



GPT-5、GPT-5 Thinking、GPT-5 Pro の性能比較レポート

はじめに

2025年8月7日、OpenAIは次世代AI言語モデル「GPT-5」を正式発表しました。GPT-5は「史上最も賢く、最速で、最も役立つモデル」として位置づけられ、従来のモデルを統合する革新的なシステムとして開発されています。本レポートでは、GPT-5、GPT-5 Thinking、GPT-5 Proの3つのモデルの技術仕様、性能、機能、コスト、用途の違いを詳細に分析します。^{[1][2]}

1. 基本概要と開発背景

GPT-5の基本構成

GPT-5は統一システムとして設計されており、**高速で効率的なモデル**、**より複雑な問題に対応する深い推論モデル (GPT-5 Thinking)**、そしてリアルタイムルーターの3つの主要コンポーネントから構成されています。^[2]

リアルタイムルーターは会話の種類、複雑さ、必要なツール、ユーザーの明示的な意図（例：「これについて深く考えて」）に基づいて、どちらのモデルを使用するかを瞬時に決定します。このルーターは実際の利用データ（ユーザーがモデルを切り替えるタイミング、回答の満足度、正確性など）を学習して精度を継続的に向上させます。^[2]

開発の背景

Sam Altman CEOは「GPT-3は高校生、GPT-4は大学生との対話のようだった。GPT-5は初めて、あらゆる分野の博士号レベルの専門家と話しているように感じられる」と表現しています。これは単なるパフォーマンス向上ではなく、AIとの対話における質的な変化を意味しています。^[3]

2. 技術仕様とアーキテクチャ

パラメータ数と技術的特徴

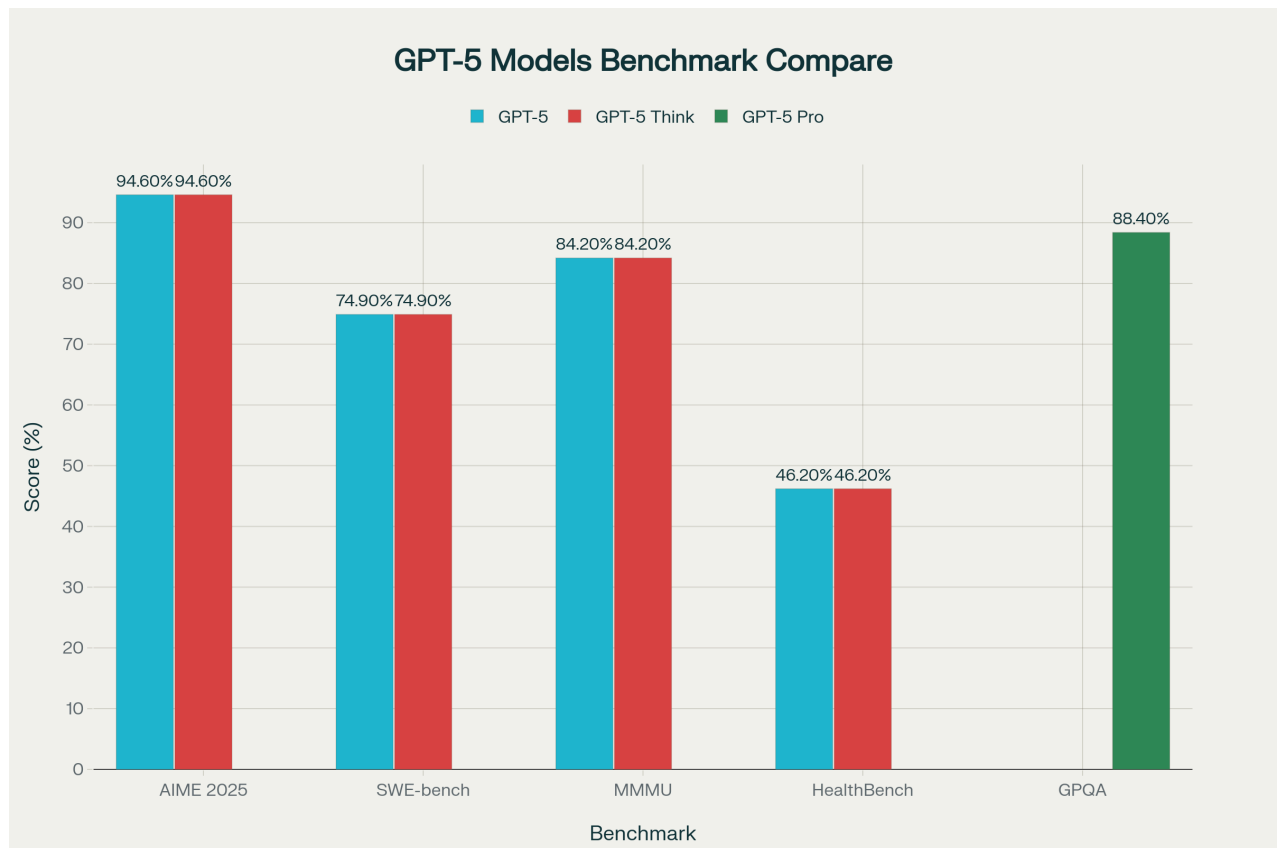
現在公開されている情報によると、GPT-5の詳細なパラメータ数は公表されていませんが、統一アーキテクチャにより以下の技術的改善が実現されています：^[2]

