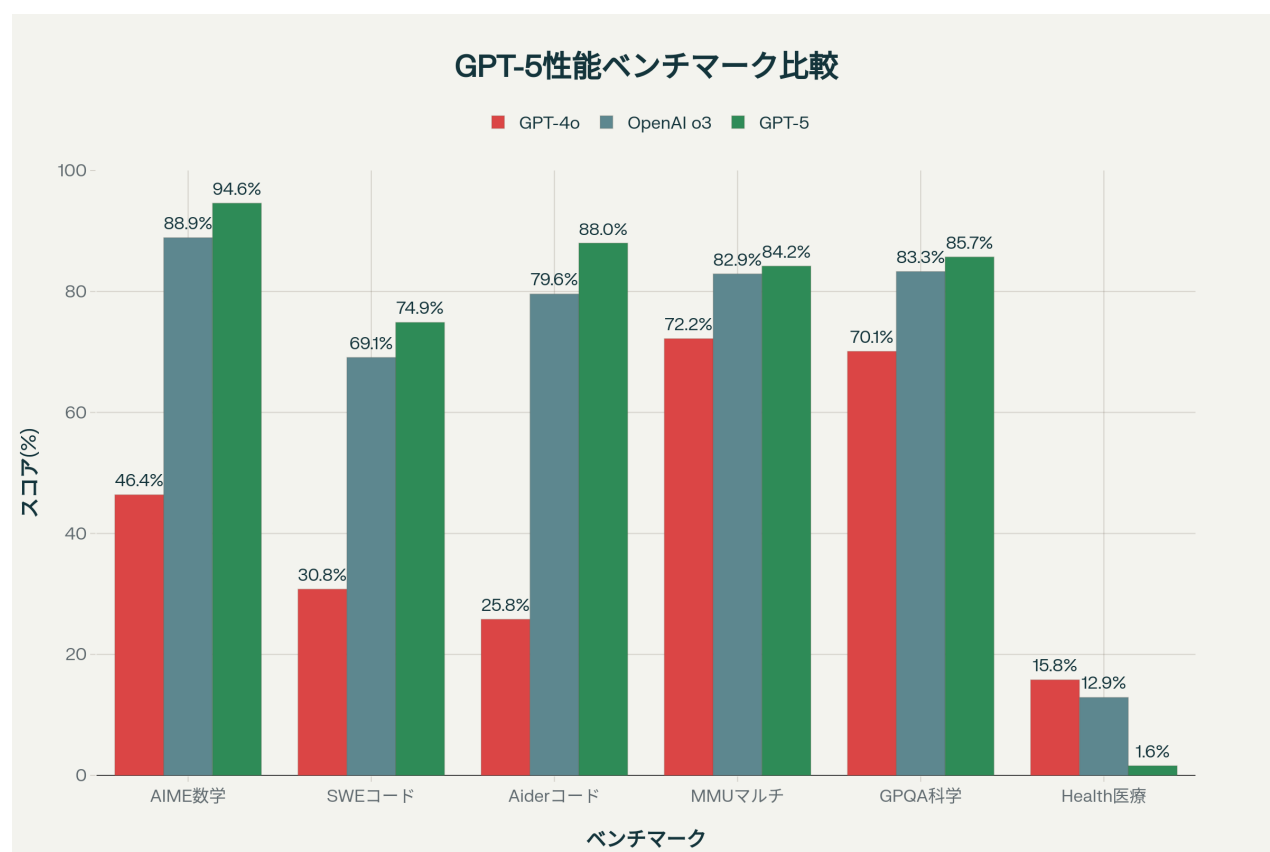




OpenAI「GPT-5」の現時点での評判・評価に関する包括的分析レポート

2025年8月7日にリリースされたOpenAIの大規模言語モデル「GPT-5」は、AI業界において注目すべき技術的マイルストーンとして位置づけられている。リリースから4日が経過した現時点で、同モデルに対する評判・評価は多様な観点から賛否両論に分かれており、その複雑な反応は現代AI開発の到達点と限界を同時に浮き彫りにしている。本レポートでは、技術ベンチマーク、専門家評価、ユーザー反応、企業導入事例を総合的に分析し、GPT-5の現在地と将来への含意を探る。^{[1] [2] [3]}



