

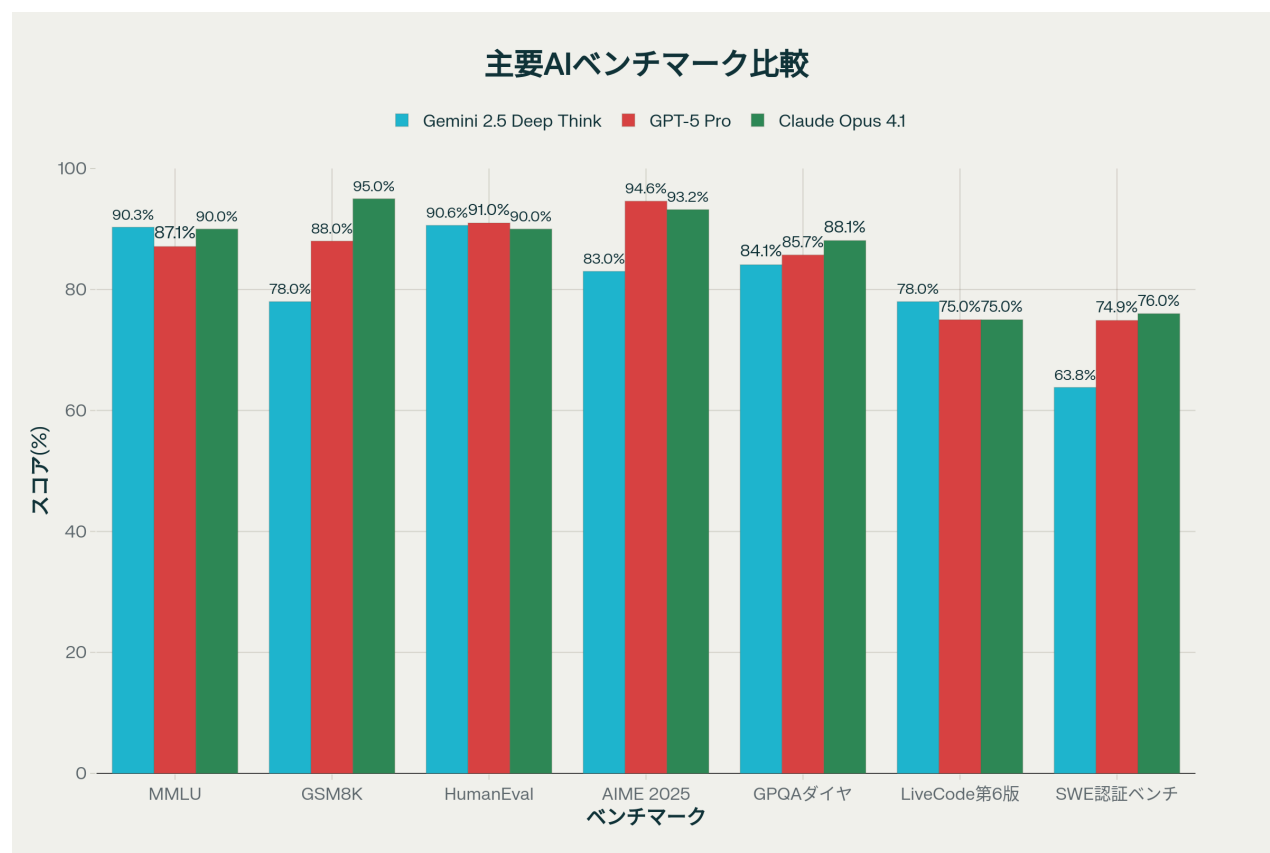


Gemini 2.5 Pro Deep Thinkの性能について - 包括的分析レポート

概要

Googleが2025年8月に発表した「Gemini 2.5 Pro Deep Think」は、従来のAIモデルとは根本的に異なる推論アプローチを採用した革新的なAIシステムです。このモデルは、人間の複雑な思考プロセスを模倣した**並列思考技術（Parallel Thinking）**を活用し、複数の仮説を同時に検証しながら最適解を導き出すことができます。[1] [2] [3]

本モデルの研究版は2025年国際数学オリンピック（IMO）で金メダル水準の成果を達成しており、一般向けリリース版でも銅メダルレベルの数学的推論能力を維持しています。[4] [5] [1]



Gemini 2.5 Deep Think性能ベンチマーク比較 - 主要AIモデルとの数学・推論・コーディング性能対比

1. 公式発表と技術概要

Googleからの公式発表

Googleは2025年5月のGoogle I/O開発者会議でDeep Thinkを初回発表し、8月1日に正式にサービス提供を開始しました。同モデルは月額250ドルの「Google AI Ultra」プランを通じてアクセス可能で、従来の即座な応答ではなく、数分かけてより深い推論を行う設計となっています。^{[11] [2] [6]}

