



DeepSeek-V3.1 性能と評判調査報告書

概要

DeepSeek-V3.1は、2025年8月19日に中国のDeepSeek AI社によってリリースされた最新の大規模言語モデルです。685億パラメータのハイブリッドモデルとして、プログラミング分野で特に優秀な性能を示し、既存の主要モデルに対して革命的なコスト効率を実現しています。 ^[1] ^[2] ^[3]

公式技術仕様と性能

アーキテクチャと技術革新

DeepSeek-V3.1はハイブリッド推論アーキテクチャを採用し、単一モデル内で「Thinking（推論）モード」と「Non-thinking（通常）モード」の両方をサポートします。主要な技術仕様は以下の通りです： ^[4] ^[5]

- **パラメータ数**: 685億（総計）、37億（トークンあたりアクティブ）
- **コンテキスト長**: 128,000トークン（約400ページの書籍相当）
- **アーキテクチャ**: Mixture-of-Experts（MoE） + Multi-head Latent Attention
- **対応形式**: BF16、F8_E4M3、F32（ハードウェア最適化対応）
- **ライセンス**: MIT（商用利用可能なオープンソース） ^[3] ^[4]

公式ベンチマーク結果

公式発表および第三者検証によるベンチマーク結果では、以下の性能向上が確認されています： ^[2] ^[6] ^[1]

- **MMLU-Pro**: 75.9 → 81.2（+5.3ポイント改善）
- **GPQA**: 59.1 → 68.4（+9.3ポイント改善）
- **AIME**: 39.6 → 59.4（+19.8ポイント改善）
- **LiveCodeBench**: 39.2 → 49.2（+10.0ポイント改善）

専門家による第三者評価

AI研究機関の分析

Artificial Analysisによる独立評価では、DeepSeek-V3.1は推論モードで**Intelligence Index 60**を記録し、前世代のR1（59）をわずかに上回りました。しかし、Alibabaの最新モデルやOpenAIのGPT-OSSには及ばない結果となっています。^{[6] [7]}

VentureBeatの技術分析では、DeepSeek-V3.1の「サイレントローンチ戦略」が、製品の技術力に対する企業の自信を示していると評価されています。^{[8] [9]}

学術研究機関による評価

学術論文による比較研究では、**英中翻訳タスク**においてDeepSeek-V3がGPT-4oより高い品質スコアを達成する一方、GPT-4oが語彙の豊かさや構文の複雑さで優れていることが報告されています。^[10]

主要競合モデルとの性能比較

プログラミング能力比較

Aiderプログラミングベンチマークにおいて、DeepSeek-V3.1は****71.6%****のスコアを記録し、以下の主要モデルを上回りました：^{[9] [11] [12]}

- **Claude Opus 4:** 70.6%
- **GPT-4:** 約65%
- **DeepSeek R1:** 前世代モデル

コスト効率比較

最も注目すべき点は圧倒的なコスト効率です：^{[11] [3] [9]}

モデル	プログラミングタスク単価	コスト効率比
DeepSeek V3.1	\$1.01	基準
Claude Opus	約\$70	68倍高い
GPT-5	約\$60-70	60-70倍高い

数学・推論能力

Math 500テストでは、DeepSeek-V3.1が****90.2%****を記録し、GPT-4o（74.6%）を大幅に上回る結果を示しています。^[12]

ユーザーコミュニティの評価と評判

プログラミングコミュニティの反応

Reddit r/LocalLLaMAコミュニティでは、プログラミング能力について高い評価が集まっています。
特に：[13] [14] [15]

- コード生成精度の高さ
- デバッグ能力の優秀さ
- 700行レベルの長いコードでもエラーフリー[12]

ある開発者は「30時間使用した結果、最高のAIコーディングアシスタント」と評価しています。[15]

創作・対話能力への批判

一方で、クリエイティブライティングや対話能力については厳しい評価が多数報告されています：
[16] [17] [18]

- V3-0324と比較して大幅な品質低下
- 定型的で創造性に欠ける応答
- プロンプトの理解不足
- 「XではなくY」という回避的な応答パターン

Hugging Faceコミュニティの反応

Hugging Faceでは、リリース直後に**第4位**の人気モデルにランクインし、80,000以上のフォロワーを獲得しています。技術的な議論は活発である一方、応答品質については意見が分かれています。
[18] [19] [1]