GPT-5の主要ベンチマークにおける性能比較 - GPT-4oとOpenAI o3との対比

GPT-5の技術的革新と統合アーキテクチャ

統合システムとしての画期的進化

GPT-5の最も重要な技術的特徴は、従来の「GPT-4o」と推論特化型「oシリーズ」を統合した**単一システムアーキテクチャ**にある。これまでユーザーは用途に応じてモデルを選択する必要があったが、GPT-5では「リアルタイムルーター」が質問の複雑さや文脈を自動判断し、最適なモデル（高速応答用の「GPT-5-main」または推論用の「GPT-5-thinking」）を動的に選択する。^{[2] [4] [5] [6]}

この統合により、OpenAI CEO Sam Altmanが言及する「博士号レベルの専門家」としての能力が実現されたとされる。特に注目すべきは、従来モデルと比較してGPT-5が示す**自律的推論能力**の向上で、問題解決において障壁に直面した際に代替案を模索する能力が評価されている。^{[11] [7] [8] [9]}

マルチモーダル機能と文脈処理の拡張

GPT-5は**400,000トークンの入力**と**128,000トークンの出力**に対応し、長文書の分析や複雑な対話においてより一貫した品質を維持できる。また、テキスト、画像、音声、動画の複合的な理解が可能となり、視覚情報を伴う推論精度が大幅に向上した。^{[10] [11] [12] [13] [14]}

ベンチマーク性能における圧倒的優位性

学術・技術ベンチマークでの世界最高水準達成

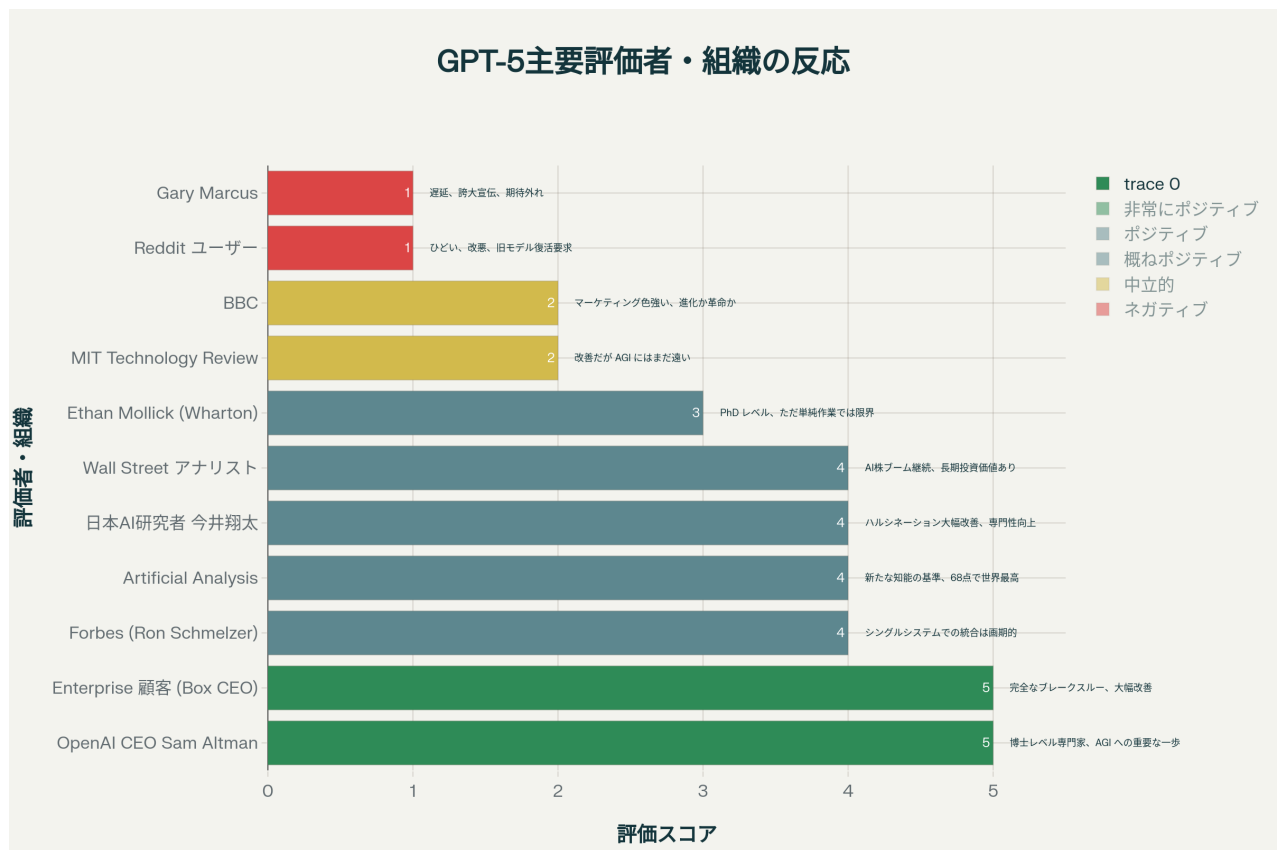
GPT-5は主要な技術ベンチマークにおいて、前世代モデルを大幅に上回る性能を示している。**AIME 2025**（数学オリンピック問題）では94.6%の正答率を記録し、ツールなしでの数学的推論能力が飛躍的に向上した。^{[10] [11] [15]}