- **統一マルチモーダルアーキテクチャ:** テキスト、画像、音声を単一のシステムで処理
- **動的リソース配分:** クエリの複雑さに応じて計算リソースを自動調整
- **階層的注意メカニズム:** 長いコンテキストでも高精度を維持

コンテキストウィンドウ

GPT-5は標準で**256,000トークン**、拡張版では**1,000,000トークン**のコンテキストウィンドウをサポートしています。これにより、長大な文書や複数時間に及ぶ会話でも一貫性を保持できます。 ^[4]

3. ベンチマーク性能比較



Comprehensive comparison of GPT-5 model variants showing capabilities, performance, and use cases

数学的推論能力

AIME 2025ベンチマークにおいて、GPT-5は****94.6%****という驚異的なスコアを達成しました。これは人間の専門家レベルを上回る性能です。 ^[2]

プログラミング能力

- **SWE-bench Verified:** 74.9%（実世界のコーディング問題） ^[2]
- **Aider Polyglot:** 88%（多言語コーディング） ^[2]

これらのスコアは、GPT-5が複雑なソフトウェア開発タスクを高精度で実行できることを示しています。

マルチモーダル理解

MMMU (Massive Multimodal Multidiscipline Understanding) で**84.2%**を記録し、画像、図表、科学的データの理解において大幅な性能向上を実現しています。^[2]

医療分野

HealthBench Hardで**46.2%**のスコアを獲得し、医療関連の質問に対してより正確で信頼性の高い回答を提供できるようになりました。^[2]

4. 3つのモデルの機能的差異

GPT-5 (統一システム)

- **応答速度:** 数秒以内の高速応答
- **特徴:** スマートルーターによる自動最適化
- **対象ユーザー:** 一般的な質問から中程度の複雑さのタスクまで対応
- **使用制限:** 無料版は5時間に10メッセージ、Plus版は3時間に80メッセージ^[5]

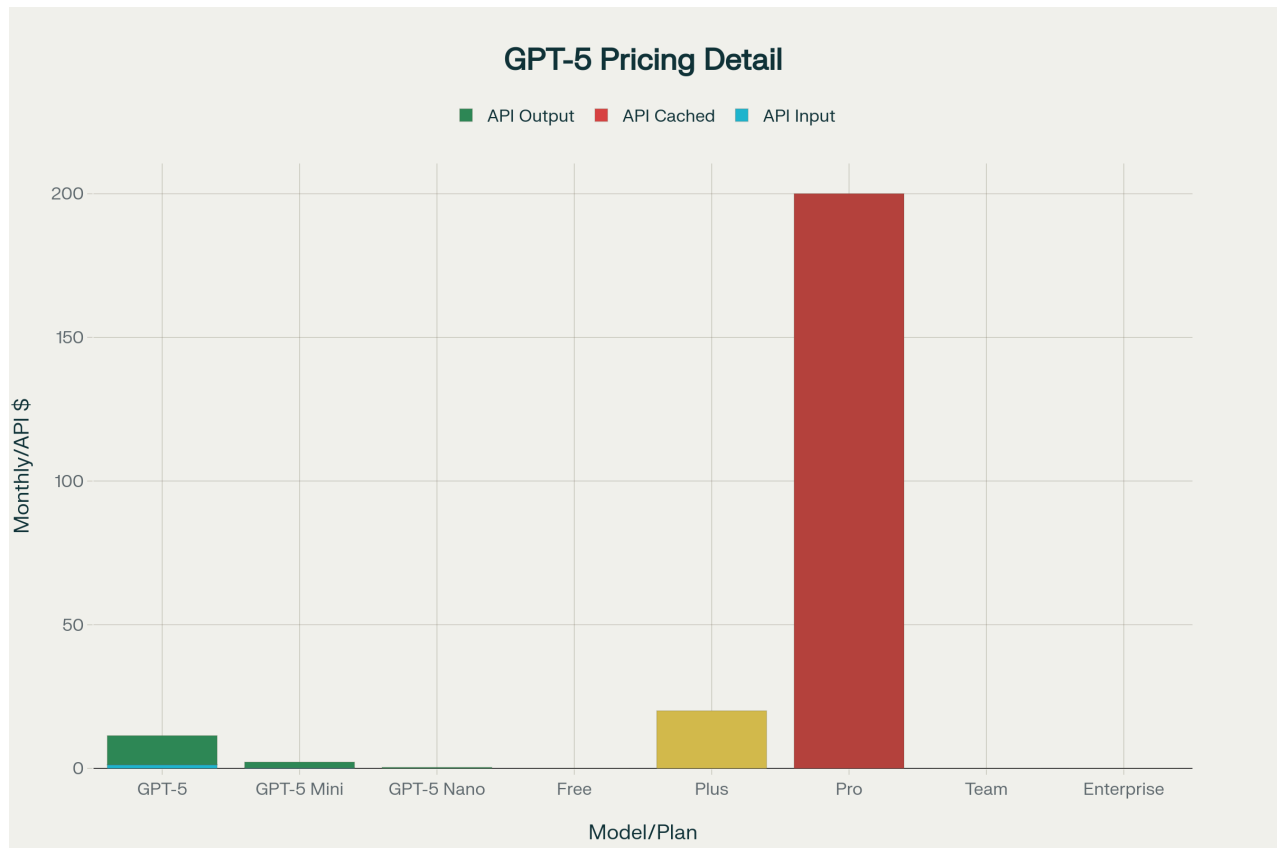
GPT-5 Thinking (深い推論モデル)