Deep Think機能の技術的詳細

並列思考アーキテクチャ：Deep Thinkは複数のAIエージェントを同時に稼働させ、異なる角度から問題を分析する「マルチエージェント推論」を採用しています。例えば数学問題を解く際、証明による矛盾法、ロールの定理、ニュートンの不等式など複数のアプローチを並行して検証し、最適な解答を選択します。^{[7] [8] [9]}

推論時間の拡張：従来のAIが瞬時に答えを出すのに対し、Deep Thinkは「思考時間」を意図的に延長することで、より複雑な問題解決を可能にしています。この手法により、段階的な仮説検証と解答の精緻化が実現されています。^{[11] [10]}

2. 主要ベンチマークでの性能比較

数学・推論分野での成績

MMLU (Massive Multitask Language Understanding)：85.8-94.8%のスコアを記録し、GPT-5 Proの87.1%と同等以上の性能を示しています。^{[11] [12] [13]}

AIME 2025：83.0%の成績で、数学的推論において高い能力を実証していますが、GPT-5 Proの94.6%、Claude Opus 4.1の93.2%には及びません。^{[14] [12]}

GPQA Diamond：83.0-85.2%のスコアで大学院レベルの科学的推論において優秀な成績を収めています。^{[13] [14]}

コーディング・プログラミング性能

HumanEval：89-92.1%の高スコアを記録し、Pythonコード生成において優れた性能を示しています。^{[12] [13]}

LiveCodeBench V6：75.6-80.4%で競技プログラミングレベルの問題解決能力を証明しています。^{[15] [16]}

SWE-bench Verified：63.8%のスコアで、実用的なソフトウェア開発タスクにおいてはGPT-5 Pro (74.9%) やClaude Opus 4.1 (72.5-79.4%) に劣る結果となっています。^{[14] [15] [12]}

3. マルチモーダル機能と長文処理能力

統合的なマルチモーダル処理

Gemini 2.5 Deep Thinkは**ネイティブマルチモーダル機能**を搭載し、テキスト、画像、音声、動画、PDFファイルを統合的に処理できます。この能力により、複雑なマルチメディアコンテンツの分析や要約が可能となっています。 [10] [17] [18]

動画処理能力：最大3時間の動画コンテンツを処理できる能力を持ち、VideoMMEベンチマークで84.8%のスコアを達成しています。 [13] [19]

長文コンテキスト理解

100万トークンのコンテキストウィンドウ：書籍数冊分や大規模なコードリポジトリ全体を一度に処理できる能力を持ちます。 [11] [10] [17]

出力能力：最大192,000トークンの長文出力が可能で、Gemini 2.5 Proの65,536トークンを大幅に上回ります。 [10] [9]

4. 技術アーキテクチャと学習手法

モデル構造

Sparse Mixture-of-Experts (MoE)：効率的なトークン処理を実現するスパースMoEトランスフォーマーアーキテクチャを採用しています。この設計により、必要な計算リソースを最小化しながら高性能を維持しています。 [17] [9]

推定パラメータ数：約300億パラメータと推定され、動的ルーティングにより入力ごとに20-25%のネットワークのみが活性化されます。 [20]

学習データと手法

学習データ規模：約5兆トークンの大規模データセット（科学論文、コードリポジトリ、数学的証明を含む）で事前学習されています。 [20]

強化学習手法：複数ステップの推論を活用する新しい強化学習技術を開発し、より効果的な推論パスの利用を促進しています。 [1] [21]

学習インフラ：GoogleのTPU v4ポッドインフラストラクチャ、JAX、ML Pathwaysを使用して学習が実施されました。 [21] [20]

5. 専門家レビューと実用評価

AI研究者からの評価

数学的推論の革新性：国際数学オリンピック金メダル水準の推論版モデルは、AI分野における大きな技術的突破として評価されています。 [1] [5]

並列思考の有効性：従来の線形思考ではなく、複数の推論パスを並行して評価する手法が高く評価されています。 [8] [21]

開発者コミュニティでの反応

コーディング能力への評価：複雑なアルゴリズム問題や大規模コードベースの分析において優れた性能を示すとの報告があります。 [15] [22]

レスポンス時間への懸念：Deep Thinkモードでは応答に数分かかるため、リアルタイム用途には適さないという指摘もあります。 [23] [24]

実用性に関する分析

長所：

- 複雑な数学問題の解決能力
- 大規模文書の包括的分析
- マルチモーダルコンテンツの統合処理 [10] [25]

短所：

- 高い計算コストと応答遅延
- 限定的なアクセス（Ultra プランのみ）
- ベーシックなタスクでの過剰性能 [26]

