

Gemini 2.5 Deep Think 徹底評判調査レポート



Genspark

Aug 02, 2025

はじめに

2025 年 8 月 1 日、Google が AI 業界に衝撃を与える新しいモデル「Gemini 2.5 Deep Think」をリリースしました。このモデルは、従来の AI の限界を超える「並列思考」技術を搭載し、複数のベンチマークで業界最高水準の性能を記録しています。本レポートでは、リリース直後から現在までの評判、技術的革新、実際の使用感、そして課題について徹底的に調査・分析いたします。

1. 基本情報とリリース概要

1.1 リリース詳細

Gemini 2.5 Deep Think は 2025 年 8 月 1 日に正式にリリースされ、Google AI Ultra¹ プラン（月額 36,800 円）の加入者向けに提供開始されました。このモデルは、今年の IMO（国際数学オリンピック）で金メダル相当の性能を達成したモデルの改良版であり、Google 公式ブログ²によると「日常使用に最適な速度と使いやすさを実現しながら、2025 年 IMO ベンチマークで Bronze レベルの性能を達成」しています。

1.2 技術的革新点

Deep Think の最大の特徴は、**並列思考 (parallel thinking)** 技術にあります。npaka 氏の Note 記事³によると、この技術は以下の仕組みで動作します：

- 複数アイデアの同時生成・検討：**人間が複雑な問題に取り組む際のように、様々な角度から検討し、潜在的な解決策を比較検討
- 推論時間の拡張：**通常より長い「思考時間」を与え、様々な仮説を探索
- 新しい強化学習技術：**拡張された推論パスを活用するように促す技術により、時間の経過とともにより優れた直感的な問題解決者へと進化

2. ベンチマーク性能の詳細分析

2.1 数学分野での圧倒的優位性

DigiRise AI Lab⁴の詳細分析によると、Gemini 2.5 Deep Think は数学分野で歴史的な高得点を記録しています：

USAMO 2025（アメリカ数学オリンピック）

- **Gemini 2.5 Pro Deep Think:** 49.4%（AI 史上前例のない高得点）
- OpenAI o3: 21.7%
- OpenAI o4-mini: 19.1%
- Gemini 2.5 Pro（通常版）：34.5%

この結果は、Deep Think が単なるパフォーマンス向上ではなく、「パターンマッチング」から「真の推論」への質的転換を示していると評価されています。

2.2 コーディング分野での最高水準

LiveCodeBench V6（競技レベルコーディング）

AIworker 社の分析⁵によると：

- **Gemini 2.5 Pro Deep Think:** 87.6%（業界最高水準）
- OpenAI o3: 71.1%
- Gemini 2.5 Pro（通常版）：71.4%

2.3 総合知識・推論能力

Humanity's Last Exam

- **Gemini 2.5 Pro Deep Think:** 34.8%
- xAI Grok 4: 25.4%
- OpenAI o3: 20.3%

MMMU（マルチモーダル推論）

- **Gemini 2.5 Pro Deep Think:** 84.0%

これらの結果から、Deep Think は「特定分野だけ強い」のではなく、AIDXLAB 検証記事⁶が指摘するように「あらゆる知的作業で人間の専門家レベル」を達成したことが確認できます。

3. ユーザーレビューと実際の使用感

3.1 肯定的評価

実際のユーザーからは以下のような高い評価が寄せられています：

開発者 A さん（Reddit 投稿より）「複雑なアルゴリズム設計で、自分では思いつかない創造的なアプローチを示してくれる。まさに優秀なチームメンバーを雇ったような感覚」

AIDXLAB[6](#)

企業 CTO の C さん「30TB + YouTube Premium + 最新 AI 機能群でこの価格なら、ストレージ代だけで元が取れる。Deep Think 機能はボーナスみたいなもの」AIDXLAB[6](#)

