

DeepSeek V4 Pro (「DeepSeek V4」)

CAISI の評価

Felo AI

Overview

2026年4月にリリースされた DeepSeek V4 Pro は、オープンウェイト AI モデルの分野において、性能とコスト効率の新たな基準を打ち立てました。米国人工知能標準化・イノベーションセンター (CAISI) の評価では、最先端モデルから約 8 ヶ月遅れているとされていますが、コーディングや数学などの特定分野では、主要なクローズドソースモデルに匹敵、あるいはそれを凌駕する性能を公開ベンチマークで示しています[3][20]。このモデルの真の破壊力は、その性能だけでなく、Mixture-of-Experts (MoE) と独自のハイブリッドアテンション機構によって実現された、既存のフロンティアモデルの数分の一から数十分の一という圧倒的なコスト効率にあります[5][13]。DeepSeek V4 Pro の登場は、高性能 AI のコモディティ化を加速させ、AI 開発と応用の経済性を根本から変える可能性を秘めています。

詳細レポート

モデルの概要とアーキテクチャ

2026年4月24日、中国の AI 企業 DeepSeek は、オープンウェイトの Mixture-of-Experts (MoE) モデルである DeepSeek V4 シリーズを公開しました[5][8]。このシリーズは、性能とコストの異なる 2 つのモデルで構成されています[5]。

- DeepSeek V4-Pro: 総パラメータ数 1.6 兆、推論時のアクティブパラメータ数 490 億を誇る、現時点で最大級のオープンウェイトモデルです。前作 V3.2 の 2 倍以上の規模を持ちます[5][17]。
- DeepSeek V4-Flash: 総パラメータ数 2840 億、アクティブパラメータ数 130 億の、より軽量でコスト効率に優れたモデルです[5][8]。

両モデルとも 32 兆トークンのデータで学習され、最大 100 万トークンという広大なコンテキスト長をサポートしています[8][13]。また、MIT ライセンスの下で公開されており、誰でも無料でダウンロードして商用利用が可能です[5][18]。

