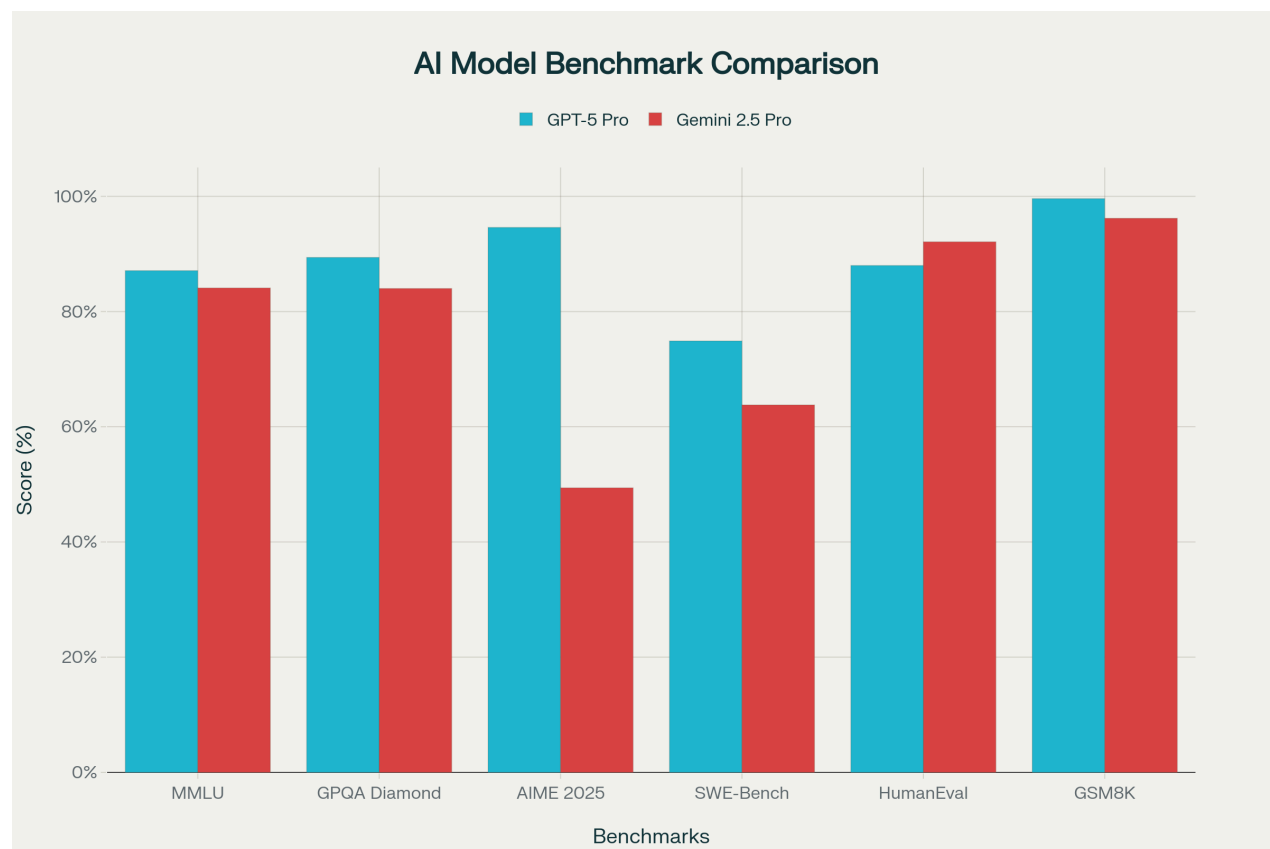




GPT-5 ProとGemini 2.5 Pro Deep Thinkの徹底比較：2025年最新AIモデルの実力を検証

2025年は次世代AI技術の飛躍的進歩を象徴する年となった。8月にリリースされたOpenAIのGPT-5 Proと、5月に発表され8月に本格提供が開始されたGoogleのGemini 2.5 Pro Deep Thinkは、それぞれ推論能力と統合システムという異なるアプローチで革新をもたらしている。本レポートでは、包括的な調査に基づいて両モデルの技術仕様、性能、実用性を徹底比較し、企業および個人ユーザーにとっての最適な選択指針を提示する。



