GPT-5シリーズ性能比較レポート

Claude Opus 4.1

OpenAIが2025年8月7日に発表したGPT-5、GPT-5 Thinking、GPT-5 Proの3モデル TechCrunch +3 は、**従来のベンチマークで測定困難だった信頼性問題を80%改善し、プログラミング分野では他モデルを20ポイント以上上回る革新的な性能**を実現した。 OpenAI +5 特筆すべきは、これらが独立したモデルではなく統合システムとして機能し、リアルタイムルーターが自動的に最適なモデルを選択する OpenAI 点である。 OpenAI +2 料金設定も前世代比50%削減と競争力が高く、無料ユーザーにも推論機能を初めて提供する画期的な展開となっている。 Il Sole 24 Ore +3

この統合アプローチにより、簡単な質問には高速で回答し、複雑な問題には深い推論を適用する最適化が実現されており、 OpenAI AI活用における新たなパラダイムシフトの始まりを示している。特に開発者向けの性能向上は劇的で、従来「不可能」とされていた複雑なWebアプリケーションの一発生成や、大規模リポジトリでのエンドツーエンドデバッグが実現可能になった。 OpenAI OpenAI

公式発表と基本概要の確認結果

全3モデルの存在が正式に確認された。 2025年8月7日午前10時（太平洋標準時）、OpenAIが「LIVE5TREAM」と題した公式ライブストリームでGPT-5シリーズを発表し、同日中にウェブサイトとAPIで利用開始となった。 TechCrunch +8

GPT-5は統合システムの中核として、ChatGPTの新しいデフォルトモデルに設定され、無料ユーザーを含む全ユーザーが利用可能 TechCrunch となった。 OpenAI VentureBeat リアルタイムルーターが質問の複雑さや文脈を判断し、高速モデルと推論モデルを自動選択する革新的アーキテクチャを採用している。 OpenAI +2

GPT-5 Thinkingは複雑な問題に対してより長時間思考する深い推論モデルで、ユーザーが「よく考えて」と指示するか明示的に選択することで起動する。 OpenAI greeden Inc. 有料ユーザーがより高い利用制限で使用でき、内部的には「gpt-5-thinking」として分類される。 OpenAI +3

GPT-5 ProはGPT-5 Thinkingの拡張版で、並列テストタイム計算を使用してさらに深い推論を実行する。 OpenAI +2 月額200ドルのProプラン限定で提供され、 Il Sole 24 Ore 研究グレードのインテリジェンスを提供する最上位モデル OpenAI として位置づけられている。 OpenAI +2

技術仕様とアーキテクチャの比較分析

統合アーキテクチャによる技術革新が最大の特徴である。従来の単一モデル方式から、推論モデルと非推論モデル、リアルタイムルーターの3要素で構成されるシステムに進化した。 OpenAI +3

コンテキスト処理能力では、GPT-5が400,000トークン（入力272,000、出力128,000）を処理可能で、 simonwillison Botpress 競合のClaude系の200,000トークンを大幅に上回る。 Bind AI IDE +3 一方、GPT-4.1の100万トークン Medium OpenAI には及ばないものの、実用的な範囲で最適化されている。

マルチモーダル対応は現時点でテキストと画像に対応し、 simonwillison 音声処理は構造的に対応可能だが未実装となっている。 Simon Willison 動画処理についても技術的基盤は整っているが、実装は今後の課題である。

パラメータ数は全シリーズで未公開だが、業界推定では3-5兆パラメータとされている。一部で69兆パラメータとの憶測もあるが、信頼性は低い。 LifeArchitect.ai 重要なのは、パラメータ数よりも統合システムによる効率的な計算資源の活用である。

APIバリエーションとして、用途別に最適化された3つのサイズを提供：gpt-5（標準版、$1.25/$10.00）、gpt-5-mini（中間版、$0.25/$2.00）、gpt-5-nano（軽量版、$0.05/$0.40）。

Botpress各バージョンとも100万トークンあたりの価格設定 OpenAIとなっている。 OpenAI

ベンチマーク性能データの定量的比較

数学的推論能力で革命的な進歩を遂げた。GSM8Kベンチマークで、基本設定では71.0%だったスコアが、thinking機能使用時には99.6%に跳上した。さらに注目すべきは、AIME 2025でGPT-5 Proがツール使用により**史上初の100%達成**を記録した OpenAIことである。 vellum