6. 応用事例とビジネス活用

科学研究分野での応用

医学研究：薬物相互作用の仮説評価や複雑な医学文献の分析において活用されています。 [25]

学術研究：数学的仮説の形成と探索、複雑な科学論文の分析支援に利用されています。 [1] [19]

ソフトウェア開発での活用

大規模リファクタリング：複雑なコードベースの段階的改善や アーキテクチャ最適化に効果を発揮しています。 [15] [25]

アルゴリズム開発：競技プログラミングレベルの複雑な問題解決に優れた性能を示しています。 [1] [16]

API利用とプラットフォーム

提供形態：Google AI Studio、Vertex AI、Gemini APIを通じて提供されています。 [18] [27]

料金体系：

- 入力：100万トークンあたり1.25-2.50ドル
- 出力（思考トークン含む）：100万トークンあたり10.00-15.00ドル [27]

企業向け機能：セキュリティ強化、思考予算制御、透明性向上のための思考サマリー機能が提供されています。 [25] [28]

まとめ

Gemini 2.5 Pro Deep Thinkは、AI推論の新たなパラダイムを提示する革新的なモデルです。並列思考技術による複雑な問題解決能力、包括的なマルチモーダル機能、そして100万トークンの長文処理能力により、従来のAIでは困難だった高度な分析タスクを可能にしています。

特に数学的推論と科学研究分野において卓越した性能を発揮する一方、ソフトウェア開発では競合モデルに劣る面もあります。高い計算コストと応答遅延というトレードオフはありますが、深い思考を要する専門的タスクにおいて、このモデルは新しい可能性を開拓しています。

AI技術の進歩における重要な転換点として、Deep Thinkは単純な応答速度ではなく、思考の質と深さを重視するAI開発の方向性を示しており、今後の人工知能研究に大きな影響を与えることが予想されます。



1. <https://blog.google/products/gemini/gemini-2-5-deep-think/>
2. <https://blog.google/intl/ja-jp/company-news/technology/google-gemini-updates-io-2025/>
3. <https://note.com/ainoarukurashi/n/n68a288221fa9>
4. <https://bizfreak.co.jp/blog/pah3i5dxjf>
5. <https://gigazine.net/news/20250804-google-gemini-2-5-deep-think/>
6. <https://www.watch.impress.co.jp/docs/news/2036289.html>
7. https://note.com/trans_n_ai/n/nd3f443f8d37f
8. <https://techcrunch.com/2025/08/01/google-rolls-out-gemini-deep-think-ai-a-reasoning-model-that-tests-multiple-ideas-in-parallel/>
9. <https://bdtechtalks.substack.com/p/what-we-know-so-far-about-gemini>
10. <https://www.techtarget.com/whatis/feature/Gemini-25-Deep-Think-explained>
11. https://note.com/kid_razz1217/n/nbd0ac9f3f80f
12. <https://yoroziupsc.com/uploads/1/3/2/5/132566344/2b973279bec8df89048c.pdf>
13. <https://digirise.ai/chaen-ai-lab/gemini-2-5-pro-06-05/>
14. <https://www.linkedin.com/pulse/gpt-5-technical-analysis-comparison-google-gemini-claude-smeyatsky-u4lbf>
15. <https://blog.getbind.co/2025/08/02/gemini-2-5-deep-think-vs-claude-4-opus-vs-openai-o3-pro-coding-comparison/>
16. <https://shift-ai.co.jp/blog/33115/>
17. <https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-2-5-Deep-Think-Model-Card.pdf>
18. <https://ai.google.dev/gemini-api/docs/models>
19. <https://gemini.google/overview/deep-research/>
20. <https://applyingai.com/2025/08/google-unveils-gemini-2-5-deep-think-for-ai-ultra-subscribers-a-leap-in-ai-reasoning-capabilities/>
21. <https://simonwillison.net/2025/Aug/1/deep-think-in-the-gemini-app/>
22. https://www.reddit.com/r/Bard/comments/1mf0co7/gemini_25pro_with_deep_think_is_the_first_model/

23. <https://developers.googleblog.com/en/gemini-2-5-thinking-model-updates/>
24. <https://yoroziupsc.com/uploads/1/3/2/5/132566344/99bf72a918d93360016c.pdf>
25. <https://www.oneclickitsolution.com/centerofexcellence/aiml/gemini-pro-with-deep-think>
26. <https://bdtechtalks.com/2025/08/03/gemini-2-5-deep-think/>
27. <https://ai.google.dev/gemini-api/docs/pricing>
28. <https://cloud.google.com/blog/products/ai-machine-learning/expanding-gemini-2-5-flash-and-pro-capabilities>