GPT-5 ProとGemini 2.5 Pro Deep Thinkの主要仕様・性能比較表

技術仕様とアーキテクチャの比較

GPT-5 Proの技術的特徴

GPT-5 Proは、OpenAIが「統合システム」として設計した革新的なアーキテクチャを採用している。このシステムは、高速・効率的に応答するモデルと、複雑な問題に対して深い推論を行うモデル（GPT-5 Thinking）を組み合わせ、リアルタイムルーターが会話の複雑さや必要なツール、ユーザー

の明示的な意図に基づいて最適なモデルを自動選択する。コンテキストウィンドウは入力272,000トークン、出力128,000トークンで、合計約40万トークンの処理が可能である。 [1] [2] [3] [4] [5] [6]

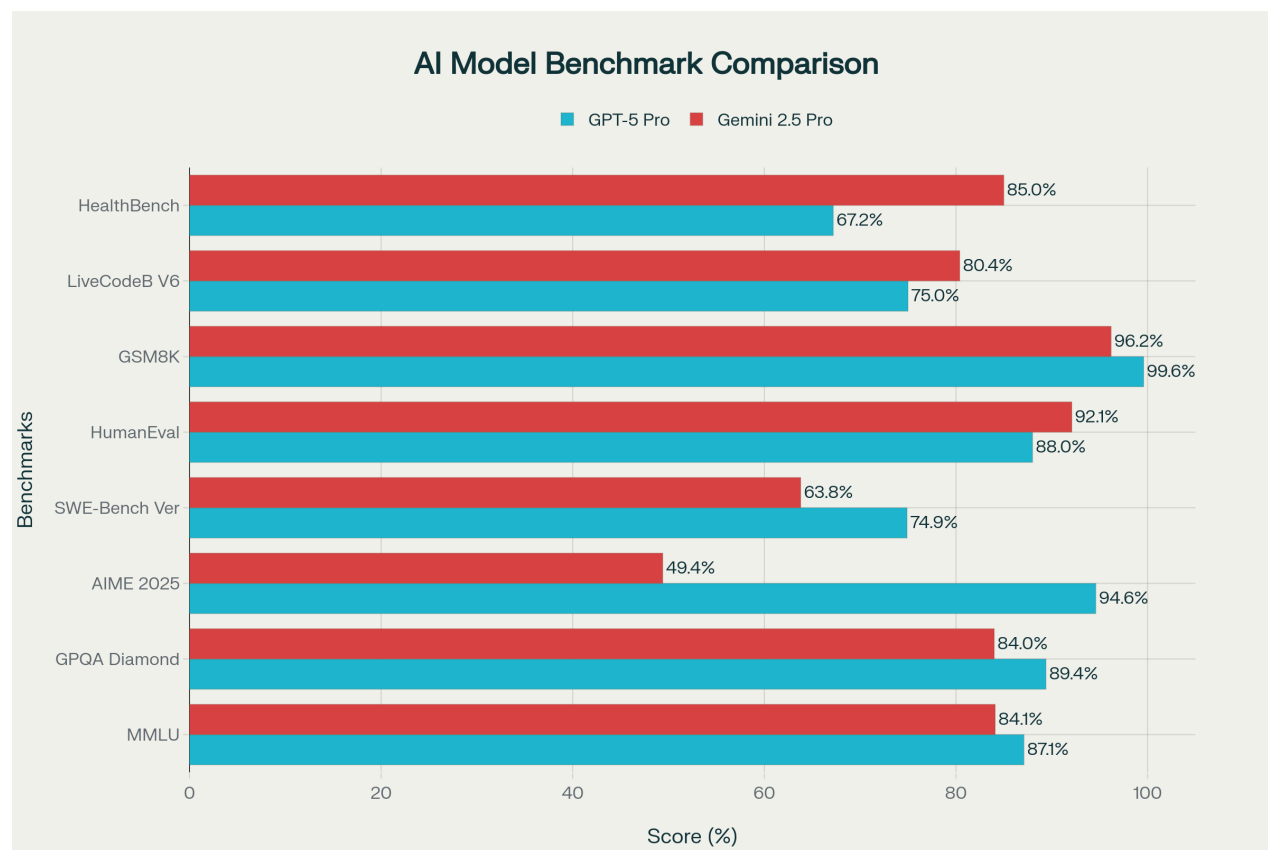
GPT-5 Proの特筆すべき特徴は「Safe Completions」機能である。従来のAIが危険な質問を単純に拒否していたのに対し、GPT-5は安全な範囲内で高レベルな回答を提供し、制限理由を説明した上で代替解決策を提案する。これにより、ハルシネーション（誤情報生成）率はGPT-4oと比較して約45%、推論モードではo3と比較して約80%削減されている。 [7] [8] [9]