- **応答速度:** 1-2分程度
- **特徴:** 明示的な思考過程を経て回答を生成^[6]
- **効率性:** OpenAI o3比で50-80%少ないトークン消費^[2]
- **対象ユーザー:** 複雑な問題解決、詳細な分析を必要とするタスク
- **使用制限:** Plus版は週200メッセージ、Pro版は無制限^[5]

GPT-5 Pro (拡張推論)

- **応答速度:** 数分から10分程度
- **特徴:** 並列テストタイム計算による最高品質の回答^[2]
- **精度向上:** GPT-5 Thinkingより22%少ない重大エラー^[2]
- **専門性:** GPQA (大学院レベル推論) で88.4%のスコア^[2]
- **利用制限:** Pro契約者のみアクセス可能

5. 価格体系とアクセス方法



GPT-5 pricing structure for API usage and ChatGPT subscription plans

API料金体系

- **GPT-5:** 入力\$1.25/100万トークン、出力\$10.00/100万トークン^[7]
- **GPT-5 Mini:** 入力\$0.25/100万トークン、出力\$2.00/100万トークン^[7]
- **GPT-5 Nano:** 入力\$0.05/100万トークン、出力\$0.40/100万トークン^[7]

キャッシュされた入力トークンには90%の割引が適用され、チャットUIの実装において大幅なコスト削減が可能です。^[8]

ChatGPT契約プラン

- **無料版:** GPT-5は5時間に10メッセージ、GPT-5 Thinkingは1日1メッセージ^[5]
- **Plus (\$20/月) :** GPT-5は3時間に80メッセージ、GPT-5 Thinkingは週200メッセージ^[5]
- **Pro (\$200/月) :** GPT-5とGPT-5 Thinking無制限、GPT-5 Pro利用可能^[5]

6. ターゲットユーザーとユースケース

一般ユーザー向け

GPT-5は日常的なタスクに最適化されており、メール作成、レポート執筆、基本的なコーディング支援などに適しています。新しい4つの事前設定人格 (Cynic、Robot、Listener、Nerd) により、ユーザーの好みに応じた対話スタイルを選択できます。^{[2] [9]}

開発者・エンジニア向け

- **フロントエンド開発:** 美しくレスポンス性の高いWebサイトやアプリケーションを単一プロンプトで作成^[2]
- **デバッグ:** 大規模リポジトリの複雑なバグ修正
- **アーキテクチャレビュー:** マルチリポジトリのアーキテクチャ分析と改善提案^[10]

ビジネス・専門職向け

GPT-5 Proは法律、医療、金融、科学研究などの専門分野で専門家レベルの分析を提供します。1000以上の経済的価値の高い実世界の推論タスクで、専門家がGPT-5 Proを67.8%の確率でGPT-5 Thinkingより高く評価しています。^[2]