コーディング分野では、**SWE-bench Verified**で74.9%（GPT-4oの30.8%から倍以上の向上）、**Aider Polyglot**で88%（60ポイント以上の向上）を達成し、「OpenAI史上最強のコーディングモデル」との評価を獲得している。^{[11] [12] [16] [10]}

ハルシネーション削減における劇的改善

最も重要な進歩の一つは、**事実誤認（ハルシネーション）の大幅削減**である。GPT-5はGPT-4oと比較して約45%、思考モードではOpenAI o3と比較して約80%の事実誤りを減少させた。特に医療分野では、HealthBench Hardでのハルシネーション率を1.6%まで低減し、GPT-4oの15.8%から劇的な改善を示した。^{[10] [17] [11] [12] [18] [19]}

専門家・業界による評価の二極化



GPT-5に対する主要評価者・組織の反応分布

高度に肯定的な評価群

企業レベルでの評価は概して非常にポジティブである。BoxのCEO Aaron Levieは「完全なブレークスルー」と評価し、複雑な文書における論理的推理能力の飛躍的向上を指摘している。AmgenやInditex等の大企業も、科学的精度と品質の向上を評価している。[\[7\]](#) [\[20\]](#) [\[21\]](#) [\[14\]](#) [\[16\]](#)

学術分野では、Wharton School のEthan Mollickが「GPT-5は単純にタスクを実行する、しばしば異常なことを」と述べ、その自律性と創造性を高く評価している。日本のAI研究者・今井翔太氏も「博士号レベルの専門性を持つという触れ込みに違わず、ハルシネーション大幅改善とプログラミング能力の飛躍的向上」を確認している。[\[22\]](#) [\[23\]](#)

懐疑的・批判的見解

一方で、一部の専門家からは冷静な評価も寄せられている。MIT Technology ReviewのClémentine Fourier (HuggingFace研究者) は、「高校生が中学生の問題を解くようなもの」として、ベンチマークの飽和を指摘している。また、Gary Marcusは「遅延、誇大宣伝、期待外れ」として、GPT-3からGPT-4ほどの劇的な飛躍がないと指摘している。[\[24\]](#) [\[25\]](#) [\[26\]](#)

ユーザー反応の複雑な分裂

初期ユーザー体験における混乱

GPT-5のローンチは、技術的成果とユーザー体験戦略の間に深刻な乖離を生み出した。最大の不満は、**従来モデル選択機能の廃止**である。有料プランユーザーは、慣れ親しんだGPT-4oやo3への直接アクセスを失い、GPT-5のルーターシステムに委ねることとなった。^{[27] [28] [29]}

RedditやX（旧Twitter）では「ひどい」「改悪」「旧モデル復活要求」等の批判が殺到し、特に創造的タスクにおける性能低下を指摘する声が目立った。一方で、「明らかに賢くなった」「信頼性が向上した」との肯定的評価も多数寄せられ、ユーザー層による評価の分裂が顕著となっている。^{[28] [23] [31]}

OpenAIの対応と改善措置

Sam Altman CEOは Reddit AMAにおいて、ローンチ時の問題を認め、**Plus ユーザーに対してGPT-4oへのアクセスを復活**させると表明した。また、透明性の向上とモデル切り替えの改善を約束し、ユーザーフィードバックを積極的に取り入れる姿勢を示している。^{[28] [29] [32]}