Gemini 2.5 Pro Deep Thinkのアーキテクチャ

Gemini 2.5 Pro Deep Thinkは、Sparse Mixture-of-Experts（MoE）トランスフォーマーアーキテクチャを基盤としている。このアーキテクチャでは、入力トークンごとに最適な「エキスパート」パラメータ群に動的にルーティングすることで、総モデル容量を維持しながらトークンあたりの計算コストを削減している。 [10] [11]

Deep Thinkモードの最大の革新は「並列思考技術」の採用である。従来のAIモデルが単一の推論パスを辿るのに対し、Deep Thinkは複数の仮説を同時に生成し、並行して検討した後、最適解を組み合わせることで導出する。コンテキストウィンドウは1,000,000トークン（将来的に2,000,000トークンに拡張予定）で、出力は192,000トークンまで対応している。 [12] [13] [14] [15] [16] [10]

主要ベンチマークにおける性能評価



GPT-5 Pro vs Gemini 2.5 Pro Deep Think ベンチマーク性能比較

数学・科学分野での性能差

数学分野において、両モデルは顕著な性能差を示している。GPT-5 ProはAIME 2025ベンチマークで94.6%という驚異的なスコアを記録し、Python ツール使用時には100%の正解率を達成している。一方、Gemini 2.5 Pro Deep ThinkはUSAMO 2025で49.4%のスコアを記録しており、これはOpenAI o3の19.1%を大幅に上回るものの、GPT-5には及ばない。^{[3] [8] [15]}

科学分野では、GPQA Diamondベンチマークにおいて、GPT-5 Pro (Pythonツール付) が89.4%を記録し、Gemini 2.5 Pro Deep Thinkの84.0%を上回っている。しかし、HealthBenchにおいては、Gemini 2.5 Pro Deep Thinkが85%と、GPT-5 Proの67.2%を大幅に超える性能を示している。^{[8] [17]}

プログラミング・コーディング能力

コーディング分野では、SWE-Bench Verifiedベンチマークにおいて、GPT-5 Proが74.9%のスコアを記録し、Gemini 2.5 Pro Deep Thinkの63.8%を上回っている。特に、GPT-5は複数のツール呼び出しを順次・並行して実行する能力に長けており、複雑な実世界タスクのエンドツーエンド実行において優位性を示している。^{[3] [8] [17]}

一方、HumanEvalベンチマークでは、Gemini 2.5 Pro Deep Thinkが92.1%を記録し、GPT-5 Proの88%を上回る結果を示している。LiveCodeBench V6では、Gemini 2.5 Pro Deep Thinkが80.4%と、GPT-5 Proの75%を超える性能を発揮している。^{[8] [18] [19] [20]}

言語理解と推論能力

MMLU (Massive Multitask Language Understanding) ベンチマークにおいて、GPT-5 Proは87.1%のスコアを記録し、Gemini 2.5 Pro Deep Thinkの84.1%を上回っている。これは両モデルが高い言語理解能力を有していることを示している。^{[6] [17] [21]}

マルチモーダル能力の比較

GPT-5 Proのマルチモーダル機能

GPT-5 Proは現時点でテキストと画像入力に対応しているが、出力はテキストのみとなっている。しかし、音声機能の改善は顕著で、カスタムGPTsとの連携や、トーンや話し方の調整が可能になっている。また、「Visual Workspace」機能により、チャートや図表との相互作用が向上している。^{[2] [3] [22]}

