

Google Gemini 3 – Performance, Capabilities & Reputation (Nov 2025)

Abstract representation of multi-modal AI

概要とリリース背景

Google は 2025 年 11 月 19 日に **Gemini 3** を発表し、歴代の Gemini モデルで培ったすべての機能を統合しました。サンダー・ピチャイ CEO は、この 2 年間で Gemini アプリの月間ユーザーが 6.5 億人に達し、13 百万人の開発者が同社の生成モデルを利用していることを強調しています¹。Gemini 3 はこうした利用実績を踏まえ、推論・マルチモーダル理解・エージェント機能を総合的に強化したモデルとして登場しました。リリース当日から Google 検索の AI モードへ統合されたのは今回が初めてで²、同時に Gemini アプリ、Google AI Studio、Vertex AI、そしてエージェント開発プラットフォーム「Google Antigravity」に展開されています³。

アーキテクチャと中核機能

Sparse Mixture-of-Experts と大規模コンテキスト

Gemini 3 は先代の Gemini 2 系列を基にしたのではなく、ゼロから設計された **ネイティブなマルチモーダル Transformer** です。アルゴリズム専門誌のレビューによれば、モデルカードでは「**スパース・ミクチャー・オブ・エキスパート (MoE)**」構造を採用し、最大 **100 万トークン**の入力（64 k トークンの出力）を受け付けると記されています⁴。この構造により従来より大きな文脈を維持したまま、画像・動画・音声・コードなど異なるモダリティの情報を同じエンコーダで処理できます⁵。多くの実装解説では従来 2.5 Pro で 2 M トークンだったコンテキスト窓がさらに拡張され、長大な資料やコードベースを読み込んでも整合性が保たれると指摘されています⁶。

エージェント機能と深い推論

Gemini 3 は「**Agentic**」な行動が可能です。Google のブログでは、本モデルが自律的に計画を立てて複雑なマルチステップのワークフローを遂行し、自身のコードを検証できると説明されています⁷。これにより、Gmail 受信トレイの整理や旅行プランの作成など、長い流れのタスクをユーザーの指示に基づいて実行することが可能になりました⁸。また、**Generative UI** 機能では検索クエリに応じてミニアプリやシミュレーションを動的に生成し、3 体問題の物理シミュレーションや住宅ローン計算ツールをその場で提供するなど、検索体験を刷新しています⁹。

100 万トークンのコンテキストとマルチモーダリティ

広大なコンテキスト窓とネイティブなマルチモーダル処理により、Gemini 3 は複数の形式の情報を統合したタスクに強みを発揮します。例えば、手書きのレシピを異なる言語から解読して家族用の料理本を生成したり、学術論文や長時間の講義を分析してインタラクティブなフラッシュカードや視覚化を作成したりすることができると公式ブログは説明します¹⁰。プラズマ実験装置トカマクの内部流動を視覚化するコード生成までこなせる例も示されており¹¹、科学技術分野での応用範囲が拡大しています。

主なベンチマーク成績

Google と外部評価機関が発表したベンチマーク結果は、Gemini 3 Pro が先行モデルや競合モデルに対して大幅なリードを示すことを裏付けています。以下は重要な指標の一部です（単位は % や Elo 等）。

ベンチマーク/カテゴリ	Gemini 3 Pro (標準)	Deep Think モード	備考
LMarena (一般的な会話品質)	1501 Elo ¹²	–	LMSYS Arena のリーダーボードでトップ
Humanity’s Last Exam (学術推論, ツールなし)	37.5 % ¹³	41.0 % ¹⁴	PhD レベルの推論を測る試験
GPQA Diamond (科学知識)	91.9 % ¹⁵	93.8 % ¹⁴	大学院レベルの科学問題
MathArena Apex (難問数学)	23.4 % ¹⁵	–	数学オリンピックレベル問題の新ベンチマーク
MMMU-Pro (マルチモーダル理解)	81 % ¹⁶	–	画像・テキストなど複数モダリティ統合
Video-MMMU (動画理解)	87.6 % ¹⁶	–	長尺動画の理解を評価
SimpleQA Verified (事実性 / 検証可能性)	72.1 % ¹⁷	–	事実の正確性を測定し、Gemini 2.5 の 54.5 % から大幅向上 ¹⁸
WebDev Arena (Web開発能力)	1487 Elo ¹⁹	–	インタラクティブなフロントエンド生成能力を評価
SWE-bench Verified (コーディングエージェント)	76.2 % ¹⁹	–	実在の GitHub 問題に対する修正成功率
Terminal-Bench 2.0 (ターミナル操作)	54.2 % ¹⁹	–	コマンドラインを使ったツール操作能力
Vending-Bench 2 (長期計画)	–	–	シミュレーションされた自動販売機ビジネスで最も高い収益を達成 ²⁰
ARC-AGI 2 (一般知能テスト, コード実行あり)	31.1 % ²¹	45.1 % ¹⁴	難易度の高い視覚パズル。Deep Think では 45.1 % に跳ね上がる

他にも、Gemini 3 Pro は **AIME 2025 (数学)** で 95 %（コード実行時は 100 %）を達成したことや ²²、**ScreenSpot-Pro** で 72.7 % の UI 理解スコアを記録し従来のコンピューター利用モデルを大きく上回ったと報告されています ²³。これらの結果は、長期計画やツール操作など実践的なエージェントタスクで強みを持つことを示しています。

学習・開発・計画を支援する機能

学習支援

Gemini 3 はテキスト・画像・動画・音声・コードをシームレスに統合するよう設計されており、**100 万トークンのコンテキスト窓**と優れた視覚理解、強力な多言語性能を組み合わせることで学習体験を向上させます²⁴。公式ブログでは、手書きレシピの解説・翻訳や学術論文・長時間のビデオ講義からインタラクティブな教材を生成するなど、個人の学習を支援する具体例が紹介されています¹⁰。