企業導入における強力な推進力

エンタープライズ市場での急速な浸透

Altman CEOは企業導入について「super strong」と評価し、GPT-5による「事業成長の大幅な加速」を予測している。現在**500万人の有料ユーザー**がChatGPTビジネス製品を利用しており、Microsoft、BNY Mellon、Morgan Stanley、SoftBank等の大企業が導入を進めている。^{[33] [20] [34] [21]}

特にMicrosoftは、GPT-5を**Copilot**や**Visual Studio Code**等の製品に統合し、開発者向けツールの性能向上を実現している。GleanのベンチマークでGPT-5がo3に対して83%の正答率を達成し、企業AIタスクでの優位性が証明されている。^{[14] [33]}

「オンデマンドソフトウェア時代」の到来

GPT-5の高度なコーディング能力により、非プログラマーでも複雑なアプリケーションを瞬時に生成する「**ヴァイブコーディング**」が現実化している。OpenAIのデモンストレーションでは、「英語話者向けフランス語学習アプリ」が数分で完成し、教育・業務効率化における革命的可能性を示唆している。^{[7] [35] [36]}

限界と技術的課題

基本的推論タスクでの継続的問題

技術的進歩にも関わらず、GPT-5は依然として**基本的な計数タスク**で問題を抱えている。

「blueberry」の文字数カウント等の単純な問題で誤答を生成することがあり、モデルファミリー内での一貫性の欠如が指摘されている。^[37]

また、GPT-5は**継続学習能力**を持たないため、2024年9月以降の情報については外部から提供されない限り対応できない。この制約は、AGI（汎用人工知能）への道のりがまだ長いことを示唆している。^{[8] [38] [25] [26]}

安全性における複雑な現実

OpenAIは「セーフコンプライエンス」技術を導入し、従来の単純拒否から安全性境界内での最大限有用な回答提供へとシフトした。しかし、内部テストでは一部カテゴリーで安全性の「統計的に有意な後退」も確認されており、安全性向上が必ずしも一様ではないことが明らかになっている。^[10]^[28]^[39]

競合環境と市場への影響

AI業界における競争激化

GPT-5のリリースは、Google、Anthropic、xAI等の競合他社に強いプレッシャーを与えている。特にAnthropicのClaude 4やGoogleのGemini 2.5 Proとの性能差は縮小しつつあり、「モデル至上主義」から「エコシステム統合力」への競争軸の移行が見られる。^[40]^[41]

金融市場の楽観的反応

Wall Streetは概してGPT-5を好意的に評価している。Palantir株が9%急騰し、SoftBankが13%上昇するなど、AI関連企業への投資熱が継続している。ただし、一部アナリストは「AIハイプサイクル」の終盤を警告し、実用化の遅れがシバブル崩壊を招く可能性も指摘している。^[42]^[43]

将来展望と戦略的含意

AGIへの段階的進歩

GPT-5は「AGIへの重要な一歩」として位置づけられているが、完全なAGI実現には依然として時間を要すると予測される。METR (Model Evaluation & Threat Research) の評価では、GPT-5の自律的タスク実行能力は2時間17分程度と推定され、破滅的リスクを伴う水準には到達していないとされる。^[7]^[8]^[44]

社会的・倫理的課題への対応

GPT-5の能力向上は、**知的労働の自動化加速**による雇用への影響、**情報生態系への脅威**、**AI格差の拡大**等の社会的課題を浮き彫りにしている。特に、高度な偽情報生成能力は民主的プロセスへの脅威となる可能性があり、継続的な監視と規制が必要となる。^[38]^[25]

結論

GPT-5は技術的ベンチマークにおいて疑いなく**新たな最高水準**を確立し、ハルシネーション削減、コーディング能力、マルチモーダル理解において前世代から大幅な進歩を遂げた。企業導入においては強い牽引力を示し、「**博士レベルAI**」としての価値を実証しつつある。^[1]^[10]^[11]^[21]^[14]

しかし、ユーザー体験戦略の失敗、基本的推論タスクでの継続的問題、そして一部専門家からの「期待ほどではない」との評価は、技術進歩と実用化の間に存在する複雑な現実を示している。^[27]^[24]^[28]