3.2 具体的な性能向上事例

同一プロンプトでの比較検証 [6](#) では、通常の Gemini 2.5 Pro と比較して以下の違いが確認されました：

- **Web 開発タスク**: 美観と機能性の双方で大幅な改善
- **科学・数学問題**: より深い洞察と創造的解決策の提示
- **コーディング**: 高度なアルゴリズムと最適化された実装

実際の比較例は通常モード [7](#) と Deep Think モード [8](#) で確認できます。

3.3 批判的評価と課題

一方で、以下のような批判も寄せられています：

IT 企業経営者 D さん (X 投稿より) 「月 3.6 万円で 1 日数回制限って、冗談でしょ？毎日大量に使えないと価値がない。これなら他の AI サービスの方がマシ」AIDXLAB[6](#)

一般ユーザー E さん (Reddit 投稿より) 「期待して契約したけど、普通の質問なら従来の Gemini で十分すぎる。オーバースペックもいいところ」AIDXLAB[6](#)

4. Reddit・ソーシャルメディアでの反応

4.1 Reddit r/singularity での議論

Reddit 上での活発な議論 [9](#) では、以下の点が注目されています：

1. **情報の透明性への要求**: ユーザーは「どの数学的予想を解いたのか」の詳細な説明を求めており、Google の発表における情報不足を批判
2. **価格体系への混乱**: Deep Think と Deep Research の違い、Ultra プランの料金設定（噂では\$250/月）について混乱と不満が表明
3. **OpenAI との競争**: Google が IMO 向けモデルを先に公開したことを評価する一方で、コストと計算リソースの負担を懸念する声
4. **技術的評価**: 単なる数値解ではなく「正式な数学的証明」を生成した点を高く評価するコメントが目立つ

4.2 専門家・研究者からの評価

Google 公式発表 [2](#) によると、「数学者・学術関係者の小グループ」に IMO 金メダル相当の完全版モデルへのアクセス権が付与されており、継続的なフィードバック収集が行われています。これらの専門家からのフィードバックは今後のモデル改良に活用される予定です。

5. 競合他社との比較分析

5.1 主要競合モデルとの性能比較

現在の AI 業界における位置づけを明確にするため、主要な競合モデルとの詳細比較を行いました：

数学推論能力 (OlympiadBench)

- **Gemini 2.5 Pro:** 総合スコア 10.1/42、精度 24.4%
- OpenAI O1-Pro: 1.2/42、精度 2.9%
- ChatGPT o3-mini: 0.9/42、精度 2.1%

マルチモーダル推論 (MMMU)

- **Gemini 2.5 Pro Deep Think:** 84.0%
- 他の最新モデル: 70-75%程度

5.2 技術アプローチの差別化

Gemini 革新分析記事 [10](#) によると、競合モデルとの主な差別化要因は：

1. **並列処理 vs 逐次思考:** Claude や GPT シリーズが逐次思考を採用する中、Deep Think は複数経路を同時探索
2. **推論時間の戦略的活用:** 従来モデルが速度を重視する中、Deep Think は品質向上のために時間を投資
3. **強化学習の新技术:** 延長された推論経路を積極利用するトレーニング手法

6. 利用上の制限と課題

6.1 アクセス制限

現在の Deep Think は以下の制限があります：

1. **Ultra 会員限定:** 月額 36,800 円の Google AI Ultra プランへの加入が必須
2. **利用回数制限:** 1 日あたりの Deep Think 利用回数に上限設定
3. **地域制限:** Latenode の分析 [11](#) によると、非米国ユーザーへの展開遅延が懸念されている

6.2 技術的課題

レスポンス時間

npaka 氏の調査 [3](#) によると、Deep Think モードは「数分～長時間」のレスポンス時間が必要で、即時性を求める用途には適していません。

安全性のトレードオフ

Google の公式モデルカード [12](#) によると：

- コンテンツの安全性と客観性は向上
- しかし無害なリクエストを拒否する傾向が高まる
- 複雑性増大に伴うリスクへの対応が課題

6.3 コストパフォーマンスへの懸念

多くのユーザーが指摘する主要な課題：

1. 高額な月額料金: 36,800 円という価格設定への批判
2. 利用制限との不整合: 高額な料金にも関わらず 1 日の利用回数制限
3. 対象ユーザーの限定: 一般ビジネスユーザーには過剰スペック