Gemini 2.5 Pro Deep Thinkのマルチモーダル優位性

Gemini 2.5 Pro Deep Thinkは、ネイティブマルチモーダル対応において明確な優位性を示している。テキスト、画像、音声、動画ファイルをすべて同一のコンテキストウィンドウ内で処理可能で、最大1,000,000トークンの長文脈理解を実現している。^{[10] [14] [23]}

特に注目すべきは、ネイティブ音声出力機能の実装である。複数話者対応のテキスト読み上げ、感情認識対話 (Affective Dialogue)、24言語以上でのシームレスな言語切り替え機能を提供している。^{[23] [24]}

Deep Think機能と推論能力の詳細分析

Gemini 2.5 Pro Deep Thinkの並列思考メカニズム

Deep Think機能の核心は、人間の思考プロセスを模倣した並列思考技術にある。システムは単一の問題に対して複数の解決アプローチを同時に検討し、100以上の論理ステップを一貫性を保ちながら実行する。さらに、自己修正機能により推論過程の誤りを発見・修正し、思考プロセスの透明性向上を実現している。^{[13] [15] [25] [26]}

この技術により、反復的な開発や設計、数学的仮説の形成、複雑な科学文献の解釈において優れた能力を発揮している。Google公式の評価では、「state-of-the-art performance（最先端の性能）」を達成したと報告されている。^{[25] [26] [27]}

GPT-5 Proの統合推論システム

GPT-5 Proの推論能力は、「GPT-5 Thinking」モードを通じて発揮される。ユーザーが「深く考えてください」と指示するか、システムが自動的に複雑さを判断して推論モードに切り替わる。この機能により、ハルシネーション率の大幅な削減と、博士レベルの専門知識を持つ回答の生成を実現している。^{[1] [4] [5] [28] [29] [30]}

API価格体系と経済性の比較

基本価格体系

両モデルの基本的な価格体系は類似している。GPT-5 Proは入力トークンが\$1.25/100万トークン、出力トークンが\$10.00/100万トークンとなっている。Gemini 2.5 Pro Deep Thinkは、200,000トークン以下の場合と同様の価格設定だが、それを超えると入力\$2.50、出力\$15.00/100万トークンとなる段階的価格設定を採用している。^{[2] [6] [31] [32]}

実コスト分析とROI

実際の使用コストでは、GPT-5 Proの統合アーキテクチャが経済的優位性を示している。多モデル構成のGPT-4セットアップからの移行により、総API費用を30-40%削減できるとの分析結果が報告されている。月間1,000万トークン処理する企業アプリケーションの場合、GPT-5 Proは約\$2,400、従来のGPT-4マルチモデル構成は\$3,200の費用となる。^[33]

専門家レビューとユーザー評価

技術メディアによる評価

技術メディアからの評価は概ね好意的である。Artificial Analysisの独立ベンチマークでは、GPT-5が新たな知能の標準を確立し、Intelligence Indexで68点を記録したと評価している。一方で、推論努力レベル間でのトークン使用量に23倍の差があることも指摘されている。^[34]

Vellum AIのレビューでは、GPT-5が「あらゆるベンチマークでトップを占めている」と評価し、特に信頼性と安全性の面での進歩を強調している。^[8]

開発者コミュニティの反応

開発者コミュニティでは、GPT-5に対する評価が二分されている。肯定的な評価では、「GPT-5は真の飛躍です。業界の他の企業は今、全力疾走する必要がある」（Matt Shumer）といった声が上がっている。^{[28] [35] [36]}

しかし、一方でユーザーからは厳しい批判も寄せられている。特にPlusプラン加入者から、「モデルセクターが廃止された」「創造的タスクで性能が劣る」「制限が厳しくなった」といった不満が表明されている。^[28]

Gemini 2.5 Pro Deep Thinkへの評価

Gemini 2.5 Pro Deep Thinkに対する専門家の評価は総じて高い。特に数学オリンピック問題での圧倒的な性能向上が評価されており、「AIが人間の知的活動を補完から代替へと進化する転換点」と位置づけられている。^{[15] [37]}