GPT-5の評判・評価は、現在のAI開発が「**技術的可能性**」と「**実用的価値**」、「**期待管理**」のバランスを如何に取るかという根本的課題を体現している。今後数カ月間のユーザー適応、競合対応、安全

性確保の取り組みが、GPT-5の長期的成功と社会への影響を決定することとなるであろう。 [42] [41]
[26]



1. <https://forbesjapan.com/articles/detail/81193>
2. <https://k-tai.watch.impress.co.jp/docs/news/2038013.html>
3. <https://www3.nhk.or.jp/news/html/20250808/k10014888861000.html>
4. <https://shift-ai.co.jp/blog/33397/>
5. <https://zenn.dev/schroneko/articles/introducing-gpt-5>
6. <https://gihyo.jp/article/2025/08/openai-gpt-5>
7. <https://www.cnbc.com/2025/08/07/openai-launches-gpt-5-model-for-all-chatgpt-users.html>
8. <https://www.cnn.com/2025/08/07/tech/openai-gpt-5-chatgpt-launch>
9. <https://yorozuipsc.com/uploads/1/3/2/5/132566344/6e69cc9c7a1a4a24ee68.pdf>
10. <https://openai.com/ja-JP/index/introducing-gpt-5/>
11. <https://openai.com/index/introducing-gpt-5/>
12. <https://www.vellum.ai/blog/gpt-5-benchmarks>
13. <https://www.iamdave.ai/blog/10-things-you-can-do-with-gpt-5/>
14. <https://www.glean.com/blog/open-ai-gpt-5>
15. <https://openai.com/index/introducing-gpt-5>
16. <https://openai.com/index/introducing-gpt-5-for-developers/>
17. <https://wandb.ai/byyoung3/ml-news/reports/GPT-5-Benchmark-Scores---VmlldoxMzkwMTYyMg>
18. <https://www.notta.ai/blog/gpt-5-complete-guide>
19. <https://momo-gpt.com/column/gpt-5thinking/>
20. <https://openai.com/index/gpt-5-new-era-of-work/>
21. <https://www.investing.com/news/stock-market-news/openais-altman-sees-strong-enterprise-adoption-of-gpt5-4180683>
22. <https://www.oneusefulthing.org/p/gpt-5-it-just-does-stuff>
23. <https://note.com/chocolats/n/nd5d1aecc2935>
24. <https://garymarcus.substack.com/p/gpt-5-overdue-overhyped-and-underwhelming>
25. https://note.com/apgd110_en/n/n9d7669585a41
26. <https://www.technologyreview.com/2025/08/07/1121308/gpt-5-is-here-now-what/>
27. <https://finance.yahoo.com/news/openai-gpt-5-met-mixed-212741623.html>
28.
30. https://www.reddit.com/r/ChatGPTPro/comments/1mk84vx/gpt5_thoughts/
31. <https://xenospectrum.com/openai-gpt5-user-dissatisfaction-gpt4o-return/>
32. <https://www.ainvest.com/news/gpt-5-rocky-start-user-discontent-spurs-return-gpt-4-openai-promises-improvements-2508/>
33. <https://news.microsoft.com/source/features/ai/openai-gpt-5/>

34. <https://openai.com/ja-JP/index/gpt-5-new-era-of-work/>
35. <https://www.wired.com/story/openais-gpt-5-is-here/>
36. <https://www.latent.space/p/gpt-5-review>
37. https://biggo.jp/news/202508090742_GPT-5_Launch_Reveals_OpenAI_Struggles
38. <https://xpert.digital/ja/gpt5が登場/>
39. <https://japan.cnet.com/article/35236488/>
40. <https://www.folio3.ai/blog/gpt-4-vs-gpt-5/>
41. <https://www.forbes.com/sites/ronschmelzer/2025/08/08/what-does-openais-gpt-5-mean-in-the-race-for-ai-model-supremacy/>
42. <https://ts2.tech/en/ai-stocks-frenzy-mega-deals-gpt-5-hype-and-bubble-warnings-fuel-aug-9-10-market-buzz/>
43. <https://www.barrons.com/articles/gpt5-ai-stock-market-other-things-to-know-today-19ae81c6>
44. <https://www.lesswrong.com/posts/SuvWoLaGiNjPDcA7d/metr-s-evaluation-of-gpt-5>