



Grok 4.3：性能とコストの均衡を破るAI市場の新たなゲームチェンジャー

Context:がリリースした新モデル「Grok 4.3」の主要機能、性能ベンチマーク、をして圧倒的なコストパフォーマンスを視覚的に伝え、開発者や企業が導入を検討するための判断材料を提供する。

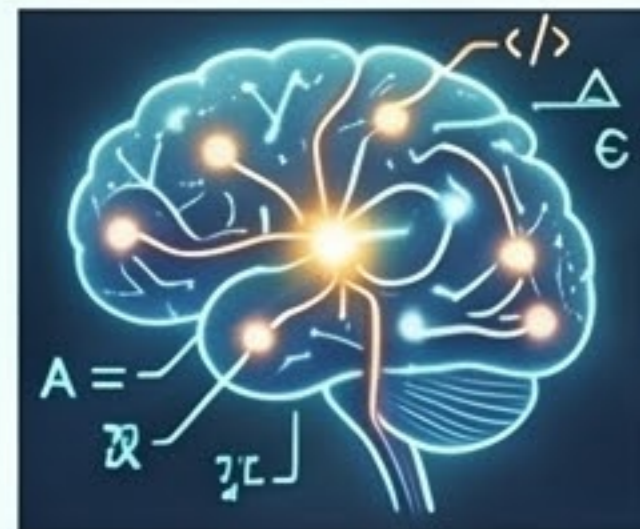
xAIが2026年4月に発表した「Grok 4.3」は、業界最高水準のコストパフォーマンスを武器に、AI開発市場の勢力図を塗り替える可能性を秘めた大規模言語モデルです。最大の特徴は、100万トークンの広大なコンテキストウィンドウと、すべての応答で内部的な「思考プロセス」を輸出する「常時推論 (Always-on Reasoning) アーキテクチャ」の導入にあります。性能面では、独立評価機関のベンチマークにおいて、OpenAIのGPT-5.5やClaude Opus 4.7といった最上位モデルには及ばないものの、Claude Sonnet 4.6を上回る実力を示しました。特に法律や金融といった専門分野での慧眼的推論や、エージェントタスクで高い評価を得ています。最大の影響は価格戦略にあり、前モデルから大幅な値下げを断行。競合のフラッグシップモデルに対し、同等タスクを約12分の1のコストで実現する圧倒的な経済性を実現しています。

主要機能と技術的特徴



100万トークンの広大なコンテキストウィンドウ

長大な文書の読解や、複数ステップを要する複雑なエージェントタスクを一度に処理する能力が大幅に向上しました。



常時推論 (Always-on Reasoning) アーキテクチャ

すべての応答で内部的な「思考プロセス」を経由することで、複雑な指示や多段階の論理的思考において安定した出力を実現します。



ネイティブ・マルチモーダル対応

テキストと画像に加え、新たに動画入力に対応。画質認識からのコード生成や、チュートリアル動画の要約が可能になりました。

性能ベンチマークと評価



Intelligence Index スコア「53」を記録

Artificial Analysisの評価で Grok 4.20 (49) から向上し、Claude Sonnet 4.6を上回る性能を証明しました。



法律・金融分野での卓越した性能

専門領域の判例推論や財務分析においてトップクラスの性能を発揮し、実務への応用が期待されています。



コーディングと自律性の課題

コーディング能力ではClaude Opus 4.7に差をつけられており、また長期タスク中に「アイドル状態 (ナルコレプシー現象)」になる課題も報告されています。

圧倒的なコストパフォーマンス



前モデルから最大83%の価格削減

入力トークンは約40%、出力トークンは約60%~83%と大幅に引き下げられ、市場のボリュームゾーンを狙う攻撃的な戦略をとっています。



競合フラッグシップの「12分の1」のコスト

高度な推論能力を要するタスクにおいて、Claude Opus 4.7と比較して劇的なコスト削減が可能です。

Grok 4.3

Grok 4.3と前モデル (Grok 4.20) のAPI価格比較

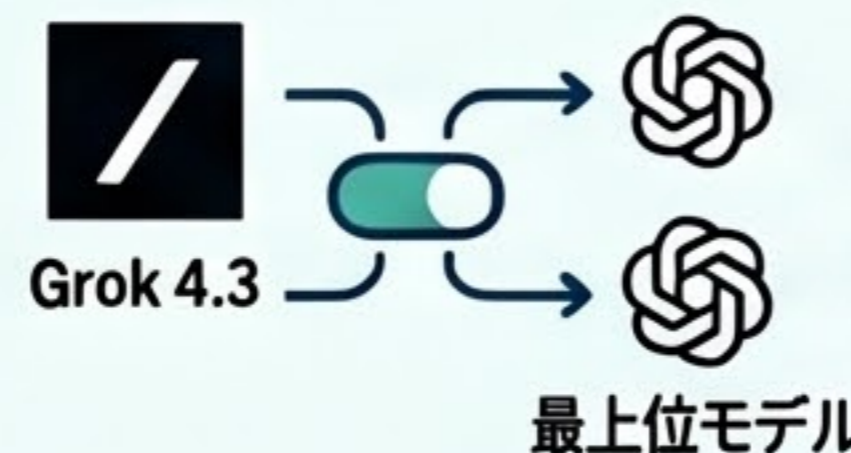
項目 (100万トークンあたり)	Grok 4.3	Grok 4.20 (旧価格)	削減率
入力トークン	\$1.25	\$2.00 - \$3.00	約37.5% - 58%
出力トークン	\$2.50	\$6.00 - \$15.00	約58.3% - 83%
キャッシュされた入力	\$0.20 - \$0.31	(新設)	-

最適なユースケースと導入戦略



コスト重視の大規模テキスト処理

大量の法的文書のレビューや、大規模なコードベースの分析など、コストとコンテキスト量の両立が必要なタスクに最適です。



ハイブリッドアーキテクチャによる最適化

単純なタスクをGrok 4.3に、極めて複雑なタスクを最上位モデルに振り分けることで、品質を維持しつつ通用コストを80%以上削減可能です。

国内外での迅速な採用事例

日本では株式会社SUPERNOVAの「Stella AI」が既に対応を開発するなど、エンタープライズ分野での採用が加速しています。