開発者からは、「ツール呼び出しの失敗率が大幅に減少」「以前問題となっていたエラー発生が劇的に改善」といった実用面での改善が高く評価されている。^[38]

ソフトウェア開発支援能力の詳細比較

GPT-5 Proの開発支援機能

GPT-5 Proは、 τ 2-bench telecomベンチマークで96.7%を達成し、長期間にわたるエージェントタスクにおいて最先端の結果を示している。複数のツール呼び出しを順次・並行して実行する能力により、複雑な実世界タスクのエンドツーエンド実行において優位性を発揮している。^[3]

フロントエンド開発においては、o3を70%上回る性能を示し、最小限のプロンプトからの高品質なコード生成と、パーソナリティや制御性の向上を実現している。^{[39] [3]}

Gemini 2.5 Pro Deep Thinkの開発能力

Gemini 2.5 Pro Deep Thinkは、WebDev Arenaで1,499.95ポイントという過去最高スコアを獲得し、Anthropic Claude 3.7 Sonnetを大幅に上回っている。特に、反復的な開発・設計において優れた能力を発揮し、美的感覚と機能性の両面での改善提案が可能である。^{[27] [38]}

1,000,000トークンの大容量コンテキストウィンドウにより、大規模コードベース全体を単一プロンプトで処理・編集することが可能で、シニアエンジニアレベルの複雑なリファクタリング作業も実行できる。^[38]

実用性とユースケース分析

企業向けアプリケーション

大規模開発プロジェクトにおいては、Gemini 2.5 Pro Deep Thinkの大容量コンテキストウィンドウと並列思考能力が優位性を示している。特に、長文書分析、動画コンテンツ理解、大規模コードベースの処理において、従来では不可能だったタスクを実現している。^{[40] [41] [42]}

一方、GPT-5 Proは、統合アーキテクチャによる運用効率性とコスト削減効果により、エンタープライズ用途での採用が進んでいる。API統合の簡便さと、複数モデル管理の必要性を排除する点で、企業IT部門からの評価が高い。^[33]

研究・学術分野での活用

研究分野では、Gemini 2.5 Pro Deep Thinkの数学的・科学的推論能力が特に注目されている。国際数学オリンピックで金メダル級の性能を示したモデルの派生版として、数学的仮説の形成・探索、複雑な科学文献の解釈において研究者の強力なツールとなっている。^{[15] [26] [43]}

GPT-5 Proは、HealthBench等の医療分野ベンチマークで改善を示しており、専門的な意思決定支援ツールとしての可能性を有している。^[8]

総合評価と推奨事項

性能面での総合評価

数値的なベンチマークでは、GPT-5 Proが数学・科学・コーディングの多くの分野でGemini 2.5 Pro Deep Thinkを上回る性能を示している。特に、AIME 2025での94.6%、SWE-Bench Verifiedでの74.9%は顕著な優位性を示している。^[8]

しかし、Gemini 2.5 Pro Deep Thinkは、HumanEvalでの92.1%、LiveCodeBench V6での80.4%、そしてHealthBenchでの85%など、特定分野での優位性を示している。また、マルチモーダル機能の充実度では明確にGemini 2.5 Pro Deep Thinkが優位である。^{[10] [23]}

コスト効率性の比較

基本価格は類似しているものの、GPT-5 Proの統合アーキテクチャによるトータルコスト削減効果は無視できない。多モデル構成からの移行により30-40%のコスト削減が期待でき、ROIの観点では3-4ヶ月での回収が見込まれている。^[33]

用途別推奨事項

数学・科学研究重視の場合： 高難度数学問題や科学的推論が主要用途であれば、GPT-5 Proの圧倒的な数学性能が適している。AIME 2025での94.6%という成績は、現時点で他の追従を許さない水準である。

大規模文書・マルチメディア処理の場合： 1,000,000トークンのコンテキストウィンドウと完全なマルチモーダル対応を有するGemini 2.5 Pro Deep Thinkが適している。書籍数冊分のテキストや数時間の動画を同時処理可能な能力は独特の価値を提供する。