7. 業界への影響と今後の展望

7.1 AI 業界への波及効果

Deep Think の登場は、AI 業界に以下の変化をもたらしています：

1. 推論モデルの新標準: 「速度重視」から「品質重視」への転換点
2. 専門分野での人間代替: 数学・科学研究分野での AI 活用が現実的に
3. 価格競争の激化: 高性能モデルの価格設定に関する業界標準の見直し

7.2 今後の展開予定

Google 公式発表 [2](#) によると：

1. API 提供拡大: 数週間以内に Gemini API 経由でトラステッドテスター向けリリース
2. 開発者・企業向け展開: 開発者とエンタープライズ利用への適用可能性を検証
3. 継続的改良: 数学者・学術関係者からのフィードバックを元にしたモデル改良

8. 推奨される利用シーン

8.1 最適な利用分野

Deep Think が特に威力を発揮する分野：

1. 高度な数学・科学研究: IMO レベルの数学問題、未解決予想の探求
2. 複雑なアルゴリズム設計: 競技プログラミングレベルの問題解決
3. 学術論文作成: 深い洞察と論理的構成を要求される研究
4. 戦略的意思決定: 多角的分析を必要とするビジネス判断

8.2 避けるべき利用シーン

以下の用途には適していません：

1. 日常的な簡単なタスク：通常の Gemini 2.5 Pro で十分
2. 即時回答を要求される業務：レスポンス時間の長さが問題
3. 大量処理を必要とする作業：利用回数制限により非効率
4. コスト重視のプロジェクト：高額な料金設定が障害

9. 総合評価と結論

9.1 技術的評価

Gemini 2.5 Deep Think は、AI 技術において明確なブレークスルーを達成しています：

優れた点：

- 並列思考技術による革新的アプローチ
- 複数ベンチマークでの業界最高水準の性能
- 真の推論能力への質的転換

課題：

- 高いコンピューティングコスト
- アクセシビリティの低さ
- 実用性と性能のバランス

9.2 市場評価

ポジティブ要因：

- 専門家レベルの問題解決能力
- 競合他社に対する明確な優位性
- 研究・学術分野での革新的可能性

ネガティブ要因：

- 高額な料金体系への批判
- 利用制限による実用性の低下
- 一般ユーザーには過剰スペック

9.3 最終結論

Gemini 2.5 Deep Think は、**技術的には革命的な進歩**を遂げているものの、**商業的・実用的な課題**を多く抱えているのが現状です。このモデルは、高度な専門性を要求される限定的な用途において、その真価を発揮する「研究者向けの特殊工具」として位置づけられます。一般的なビジネス利用や日常的な AI 活用を考えているユーザーにとっては、コストパフォーマンスの観点から推奨できませんが、数学研究、科学的発見、高度なアルゴリズム開発に従事する専門家にとっては、その圧倒的な性能は投資に値する可能性があります。

今後の API 展開や価格体系の見直し、利用制限の緩和により、より幅広いユーザーがこの革新的技術の恩恵を受けられることが期待されます。

本レポートは 2025 年 8 月 1 日のリリースから現在までの公開情報、ユーザーレビュー、専門家分析を総合的に調査・分析したものです。最新の情報については、[Google 公式ブログ 2](#) および公式モデルカード [12](#) をご確認ください。

Appendix: Supplementary Video Resources



Gemini 2.5 Deep Think が登場！実力を試してみた

13 hours ago



人間的ように思考！Google AI Ultra「Gemini 2.5 Deep Think ...

1 day ago



Gemini Deep Think Review

17 hours ago

もっと詳しく

www.watch.impress.co.jp

2
blog.google

3
note.com

4
digirise.ai

5
note.com

6
note.com

7
gemini.google.com

8
gemini.google.com

9
www.reddit.com

10
note.com

11
latenode.com

12
storage.googleapis.com