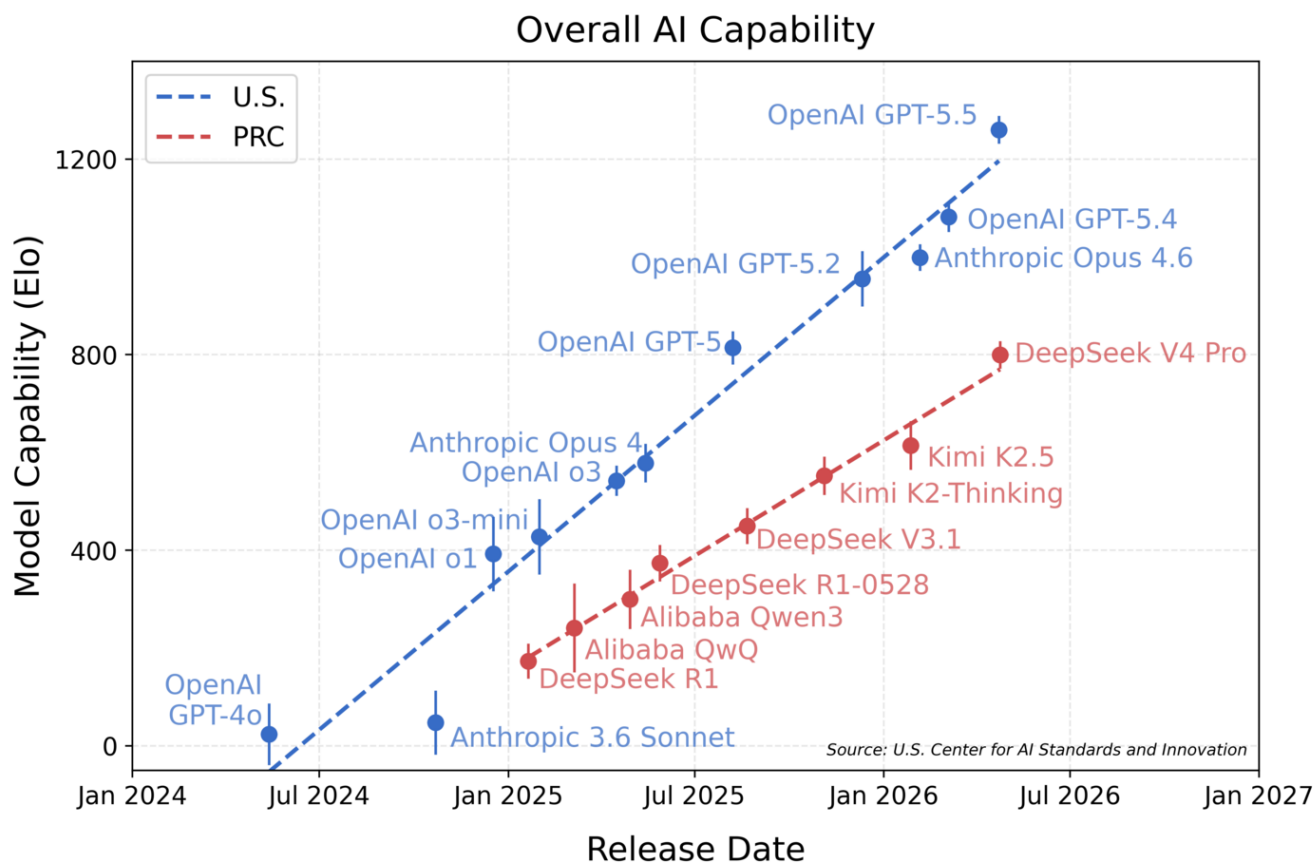
Hybrid Attention Architecture (ハイブリッドアテンション機構) V4 シリーズの画期的な効率性の核となるのが、新開発の「ハイブリッドアテンション機構」です[5][13]。これは、KV キャッシュを圧縮してスパースアテンションを行う「CSA (Compressed Sparse Attention)」と、より強力な圧縮を施す「HCA (Heavily Compressed Attention)」を組み合わせたものです[20]。この技術により、100 万トークンのコンテキスト長設定時において、前世代の DeepSeek-V3.2 と比較して、推論に必要な計算量 (FLOPs) を約 73%、KV キャッシュサイズを約 90%削減することに成功しました[20][24]。この効率化が、理論上だけでなく、実用的なコストで 100 万トークンのコンテキストウィンドウを利用可能にしています[13]。

性能評価：CAISI と公開ベンチマークの比較

DeepSeek V4 Pro の性能については、評価機関やベンチマークによって見解が分かれており、多角的な分析が必要です。

CAISI による評価 米国国立標準技術研究所 (NIST) 内の人工知能標準化・イノベーションセンター (CAISI) は、2026 年 4 月に DeepSeek V4 Pro を評価しました[1][3]。その報告によると、V4 Pro はこれまでに評価された中国製 AI モデルの中で最も高性能であるとしながら

も、全体的な能力では最先端の米国製モデルに約 8 ヶ月遅れていると結論付けています[3][7]。この評価は、サイバー、ソフトウェア工学、自然科学、抽象的推論、数学といった 5 つのドメインにわたる非公開ベンチマークを含んでおり、DeepSeek 社自身の評価よりも低い結果となっています[3][9]。



公開ベンチマークにおける性能一方で、広く公開されている標準的なベンチマークでは、DeepSeek V4 Pro は最先端モデルに肉薄する、あるいはそれを上回る非常に高い性能を示しています。特にコーディングと数学の分野でその能力が際立っています。

ベンチマーク	DeepSeek V4-Pro	Claude Opus 4.7	GPT-5.5	Gemini 3.1 Pro
SWE-bench Verified	80.6% [10]	87.6% [18]	(推定 ~85%) [18]	80.6% [18]

ベンチマーク	DeepSeek V4-Pro	Claude Opus 4.7	GPT-5.5	Gemini 3.1 Pro
HumanEval	91.8% [18]	93.4% [18]	91.5% [18]	88.2% [18]
AA Intelligence Index	54 点 [18]	57 点 [18]	60 点 [18]	57 点 [18]
FrontierMath Tier 4	28.4% [18]	22.9% [18]	39.6% [18]	16.7% [18]
Codeforces Rating	3,206 [20]	-	(GPT-5.4: 3,168) [20]	(3,052) [20]

