

# 2025年8月7日にリリースされた OpenAI の新しい大規模言語モデル「GPT-5」に関する現時点での評判・評価

Grok-4

## GPT-5 の概要とリリース背景

OpenAI は 2025 年 8 月 7 日に GPT-5 をリリースしました。これは前世代の GPT-4 シリーズや o3 を統合した新しい大規模言語モデルで、自動ルーティング機能(クエリに応じて高速モードと思考モードを切り替える)を搭載し、コーディング、推論、正確性で改善を謳っています。API 価格は入力トークンあたり\$1.25/百万と低価格化され、企業ユーザー向けの活用が強調されています。しかし、リリース直後からハイプの過大さと旧モデルの廃止が議論を呼んでいます。以下に、現時点(2025年8月10日)での評判・評価を、肯定的・否定的・中立的観点からまとめます。情報源は主にニュースメディア、テックレビュー、X(旧 Twitter)のユーザー投稿に基づきます。

## 肯定的な評価

- **性能向上のポイント:** ベンチマークでコーディング(SWE-bench Verified: 74.9%)や医療推論(46.2%)で前モデルを上回り、ハルシネーション率が45%低減。思考モードでは出力トークンを50-80%削減しつつ優位。速さと正確性が日常業務で生産性を高めるとの声が多い。
- **コーディングと実務活用:** 開発者レビューでは、複雑なプロトタイプを1時間で作成可能で、UI クローンやバックエンドの自動スケールリングが「人間レベルに近い」。長文コンテキスト(400kトークン)の扱いが優秀で、長期セッションでの一貫性が向上。企業(例: Box, Microsoft)ではドキュメント分析でブレークスルー。
- **価格とアクセシビリティ:** API が安価で、無料ユーザーも利用可能。開発者コミュニティでは「インテリジェンス per dollar」が増加と好評。一般ユーザー向けに「PhD レベルの専門家」との宣伝が一部で響いている。
- **全体的な印象:** 「ベストオーバーオールモデル」で業界の基準を上げたとの意見。ニッチなトピックでの議論が自然で、速さがフローを維持。

## 否定的な評価

- **ハイプの失望:** Sam Altman CEO の事前宣伝(「AGIに近い」)が過大で、実際はインクリメンタル改善のみ。ベンチマークで SimpleBench 56.7%(5位)や ARC-AGI で期待外れ、Claude 4 Opus に劣る部分あり。「ダウングレード」と呼ぶユーザーも。
- **旧モデルの廃止とユーザー不満:** GPT-4o の廃止で Reddit スレッドが 3000 アップボート超えの請願。モデルルーターが不便で、応答が短く汎用的になり、クリエイティブライティングや研究で弱い。知識カットオフが 2024 年で古く、視覚理解やチェスルールで失敗報告。「シュリンクフレーション」(機能縮小)と批判。
- **パフォーマンスの限界:** ハルシネーションは減ったが残存。感情タスクや微妙な指示追従で GPT-4.5 に劣り、過信によるエラーが目立つ。ライブデモの失敗や誤ラベルチャートで信頼性疑問視。
- **業界全体への影響:** スケーリング神話の限界を示唆。Arizona State 大学の研究で「チェーン・オブ・ソート推論が訓練分布外で崩壊」と指摘され、AGI への道が遠いとの声。OpenAI の市場信頼が急落(75%→14%)。

## 中立的・混合的な評価

- **比較と競争:** Claude 4 Opus がコーディングで優位、Grok 4 が一般分析で上回るとの比較レビュー。GPT-5 は「速さと信頼性のバランスが良い」が、専門モード(Pro)未評価でポテンシャル不明。xAI の Grok 4 無料化で競争激化。
- **ユーザーアタッチメント:** 旧モデルへの強い愛着が顕在化。Altman は「AI 依存のリスク」を指摘し、長期ウェルビーイングを考慮した設計を強調。安全性で METR 評価は中間(1-4.5 時間ホライゾン)。
- **将来性:** 企業向けは好調だが、一般ユーザー調整中。オープンソース批判 (GPT-OSS が阉割版)もあり、コミュニティ駆動 AI(例: SenticentAGI)の優位を指摘する声。全体として「実用的アップデート」だが、ハイプが裏目に出た形。

カテゴリ	肯定的ポイント	否定的ポイント	スコア例(ベンチマーク)
コーディング	74.9% (SWE-bench)	誤り修正が必要	88% (Aider polyglot)

カテゴリ	肯定的ポイント	否定的ポイント	スコア例(ベンチマーク)
推論/正確性	ハルシネーション 45% 減	分布外で崩壊	42.0% (Humanity's Last Exam Pro)
速度/価格	秒単位応答、安価 API	ルーター不便	\$1.25/M 入力トークン
全体評価	生産性向上	期待外れ	5-10%改善 (PhD-level タスク)

総括として、GPT-5 は技術的に進化しているものの、ハイプとユーザー体験のギャップが目立ち、評価は分かれています。開発者/企業は歓迎、一般ユーザーは旧モデル復活を望む傾向です。今後、フィードバックに基づく調整が鍵となりそうです。

33 X posts

20 web pages