特定タスクでの長所と短所

長所

1. プログラミング・コーディング

- Aider benchmark: 業界トップレベルの71.6%
- エラーフリーな長文コード生成
- 優秀なデバッグ・リファクタリング能力

2. 数学・論理的推論

- Math 500: 90.2% (GPT-4oの74.6%を大幅上回る)
- 複雑な物理シミュレーション問題の解決

3. ツール使用・エージェント機能

- SWE-bench Verified: 66.0
- Terminal-bench: 31.3
- 大幅に改善されたAPI連携能力

4. 翻訳能力

- 学術論文翻訳でGPT-4oより高品質
- 多言語対応の向上

短所

1. クリエイティブライティング

- 前バージョン（V3-0324）から大幅劣化
- 創造性・オリジナリティの不足

2. 対話・会話能力

- 定型的で自然さに欠ける応答
- プロンプトの細かいニュアンス理解不足

3. 機能制限

- 推論モードでのファンクション呼び出し未対応
- 一部のエージェントワークフローで制約

市場での位置付けと将来性

市場インパクト

DeepSeek-V3.1のリリースは**AI業界に甚大な影響**を与えました：[\[12\]](#) [\[20\]](#)

- **NVIDIA株価**: \$600億の下落を引き起こす要因の一つ
- **価格破壊**: 既存のAIサービス価格モデルに挑戦
- **オープンソースAIの新基準**: MIT ライセンスでの完全公開

エンタープライズ採用の可能性

企業レベル導入の優位性：[\[20\]](#) [\[21\]](#)

- **コスト削減**: 数百万ドル規模の運用費削減が可能
- **カスタマイズ性**: オープンソースによる柔軟な改良
- **インフラ制御**: 自社環境での完全運用可能

採用の課題：

- **ハードウェア要件**: 700GB、高メモリ容量GPU必要
- **技術サポート**: エンタープライズレベルサポート不足
- **コンプライアンス**: データ統制・バイアス対策の課題

競合他社への影響

OpenAI・Anthropicへの挑戦：^[12] ^[20]

- 従来のクローズドAIビジネスモデルに対する根本的挑戦
- API中心アプローチからオープンソース移行の加速
- イノベーションサイクルの高速化圧力

総合評価と結論

技術的成果

DeepSeek-V3.1はプログラミング・数学分野で業界最高水準の性能を実現し、同時に革命的なコスト効率（68倍の優位性）を達成しました。ハイブリッドアーキテクチャの成功により、単一モデルで多様なタスクに対応可能となっています。^[1] ^[3] ^[9]

市場ポジション

オープンソースAI分野の新リーダーとして確立され、エンタープライズ市場で既存プロプライエタリモデルに対する強力な代替選択肢を提供しています。特にコスト意識の高いスタートアップや研究機関にとって最適な選択肢となっています。^[22] ^[12]

将来性と課題

強み：

- 急速な開発ペース（V4の早期リリース予測）
- 中国AI技術力の世界的認知向上
- オープンソースエコシステムの拡大

解決すべき課題：

- クリエイティブ・対話能力の根本的改善
- エンタープライズサポート体制の構築
- 地政学的制約下でのグローバル展開

最終提言

DeepSeek-V3.1は技術・コスト効率で画期的な成果を達成し、AI業界の競争構造を根本的に変える可能性を秘めています。プログラミング・数学タスク中心の用途では即座に採用を検討する価値がある一方、創作・対話が重要な用途では慎重な評価が必要です。今後のアップデートによる改善と、エンタープライズ向けサポート体制の整備が、さらなる市場浸透の鍵となるでしょう。

✻

2. <https://dev.to/czmilo/deepseek-v31-complete-evaluation-analysis-the-new-ai-programming-benchmark-for-2025-58jc>
3. https://www.reddit.com/r/LocalLLaMA/comments/1mwexgd/deepseek_v31_thinking_aggregated_benchmarks_vs/
4. <https://huggingface.co/deepseek-ai/DeepSeek-V3.1>
5. https://wandb.ai/wandb_fc/genai-research/reports/DeepSeek-V3-1-tutorial-How-to-use-the-open-source-LLM-with-Python-and-W-B--VmlldzoxNDA4NTA2NQ
6. <https://bdtechtalks.substack.com/p/deepseek-v31-is-here-heres-what-you>
7. <https://api-docs.deepseek.com/news/news250821>
8. <https://openrouter.ai/deepseek/deepseek-chat-v3.1:thinking>
9. <https://zenn.dev/czmilo/articles/d6156037b0dfbe>
10. <https://www.youtube.com/watch?v=i0erK5-8vZI>
11. <https://www.reuters.com/world/china/chinese-ai-startup-deepseek-releases-upgraded-model-with-domestic-chip-support-2025-08-21/>
12. <https://www.implicator.ai/deepseeks-v3-1-lands-quietly-and-pushes-the-open-weight-fight/>
13. <https://the-decoder.com/deepseeks-first-hybrid-model-v3-1-outperforms-its-r1-reasoning-model-on-benchmarks/>
14. <https://www.linkedin.com/pulse/new-leader-open-source-ai-deepseek-v3-benchmarking-1sryc>
15. <https://venturebeat.com/ai/deepseek-v3-1-just-dropped-and-it-might-be-the-most-powerful-open-ai-yet/>
16. <https://mpgone.com/deepseek-v3-1-0324-redefining-efficiency-in-large-language-models/>
17. https://www.reddit.com/r/LocalLLaMA/comments/1jkhk6/mismatch_between_official_deepseekv31_live_bench/
18. https://www.reddit.com/r/LocalLLaMA/comments/1mw3c7s/deepseekaideepseekv31_hugging_face/
19. <https://arxiv.org/pdf/2412.19437.pdf>
20. <https://artificialanalysis.ai/models/deepseek-v3-0324>
21. <https://github.com/deepseek-ai/DeepSeek-V3>
22. <https://www.bittrue.com/blog/deepseek-v3-1-vs-chatgpt-5>
23. <https://docsbot.ai/models/compare/deepseek-v3/llama3-1-70b-instruct>
24. <https://www.creolestudios.com/deepseek-v3-1-vs-gpt-5-vs-claude-4-1-compared/>
25. <https://docsbot.ai/models/compare/deepseek-v3/claude-3-opus>
26. <https://docsbot.ai/models/compare/deepseek-v3/llama-3-1-405b-instruct>
27. <https://www.prompthackers.co/compare/deepseek-v3/gpt-5>
28. <https://artificialanalysis.ai/models/comparisons/deepseek-v3-1-reasoning-vs-claude-3-opus>
29. <https://www.digitalocean.com/community/tutorials/deepseek-vs-llama-chatbot-gradientai-platform>
30. <https://artificialanalysis.ai/models/comparisons/deepseek-v3-1-vs-gpt-5>
31. <https://sourceforge.net/software/compare/Claude-Opus-4.1-vs-DeepSeek-V3.1/>
32. <https://www.glbgt.com/blog/deepseek-v3-vs-gpt4-llama31-comparison/>
33. <https://sourceforge.net/software/compare/DeepSeek-V3.1-vs-GPT-5/>