- コーディング: ソフトウェア開発タスクのベンチマーク「SWE-bench Verified」では、Gemini 3.1 Pro と同等の 80.6%を達成し、Claude Opus 4.6 (80.8%) に迫っています [18][20]。コーディング競技ベンチマーク「Codeforces」では、GPT-5.4 を上回るレーティングを記録しました[20]。
- 数学: 難解な数学の問題解決能力を測る「FrontierMath Tier 4」では、Claude Opus 4.7 や Gemini 3.1 Pro を上回るスコアを叩き出しています[18]。
- 総合知能: 総合的な知能を測る「AA Intelligence Index」では 54 点を記録し、Claude Opus 4.7 や Gemini 3.1 Pro の 57 点、GPT-5.5 の 60 点にわずかに及ばないものの、オープンウェイトモデルとしては Kimi K2.6 に次ぐ第 2 位のスコアです[18]。

評価の乖離に関する分析 CAISI の評価と公開ベンチマークの結果に乖離が見られるのは、評価の尺度が異なるためと考えられます。CAISI が用いる非公開ベンチマークは、まだ一般的に対策されていない、より困難なタスクを含んでいる可能性があります。一方で、V4 Pro は特定のタスク（特にコーディングと数学）に最適化されており、それらの分野で卓越した性能を発揮します。このことは、モデルの能力を単一の指標で評価することの難しさを示唆しています。

価格破壊：経済的インパクト

DeepSeek V4 シリーズの最も衝撃的な側面は、その圧倒的な価格性能比です[13][18]。

API 価格設定 V4 シリーズの API 価格は、同等性能の競合モデルと比較して桁違いに安価です [13]。

モデル	入力価格/1M トークン	出力価格/1M トークン
DeepSeek V4-Flash	\$0.14 [18]	\$0.28 [13]
DeepSeek V4-Pro	\$1.74 [12]	\$3.48 [12]
Gemini 3.1 Pro	\$2.00 [18]	-
Claude Opus 4.7	\$5.00 [18]	\$75.00 [13]
GPT-5.5	\$5.00 [18]	\$75.00 [13]

注: 価格はリリース時期やプロモーションにより変動する可能性があります。V4-Pro は 2026 年 5 月 31 日まで 75%割引が適用されていました[13]。

例えば、月間 1 億トークンを出力するユースケースを想定すると、Claude Opus 4.7 では月額 7,500 ドルかかる計算が、DeepSeek V4-Pro ではわずか 348 ドルで済み、実に 22 倍ものコスト差が生まれます[13]。

コストパフォーマンス分析 性能をコストで割った「1 ドルあたりの知能」で比較すると、V4 シリーズの優位性はさらに明白になります[18]。

V4-Flash は同価格帯のクラウドモデルの 11 倍のコストパフォーマンスを誇る[18]。

この価格設定は、これまでコストの観点から高性能 AI の利用をためらっていた多くの開発者や企業にとって、大きな福音となります。特に、大量のデータを処理するバッチ処理や、低

レイテンシーが求められない分析タスクにおいて、V4 シリーズは新たな標準となる可能性があります。

市場への影響と戦略的意義

DeepSeek V4 のリリースは、単なる新モデルの登場以上の意味を持ちます。OpenAI が GPT-5.5 をリリースしたわずか翌日に、ほぼ同等の性能を持つモデルをオープンソースかつ圧倒的な低価格で公開するという動きは、明らかに計算された戦略的なものと見なされています [18]。

この出来事は、以下の 3 つの事実を市場に突きつけました [18]。

1. 性能のキャッチアップ: 中国発のオープンウェイトモデルが、米国の最先端クローズドモデルの性能にほぼ追いついた。
2. 価格の破壊: 性能が同等でありながら、コストは数十分の一にまで低下した。
3. アクセスの民主化: MIT ライセンスにより、高性能 AI が誰でも自由に利用・改変できるようになった。