プログラミング能力では圧倒的優位性を確立した。SWE-bench Verifiedで74.9%を記録し、 OpenAIClaude 4の72.7%、GPT-4oの52%を大幅に上回った。 NBC News +5Aider Polyglotコーディングテストでは、基本版26.7%からthinking版88%への劇的な向上 OpenAIを示している。 OpenAI +3

科学的推論分野でも顕著な改善が見られた。GPQA Diamondベンチマークで、GPT-5が77.8%、GPT-5 Proがツール使用時89.4%を記録し、 OpenAIGPT-4oの70.1%を大幅に超越している。 Getpassionfruit +3

信頼性指標で特筆すべき改善を実現した。**幻覚率を従来モデルの5分の1以下**（4.8%対22.0%）に削減し、医療分野のHealthBench Hardでは、エラー率を1.6%まで低下させた（ OpenAIGPT-4oは15.8%）。 OpenAI +3

総合性能をMMLUで評価すると、GPT-5が87.1%、GPT-5 Proが推定90.2%を記録し、 OpenAI Claude 4の86.5%、Gemini 2.5 Proの85.8%を上回る性能を示している。 Bind AI IDE OpenAI

ターゲットユーザーと想定ユースケースの分析

開発者コミュニティがGPT-5の最大の受益者となっている。複雑なWebアプリケーションの一発生成、大規模リポジトリでのエンドツーエンドデバッグ、依存関係競合の自動解決など、従来「不可能」とされていたタスクが実現可能になった。 Substack +3Cursor、Windsurf、Vercelなどの開発ツールプロバイダーが「使用した中で最も賢いモデル」と評価 OpenAIしている。 OpenAI

企業向け業務自動化では、40以上の職業分野で専門家レベルの性能を発揮する。 OpenAI特にGPT-5 Proは並列テストタイム計算により、法務、物流、営業、エンジニアリングの各分野で人間専門家に匹敵する支援を提供する。 OpenAI TechCrunchMicrosoft 365 Copilot、GitHub Copilot、Azure AI Foundryとの統合により、エンタープライズ環境での実用性が大幅に向上している。 Microsoft News +4

一般ユーザー向け対話システムとして、無料ユーザーにも推論機能を初めて提供する画期的な展開となった。 Il Sole 24 Ore +4Google CalendarやGmail統合、カスタマイズ可能な性格設定（Cynic、Robot、Listener、Nerd）により、パーソナライズされた体験を実現している。 OpenAI

研究者と学術機関向けでは、272,000トークンの入力制限により大規模文書の並列分析処理が可能になった。 OpenAI +3実時間での研究文献調査機能と組み合わせ、研究プロセスの効率化に貢献している。

応答品質と機能面での具体的違い

プログラミング支援での質的変化が最も顕著である。単にツールを使用するのではなく、ツールと共に「思考」する能力を獲得した。並列ツール呼び出しの知的実装により、65%から72%へのソフトウェアエンジニアリング自動化進展を実現している。 OpenAI +2美的感覚を持ったUI生成（スペーシング、タイポグラフィ、余白の理解）や、プロダクション準備完了レベルのコード生成が可能になった。 OpenAI +2

推論の質と論理的一貫性では、専門家から「劇的な飛躍ではないが、一貫して有能で、ほとんど失敗しない」との評価を受けている。 simonwillison +2MIT Technology Reviewは「o1が技術的進歩

だったのに対し、GPT-5は洗練されたプロダクト」と評価している。 Simon Willison MIT Technology Review

創造性と文章生成では分野による差が明確になった。技術文書、API文書、構造化コンテンツでは非常に優秀だが、クリエイティブライティングや個人的なトーン維持では限界がある。「OpenAILinkedIn-slop」スタイルになりがちという指摘もあり、GPT-4.5やClaude 4の方が創造的文章で優位性を持つ。 Substack latent

応答速度の最適化では、リアルタイムルーターによる自動切り替えで、簡単な質問には高速レスポンス、複雑な問題には深い推論という最適なバランスを実現している。 OpenAI OpenAIAPI版では4段階の推論レベル（minimal、low、medium、high）から選択可能で、コストと品質のトレードオフを柔軟に調整できる。 OpenAI +2

アクセス方法と料金体系の詳細

攻撃的な価格設定により市場の競争構造を変革した。GPT-5のAPI利用料金は入力$1.25、出力$10.00（100万トークンあたり） OpenAIで、GPT-4oの入力$2.50から50%削減を実現している。 simonwillison +3競合のClaude Opus 4（$15.00/$75.00） Anthropicと比較すると圧倒的な価格競争力を持つ。 Platformer