34. https://www.reddit.com/r/LocalLLaMA/comments/1mv7zdl/deepseek_v31_is_bad_at_creative_writing_way_worse/
35. https://www.reddit.com/r/ChatGPTCoding/comments/1hvavio/the_performance_of_the_deepseek_v3_model_must_be/
36. <https://bdtechtalks.com/2025/08/20/deepseek-v3-1/>
37. <https://www.youtube.com/watch?v=cvH6xjpT1PA>
38. <https://www.reddit.com/r/DeepSeek/>
39. <https://www.taneyats.com/entry/deepseek-v31-release>
40. <https://www.youtube.com/watch?v=ufXZl6aqOU8>
41. <https://www.reddit.com/r/DeepSeek/new/>
42. <https://aclanthology.org/2025.pslt-1.1.pdf>
43. <https://huggingface.co/deepseek-ai/DeepSeek-V3.1-Base>
44. <https://www.bentoml.com/blog/the-complete-guide-to-deepseek-models-from-v3-to-r1-and-beyond>
45. <https://huggingface.co/deepseek-ai/DeepSeek-V3-0324>
46. <https://ossels.ai/deepseek-v3-1-ai-model-update/>
47. <https://api-docs.deepseek.com/updates>
48. <https://xpert.digital/en/deepseek-v3-1/>
49. <https://www.bittrue.com/blog/deepseek-v3-1-game-changing-ai>
50. <https://scanx.trade/stock-market-news/global/deepseek-unveils-v3-1-ai-model-with-enhanced-capabilities/17157668>
51. <https://centific.com/blog/deepseek-v3-makes-enterprise-ai-smarter-faster-and-more-affordable>
52. <https://www.mitrade.com/insights/news/live-news/article-3-1059503-20250822>
53. <https://www.ainvest.com/news/deepseek-unveils-v3-1-ai-model-enhanced-reasoning-pricing-strategy-2508/>
54. <https://www.janushenderson.com/en-se/investor/article/deepseek-marks-a-potential-shift-in-the-ai-competitive-landscape/>
55. <https://ppl-ai-code-interpreter-files.s3.amazonaws.com/web/direct-files/bd1a24b10406c5066db117a36079b590/fc5a9887-e342-441c-a54c-3c08d6b59283/26ff6496.csv>
56. <https://ppl-ai-code-interpreter-files.s3.amazonaws.com/web/direct-files/bd1a24b10406c5066db117a36079b590/fc5a9887-e342-441c-a54c-3c08d6b59283/99a12bbc.csv>
57. <https://ppl-ai-code-interpreter-files.s3.amazonaws.com/web/direct-files/bd1a24b10406c5066db117a36079b590/fc5a9887-e342-441c-a54c-3c08d6b59283/247faa8a.csv>
58. <https://ppl-ai-code-interpreter-files.s3.amazonaws.com/web/direct-files/bd1a24b10406c5066db117a36079b590/fc5a9887-e342-441c-a54c-3c08d6b59283/406d619a.csv>
59. <https://ppl-ai-code-interpreter-files.s3.amazonaws.com/web/direct-files/bd1a24b10406c5066db117a36079b590/fc5a9887-e342-441c-a54c-3c08d6b59283/50e88182.csv>