これは、米中の AI 開発競争が新たな局面に入ったことを示唆すると同時に、「高性能 AI は無料に近づく」という未来が、予想より早く現実になったことを意味します [18]。開発者はもはや、単一のプロプライエタリな API にロックインされる必要はなく、用途やコストに応じて最適なモデルを自由に選択し、あるいは自前でホスティングするという選択肢を手に入れました。

結論と今後の展望

DeepSeek V4 Pro は、CAISI の評価が示すように総合的な能力で最先端モデルにまだ及ばない部分があるものの、その評価だけでは本質を見誤る可能性があります。特定の専門分野における卓越した性能、そして何よりも市場の常識を覆すほどの圧倒的なコストパフォーマンスは、AI 業界に地殻変動をもたらすインパクトを持っています [3][13][18]。

このモデルの登場により、AI アプリケーション開発の焦点は、最高の性能を持つ単一のモデルをいかに利用するかという点から、タスクの要件に応じて性能とコストが最適なモデルをいかに組み合わせ、運用するかという点へとシフトしていくでしょう。DeepSeek V4 は、その変化を象徴する存在であり、今後の AI エコシステムの発展において重要な役割を担うことは間違いありません。

1. [CAISI Evaluation of DeepSeek V4 Pro](#)
2. [DeepSeek V4 Pro が今出たばかりだけど、Copilot っぽい ... - Reddit](#)
3. [CAISI Evaluation of DeepSeek V4 Pro](#)
4. [I ran a personal AI benchmark across 6 models, DeepSeek ...](#)
5. [DeepSeek が V4 を公開——米国 AI にとってこれは本当にヤバいのか？](#)
6. [DeepSeek V4 のベンチマーク！ : r/singularity](#)
7. [CAISI Evaluation of DeepSeek V4 Pro finds it to be on par ...](#)
8. [ついに「DeepSeek-V4」が登場、Claude Opus 4.6 を超える性能の ...](#)
9. [DeepSeek V4 Pro は Opus 4.6 や GPT-5.4 並みとされていますが](#)
10. [DeepSeek V4 Alters Everything We Knew About Price- ...](#)
11. [I just wrote a deep dive on this topic:](#)
12. [DeepSeek V4 Pro Benchmarks 2026: Scores, Rankings & ...](#)
13. [DeepSeek V4 Just Dropped — And It's the Cheapest Frontier ...](#)
14. [Deepseek V4 Pro はすごいよ。必要だったのはちゃんとした ... - Reddit](#)
15. [deepseek-ai/DeepSeek-V4-Pro · Technical Report Summary](#)
16. [DeepSeek V4 Pro \(Max\) - Intelligence, Performance & ...](#)
17. [DeepSeek V4 Pro: Model Overview, Features & ...](#)
18. [【完全解説】 DeepSeek V4 登場——オープンソースで GPT-5.2 ...](#)
19. [DeepSeek V4: The Open-Source Model That Rivals ...](#)
20. [Claude Opus 4.6 匹敵、100 万トークン対応「DeepSeek-V4」無償 ...](#)
21. [DeepSeek V4 Flash: Review, Pricing & When to Use It \(2026\)](#)
22. [DeepSeek V4](#)
23. [DeepSeek V4 Pro は Arena で拍子抜けだった（クラウドで集めた ...](#)

24. [DeepSeek V4 Review - by Barnacle Goose](#)
25. [ついに「DeepSeek-V4」が登場、Claude Opus 4.6 を超える ...](#)
26. [deepseek-ai/DeepSeek-V4-Pro - Hugging Face](#)
27. [DeepSeek V4: Features, Benchmarks, and Comparisons](#)
28. [DeepSeek-V4-Pro-Max Benchmarks, Pricing & Context ...](#)
29. [「DeepSeek-V4」登場 オープンながら“世界トップのクロースド ...](#)
30. [I reviewed how DeepSeek V4-Pro, Kimi 2.6, Opus 4.6, and ...](#)
31. [同等のベンチマーク付きの新しい Deepseek Pro の料金を ...](#)
32. [DeepSeek v4 Pro のレビュー能力凄いな。衝撃を受ける。](#)
33. [deepseek v4 pro ってコード書きには良いの？率直なレビューが聞き ...](#)
34. [deepseek.com についてカスタマーサービスのレビューをご覧ください](#)