

Gemini 3 徹底調査レポート：性能、評判、競合比較

作成日: 2025年11月19日 作成者: Manus AI

エグゼクティブサマリー

2025年11月18日にGoogleが発表したGemini 3は、AIの能力を新たな次元へと引き上げる画期的なモデルです。本レポートでは、公式発表、ベンチマーク分析、YouTubeでの解説、ソーシャルメディアでの評判など、多角的な情報源を基にGemini 3の全貌を徹底的に調査・分析しました。

調査の結果、Gemini 3 Proは、特に博士号レベルの科学的推論能力 (GPQA Diamond: 91.9%)、高度な数学的能力 (MathArena Apex: 23.4%)、そしてエージェントとしてのコーディング能力 (Terminal-Bench 2.0: 54.2%)において、GPT-5.1やClaude Sonnet 4.5といった主要な競合モデルを明確に上回る性能を達成していることが明らかになりました。これは、単なる性能向上ではなく、「知性の新時代」と呼ぶにふさわしい技術的ブレークスルーと言えます。

本レポートでは、これらの性能を支える技術的背景、具体的な機能、ユーザーからの初期評価、そして潜在的な弱点について詳細に解説し、Gemini 3がAI業界および社会全体に与える影響を考察します。

1. 公式発表と「知性の新時代」の背景

Googleは、2025年11月18日に公式ブログを通じてGemini 3を発表しました¹。この発表は、単なるモデルのアップデートではなく、「知性の新時代」の幕開けとして位置づけられています。その背景には、以下の技術的進歩があります。

Gemini 3は、あらゆるアイデアを実現するのに役立つ、最も高性能なモデルです。創造的なアイデアの微妙なヒントを認識したり、困難な問題の重なり合う層を剥がしたりと、深さとニュアンスを理解するように構築されています。¹

この「深さとニュアンスの理解」は、Gemini 3が単なる情報処理ツールから、ユーザーのコンテキストと意図を深く理解する思考パートナーへと進化したことを示しています。この設計思想の転換こそが、「知性の新時代」の核となる概念です。

主要な新機能

- Gemini 3 Pro:** 発表と同時にプレビュー版がリリースされた主力モデル。
- Gemini 3 Deep Think:** より深い推論能力を持つ高性能モード。GPQA Diamondで93.8%という驚異的なスコアを記録。

- **Google Antigravity**: エージェント機能に特化した新しい開発プラットフォーム。
- **Thinking Level**: APIでモデルの思考深度を制御できる新機能。
- **Media Resolution**: 画像や動画の解析解像度をタスクに応じて調整できる機能。

2. GPQAスコア91.9%の衝撃と性能分析

Gemini 3 Proの性能を象徴するのが、**GPQA Diamondベンチマーク**で達成した91.9%というスコアです。このスコアが持つ意味を、競合モデルとの比較を通じて分析します。

GPQA Diamondベンチマークとは?

GPQA (Graduate-Level Google-Proof Q&A) は、生物学、物理学、化学の分野における**博士号レベルの科学的知識**を問う、極めて難易度の高いベンチマークです²。Google検索でも容易に答えが見つからないよう設計されており、人間のPhD専門家でさえ正答率は65%～69.7%に留まります³。

競合モデルとの性能比較

Gemini 3 Proのスコアは、人間の専門家を大きく凌駕するだけでなく、主要な競合モデルに対しても明確な優位性を示しています。

モデル名	GPQA Diamond スコア (%)	Gemini 3 Proとの差分
Gemini 3 Deep Think	93.8	+1.9
Gemini 3 Pro	91.9	-
GPT-5.1	88.1	-3.8
Gemini 2.5 Pro	86.4	-5.5
Claude Sonnet 4.5	83.4	-8.5
人間 (PhD専門家)	65.0 - 69.7	-22.2～-26.9

この結果は、Gemini 3 Proが複雑な科学的概念を理解し、高度な推論を適用する能力において、**超人的なレベル**に達していることを客観的に示しています。特に、前モデルであるGemini 2.5 Proからの5.5ポイントの向上は、この難易度のベンチマークにおける**大きな技術的飛躍**を意味します⁴。

3. YouTubeでの評判：「思考」能力と「次元の違い」

提示されたYouTube動画のタイトルにある「思考は新次元へ」「次元違いすぎた」といった評価は、単なる性能向上だけでなく、Gemini 3の質的な変化を指しています。調査の結果、この「次元の違い」は以下の2点に集約されます。

- 設計思想の転換:** ユーザーが「聞きたいこと」に答えるのではなく、問題解決のために「聞くべきこと」を提示する思考パートナーへと進化した点⁵。
- 高度なマルチモーダル統合:** テキスト、画像、動画、音声、コードをシームレスに統合し、人間のように多角的な情報から推論する能力¹。

例えば、ピックルボールの試合動画を分析してフォーム改善のトレーニングプランを生成したり、手書きのメモからインタラクティブなWebアプリを生成したりするなど、これまで不可能だったレベルのタスクを実行可能です^{1 6}。