企業統合・運用効率重視の場合： GPT-5 Proの統合アーキテクチャとコスト効率性が企業採用において優位である。単一APIでの複数機能提供と、運用管理の簡素化は企業IT部門にとって重要な価値となる。

創造的・デザイン作業の場合： WebDev Arenaでの圧倒的な成績を示すGemini 2.5 Pro Deep Thinkが、美的感覚と機能性を両立したアウトプット生成において優位性を示している。

両モデルは、それぞれ異なる強みを持つ革新的なAIシステムであり、用途に応じた適切な選択が重要である。GPT-5 Proは統合性と効率性、Gemini 2.5 Pro Deep Thinkは専門性と多様性をそれぞれの核心价值として提供している。

✽

1. <https://openai.com/ja-JP/index/introducing-gpt-5/>
2. <https://shift-ai.co.jp/blog/33397/>
3. <https://simonwillison.net/2025/Aug/7/gpt-5/>
4. <https://ensou.app/blog/introducing-gpt-5/>
5. <https://zenn.dev/schroneko/articles/introducing-gpt-5>
6. <https://openai.com/index/introducing-gpt-5-for-developers/>
7. <https://www.cnn.com/2025/08/07/openai-launches-gpt-5-model-for-all-chatgpt-users.html>
8. <https://note.com/yusukexz777/n/n3a4ddac87694>
9. <https://openai.com/index/gpt-5-system-card/>
10. <https://zenn.dev/galirage/articles/gpt-5-release>
11. <https://momo-gpt.com/column/chatgpt-5-pro/>
12. <https://openai.com/gpt-5/>
13. <https://openai.com/index/introducing-gpt-5/>
14. <https://news.yahoo.co.jp/articles/dc4c0dae6092ceba8ff3339e38b33f51d0479f95>
15. <https://news.microsoft.com/source/features/ai/openai-gpt-5/>
16. https://note.com/masa_wunder/n/na26d568e5588
17. <https://news.yahoo.co.jp/articles/57d318f18f44087558794d26b393f91ce60ea90a>
18. https://note.com/ai_worker/n/n776d8715cdc4
19. <https://japan.zdnet.com/article/35236491/>
20. <https://yorozuipsc.com/blog/gpt-5gpt-5-thinkinggpt-5-pro>
21. <https://techcrunch.com/2025/08/01/google-rolls-out-gemini-deep-think-ai-a-reasoning-model-that-tests-multiple-ideas-in-parallel/>
22. <https://japan.zdnet.com/article/35233231/>
23. <https://jobirun.com/gemini-2-5-deep-think-enhanced-reasoning-model-explained/>
24. https://gigazine.net/gsc_news/en/20250804-google-gemini-2-5-deep-think/
25. <https://blog.google/intl/ja-jp/company-news/technology/google-gemini-updates-io-2025/>
26. <https://ai.google.dev/gemini-api/docs/models>
27. <https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/>
28. <https://digirise.ai/chaen-ai-lab/gemini-2-5-pro-deepthink/>
29. <https://deepmind.google/models/gemini/>
30. <https://cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-pro>
31. <https://www.itmedia.co.jp/aiplus/articles/2508/02/news026.html>
32. <https://gemini.google/subscriptions/>
33. <https://gemini.google/release-notes/>

34. <https://note.com/aidxlab/n/nc9ad550b5f92>
35. <https://one.google.com/about/google-ai-plans/>
36. <https://www.watch.impress.co.jp/docs/news/2036289.html>
37. <https://pc.watch.impress.co.jp/docs/news/2036701.html>
38. <https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-2-5-Deep-Think-Model-Card.pdf>
39. <https://shift-ai.co.jp/blog/33115/>
40. <https://blog.google/products/gemini/gemini-2-5-deep-think/>
41. <https://artificialanalysis.ai/articles/gpt-5-benchmarks-and-analysis>
42. <https://bito.ai/blog/gemini-1-5-pro-vs-gpt-4-turbo-benchmarks/>
43. <https://www.cursor-ide.com/blog/gpt-5-api-guide>