多様なアクセス経路を提供している。OpenAI APIでは3つのサイズバリエーション（gpt-5、gpt-5-mini、gpt-5-nano）を提供し、用途に応じた最適化が可能 OpenAIである。 simonwillison +2 ChatGPTでは無料プランでも限定的なアクセスが可能で、Plus（$20/月）、Pro（$200/月）、Team（$25-30/月/ユーザー）、Enterprise（推定$60-100/月/ユーザー）の階層設定 TechCrunchとなっている。 Il Sole 24 Ore +2

トークンキャッシング機能により、数分以内の再利用時に90%の入力トークン割引が適用される。 simonwillisonチャットUIでの大幅なコスト削減効果があり、実用的な運用コストをさらに下げる効果がある。 Simon Willison

企業向けプランでは、Microsoft統合によりエンタープライズグレードのセキュリティとcompliance機能を提供している。 OpenAITeam、Enterprise、Eduプランでは管理機能、セキュリティ機能、内部ツールとの連携が可能である。 Microsoft News +3

統合的な性能・機能・コスト・用途分析

開発者向け用途では圧倒的な優位性を確立した。従来「4つのエージェントを同時実行」する開発手法が普及し、より高度な抽象化レベルでの開発作業が可能になっている。 OpenAI価格面でも、50-80%少ないトークンで同等以上の性能を実現するため、実質的な大幅コスト削減となっている。 OpenAI 9to5Mac

企業業務自動化では、信頼性の大幅向上（幻覚率80%削減）により、より安全で確実な自動化システムの構築が可能になった。「 OpenAI TechCrunch安全完了」機能による適切なリスク管理と透明性の高い回答により、企業レベルでの導入に適した品質を達成している。 OpenAI +3

一般ユーザー向け対話では、無料プランでも推論機能を提供する画期的な展開 TechCrunchにより、幅広いユーザー層にとって実用的な価値を提供している。 Il Sole 24 Ore +3自動的な推論レベル調整により、ユーザーは複雑な設定なしに最適な体験を得られる。 OpenAI

研究・学術利用では、大規模文書処理能力と並列分析機能により、従来のワークフローを大幅に効率化できる。 Botpress +2特に文献調査や大量データ分析において、人間研究者の生産性を大幅に向上させる可能性がある。 OpenAI

コスト効率性では、各モデルが明確に差別化されている。軽量なタスクにはgpt-5-nano（$0.05/$0.40）、バランス重視にはgpt-5-mini（$0.25/$2.00）、最高性能にはgpt-5

（$1.25/$10.00） OpenAIまたはGPT-5 Pro（$200/月定額） OpenAIという階層的選択が可能である。 OpenAI Botpress

結論と今後の展望

GPT-5シリーズは、単なる性能向上を超えて**AI活用パラダイムの根本的変化**を促進している。 OpenAI統合アーキテクチャによるリアルタイム最適化、信頼性の劇的向上、攻撃的な価格設定により、AI技術の実用化における重要な転換点を形成している。 OpenAI OpenAI

最も重要な成果は、**開発者コミュニティでの「不可能」の壁の突破**である。複雑なWebアプリケーションの一発生成、大規模システムのエンドツーエンドデバッグなど、従来は人間の専門知識を必要としていた高度なタスクが自動化可能になった。 Substack +3これにより、ソフトウェア開発における生産性革命が現実のものとなっている。 OpenAI OpenAI

企業導入における実用性では、幻覚率の80%削減という信頼性向上が決定的要因となっている。 OpenAI OpenAIこれまで「実験的」とされていたAI活用が、「本格運用」レベルの信頼性を獲得し、幅広い業務プロセスでの実装が加速すると予想される。 9to5Mac +3

価格競争力の確立により、AI技術の民主化が大幅に進展する。無料ユーザーへの推論機能提供、 TechCrunch TechRadar従来モデル比50%のコスト削減、階層的な料金設定により、個人から大企業まで、それぞれの需要に適したソリューションが提供されている。 Il Sole 24 Ore +5

今後の展望として、GPT-5が確立した「統合システム」アプローチは、AI業界全体の新しい標準となる可能性が高い。 OpenAI単一モデルの性能競争から、複数モデルの知的協調による最適化競争への移行が予想され、2025年はAI活用における新時代の始まりとして記録されるであろう。