL 未検証
- [17] VentureBeat, "Grok 4.7 pairs coding gains with the same affordable pricing — but high token consumption threatens real-world ROI". <https://venturebeat.com/technology/grok-4-7-pairs-coding-gains-with-the-same-affordable-pricing-but-high-token-consumption-threatens-real-world-roi>
- [18] The Decoder, Grok 4.7 発表に関する記事（「格安だが Claude・GPT-6 との差は大きい」）. ※URL 未検証
- [19] Hacker News, "Grok 4.7"（スレッド）. <https://news.ycombinator.com/item?id=49788838>
- [20] SmartAsset, "SpaceX IPO: Valuation, Timeline and Investment Options". <https://smartasset.com/investing/spacex-ipo>
- [21] Intellectia.AI, "SpaceX IPO 2026: SPCX Stock Analysis, Valuation & Investment Guide". <https://intellectia.ai/blog/spacex-ipo-2026-investment-analysis>
- [22] SHAttered, "Grok 4.7's 80% Discount Squeezes Coding Rivals". <https://shattered.io/grok-4-7-price-war-coding-market-2026/>
- [23] USPTO, "Revised Inventorship Guidance for AI-Assisted Inventions", Federal Register 通知 2025-21457（2025 年 11 月 28 日発効）. ※URL 未検証
- [24] AIPPI, USPTO 改訂インベンターシップ・ガイダンスに関する解説記事. ※URL 未検証
- [25] DataCamp, "Grok 4.7: Features, Benchmarks, and Pricing". <https://www.datacamp.com/blog/grok-4-7>
- [26] Promptfoo, "Evaluating political bias in LLMs"（Grok 4 を対象とした評価）. ※URL 未検証

※本文中の上付き番号は上記参考文献の番号に対応する。「URL 未検証」の出典は本調査で記事の存在を二次的に確認したもの、URL を直接確認できなかったものである。