4. 技術仕様と実用面での変更点

Gemini 3は、実用面でも多くの重要な変更が加えられています。

項目	詳細	実用面でのインパクト
入力コンテキストウィンドウ	1,000,000トークン	大量のドキュメント、コードベース、長時間の動画全体を一度に読み込み、文脈を維持したまま分析可能。
API価格	\$2/1Mトークン（入力）、\$12/1Mトークン（出力）	性能向上にもかかわらず、競争力のある価格設定。コンテキストサイズに応じた段階的料金体系。
Thinking Level	APIで思考深度を High と Low で制御可能	速度が求められるタスクと、深い思考が求められるタスクで、コストと性能のバランスを最適化できる。
Media Resolution	APIで画像・動画の解析解像度を制御可能	トークン消費量とレイテンシをタスクに応じて調整し、コスト効率を向上させる。
Thought Signatures	API呼び出し間で推論コンテキストを維持	複雑なエージェントワークフローにおいて、一貫した思考プロセスを保証する。

5. ソーシャルメディアでの初期評価（強みと弱点）

RedditやHacker Newsなどの技術コミュニティでは、Gemini 3に対して非常に高い興奮と期待が寄せられています。特に、その圧倒的な推論能力とエージェントとしての可能性が注目されています。

"guys, we will get a monster of a computer use agent with Gemini 3." (Redditユーザー) 7

強みとして評価されている点

- **圧倒的な数学・科学能力:** MathArena Apexでのスコア（23.4%）は、競合（1%台）を大きく引き離し、研究開発への応用が期待されている 8。
- **ハルシネーションの大幅な削減:** SimpleQAベンチマークでGPT-5.1の2倍以上の信頼性を達成 8。
- **エージェント能力:** ターミナル操作やWeb開発のベンチマークで最高スコアを記録し、自律的なタスク実行能力の高さが評価されている。

弱点として指摘されている点

- **回答の誤り:** 一部のユーザーから、特定のプログラミング問題で誤った回答が生成されたとの報告がある 8。
- **検索ツールの挙動:** 検索機能が有効な場合、問題解決よりも検索結果の要約を優先する傾向があるとの指摘 8。
- **プライバシー懸念:** 一部の環境でGoogle Driveへのアクセス許可を求めるUIが表示され、ユーザーに混乱と懸念を与えている 8。

6. 競合モデルとの性能比較

Gemini 3 Proは、多くの分野で競合モデルを凌駕していますが、各モデルには依然として得意分野が存在します。

分野	Gemini 3 Pro	GPT-5.1	Claude Sonnet 4.5
論理推論・科学	◎	○	△
高度な数学	◎	△	△
コーディング（エージェント）	◎	○	○
コーディング（バグ修正）	○	○	◎
創造的タスク（長文）	○	◎	○

総括:

- **Gemini 3 Pro:** 総合的な推論能力、高度な数学・科学、マルチモーダル理解、エージェント能力において明確なリーダー。
- **GPT-5.1:** 長文ライティングや汎用性に優れ、安定した性能を発揮。
- **Claude Sonnet 4.5:** 特定のコーディングタスク（SWE-Bench）において最高の性能を維持。

結論

Gemini 3は、単なる性能向上に留まらず、AIの「思考」の質を根本的に変える可能性を秘めたモデルです。特に、博士号レベルの科学的推論能力や、人間のように複数の情報源から統合的に判断するマルチモーダル機能は、研究開発、医療、教育、ソフトウェア開発など、専門的な分野に革命をもたらす潜在能力を秘めています。

一方で、ユーザーから指摘されている回答の信頼性やプライバシーに関する課題は、今後の改善が求められる領域です。Googleがこれらの課題にどう対応し、Gemini 3の圧倒的な性能を誰もが安全かつ便利に利用できる形に昇華させていくのか、その動向が注目されます。

Gemini 3の登場により、AI開発の競争は新たなステージに突入しました。今後、競合他社がどのように追随・対抗していくのかを含め、AI業界全体の技術革新がさらに加速することは間違いないでしょう。

参考文献

- [1] Google. (2025, November 18). A new era of intelligence with Gemini 3. Google Blog.
- [2] Rein, D., et al. (2023). GPQA: A Graduate-Level Google-Proof Q&A Benchmark. arXiv.
- [3] ArtificialAnalysis.ai. (2025). GPQA Diamond Benchmark Leaderboard.
- [4] 9to5Google. (2025, November 18). Google launches Gemini 3 with state-of-the-art reasoning, '....
- [5] ZDNET Japan. (2025, November 19). グーグル、最新AIモデル「Gemini 3」を公開--博士号レベルの推論も可能に。
- [6] BinaryVerse AI. (2025, November 18). Gemini 3: 11 Proven Wins Over GPT-5.1, Claude In Benchmarks.
- [7] Reddit. (2025, November 18). Gemini 3.0 Pro benchmarks leaked. r/singularity.
- [8] Hacker News. (2025, November 18). Gemini 3.