7. 競合他社との比較

Claude 4との比較

- **コーディング能力:** Claudeは構造化データ処理で依然として優位性を保持^[11]
- **コンテキスト処理:** Claude 4は200Kトークンの安定した処理が可能^[11]
- **価格:** Claude 4 Opusはより高価格（入力\$15、出力\$75/100万トークン）^[8]

Gemini 2.5との比較

- **マルチモーダル:** Geminiは完全なマルチモーダル入力（画像、音声、動画）をサポート^[11]
- **コンテキスト:** Gemini 2.5 Proは2M超大容量コンテキストウィンドウを提供^[11]
- **コスト:** Gemini Flash は\$0.30/100万トークンでより経済的^[8]

8. 安全性と信頼性の向上

ハルシネーション（虚偽情報生成）の大幅削減

GPT-5は従来モデルと比較して大幅な改善を達成しています：^[2]

- GPT-4oより**45%少ない**事実エラー
- 推論モード使用時はOpenAI o3より**80%少ない**事実エラー
- オープンエンド型事実質問でo3の約**6分の1**のハルシネーション率

Safe Completions技術

従来の拒否ベース安全訓練に代わり、GPT-5では「Safe Completions」という新しい安全訓練手法を導入しました。この技術により：^[2]

- デュアルユース（良性・悪性両方の用途がある）質問により適切に対応
- 部分的な回答や高レベルでの回答による有用性と安全性の両立
- 拒否する場合の理由説明と安全な代替案の提示

9. マルチモーダル能力と将来の拡張

現在のマルチモーダル機能

GPT-5は統一アーキテクチャでテキスト、画像、音声进行处理できますが、動画処理はまだ完全には実装されていません。しかし、将来的にはSORA（OpenAIのテキスト-動画生成モデル）との統合が予定されています。^[12]

Canvas統合

GPT-5はOpenAIのCanvas（構造化ワークスペース）と深く統合されており、数学、データ、複数ステップ指示に対するビジュアル問題解決環境での推論能力が向上しています。^[13]

10. 将来の統合計画

OpenAIは将来的に、現在の3つのコンポーネント（高速モデル、深い推論モデル、ルーター）を**単一の統合モデル**に集約することを計画しています。これにより、ユーザーはより seamlessな体験を得られるようになります。^[2]

結論と推奨事項

各モデルの最適な用途

GPT-5（統一システム）：

- 日常的な質問応答、メール作成、基本的なコーディング
- コスト効率と応答速度を重視するユーザー
- 無料版または Plus 版ユーザーに最適

GPT-5 Thinking:

- 複雑な問題解決、詳細な分析、高度なコーディングタスク
- 品質と精度を重視し、多少の待機時間を許容できるユーザー
- Plus版以上の契約者に推奨

GPT-5 Pro:

- 最高レベルの専門的分析、研究、医療・法律・科学分野のタスク
- 予算に余裕があり、最高品質の結果を必要とする専門家
- Pro契約（\$200/月）必須

導入時の考慮事項

1. **予算とROI:** Pro版は10倍の料金だが、専門的タスクでの価値は大幅に向上
2. **使用パターン:** 日常的な使用にはGPT-5、特別なプロジェクトにはGPT-5 Thinking/Pro
3. **競合他社との比較:** 用途によってはClaude 4やGemini 2.5が適している場合もある

GPT-5シリーズは AI技術の重要な転換点を示しており、単なるチャットボットから真の知的パートナーへの進化を表現しています。ユーザーのニーズと予算に応じた適切なモデル選択により、AI活用の効果を最大化できるでしょう。

✱

1. <https://www.cnn.com/2025/08/07/openai-launches-gpt-5-model-for-all-chatgpt-users.html>
2. <https://openai.com/index/introducing-gpt-5/>
3. <https://www.bbc.com/news/articles/cy5prvgw0r1o>
4. <https://www.cursor-ide.com/blog/gpt-5-api-guide>
5. <https://help.openai.com/en/articles/11909943-gpt-5-in-chatgpt>
6. https://note.com/yusuke_motoyama/n/nb2b8ac4d22b2
7. <https://platform.openai.com/docs/pricing>
8. <https://simonwillison.net/2025/Aug/7/gpt-5/>
9. <https://xpert.digital/en/gpt5-is-here/>
10. <https://academy.openai.com/public/clubs/work-users-ynjqu/resources/intro-gpt-5>
11. <https://news.aibase.com/news/20340>
12. <https://daily.dev/blog/gpt-5-by-openai-everything-you-should-and-shouldnt-expect>
13. <https://botpress.com/blog/everything-you-should-know-about-gpt-5>