開発支援

Gemini 3 は Google が「**Vibe Coding**」と呼ぶ自然言語からのアプリ構築能力を備えており、複雑なプロンプトからインタラクティブな Web UI を生成できます。WebDev Arena での 1487 Elo スコアは Gemini 2.5 を大きく超え、Terminal-Bench 2.0 での 54.2 % スコアもツール操作能力の向上を示しています¹⁹。さらに、Antigravity はエージェントがエディター・ターミナル・ブラウザへ直接アクセスし、フルスタックアプリを自律的に計画・実行する新しい IDE であり⁷、Gemini 3 と組み合わせることでエンドツーエンドなソフトウェア開発支援が可能になっています。

計画支援

Vending-Bench 2 での高得点に示されるように、Gemini 3 は長期的な計画と意思決定に優れています。ブログによると、モデルは1年間の自動販売機ビジネスシミュレーションで一貫したツール使用と意思決定を維持し、競合モデルよりも高い収益を上げました²⁰。この長期計画能力は、メール整理や旅行プランニングなど日常的な複数ステップのワークフローでも活用できると説明されています⁸。

Deep Think モード

Google は Gemini 3 の強化推論モードとして **Deep Think** を導入しました。Deep Think は標準モデルに比べてさらなる計算リソースを使用し、多段階の推論や複数の解法探索を行います。Humanity's Last Exam で 41.0 %、GPQA Diamond で 93.8 %、ARC-AGI 2 で 45.1 % を達成しており¹⁴、特に ARC-AGI 2 のスコアは他モデルの 2〜3 倍であると専門誌が驚きをもって報じています²⁵。ただし Deep Think は安全性評価中であり、Google AI Ultra のサブスクリプションユーザー向けに段階的に提供される予定です²⁶。

料金・利用可能性

The Decoder によると、Gemini 3 Pro の API 料金は 200 k トークン以下の場合で **入力 1 M トークンあたり 2 ドル、出力 1 M トークンあたり 12 ドル**とされています²⁷。これは Gemini 2.5 Pro より高価ですが、Claude 4.5 Sonnet や GPT-5 Pro など他のハイエンドモデルよりは安価です²⁸。200 k トークンを超えるコンテキストでは価格が 4 ドル/18 ドルに上昇し、Deep Think モードはさらに高価になる見込みです²⁹。Gemini 3 は Gemini アプリを通じて一般ユーザーに提供され、開発者は Google AI Studio や Vertex AI、Gemini CLI、Antigravity から利用できます³。²⁶で示されているように、Deep Think は安全性評価後に Ultra サブスクライバー向けに展開される予定です。

安全性・制限事項

Gemini 3 は前モデルより安全性が強化されており、迎合性の低下やプロンプトインジェクションへの耐性向上、サイバー攻撃からの保護が図られています³⁰。Google は英国の AI Safety Institute など外部機関と協力し独立した評価を実施しています³¹。ただし独立アナリストによれば、Gemini 3 Pro は高い知識テスト精度を示す一方、競合モデルより **幻覚率が高い** 可能性があると指摘されています³²。アルゴリズム専門誌

でも、ベンチマークはモデルの本質的能力を完全には表さないことや³³、ARC-AGI のタスクがコンテキスト窓サイズに依存する可能性があることなど、評価の解釈に注意を促しています³⁴。

評判と評価

正の評価

- **技術メディアの賞賛** – WIRED は Gemini 3 を Google 史上最もインテリジェントなモデルと呼び、GPT-5 など他モデルを上回る推論力と長期計画能力を備えていると報じています³⁵。Tom's Guide のレビューでは、Gemini 3 が Gemini 2.5 に比べて驚くほど速く賢くなったと述べ、検索の「Thinking」モードで複雑な質問に深い文脈的回答を返したことが紹介されています³⁶。
- **ベンチマークでの優位** – 多数の公式ベンチマークで Gemini 3 Pro がトップスコアを獲得したことが専門家や開発者の注目を集めています。Artificial Analysis のインテリジェンス指数でも GPT-5.1 を 3 ポイント上回り、10 評価のうち 5 で首位となったと報じられています³⁷。
- **エコシステム統合** – リリース初日から Google 検索に統合されたほか、Antigravity など開発者向けプラットフォームを通じてエージェント型のワークフローが体験できる点が高く評価されています²。

批判的な評価・課題

- **料金の高さ** – Gemini 3 Pro の API 料金は Gemini 2.5 Pro や GPT-5.1 に比べて高く²⁷、Deep Think モードではさらにコストがかさむため、コスト効率を重視する用途では競合モデルが選ばれる場合があります。第三者の記事では、Gemini 3 Pro の利用により Artificial Analysis のベンチマーク実行コストが 12 % 増加したと指摘されています³⁸。
- **幻覚率と安定性** – Artificial Analysis は Gemini 3 Pro の幻覚率が競合より高いと報告し、Google はモデルカードで hallucination metrics を明示していません³⁹。緩やかな安全性向上にもかかわらず完全な無害化は達成していないため、ユーザーは重大な判断を下す際には注意が必要です。
- **ベンチマーク依存への警告** – The Algorithmic Bridge はベンチマークの得点がモデル品質の全てを示すわけではないと強調し、実際の業務においては既存モデルでも十分な性能を持つ場合が多いと述べています⁴⁰。研究者の間でも、ARC-AGI の成績向上がコンテキスト窓の拡大によるものである可能性が指摘されています⁴¹。
- **競合の存在** – GPT-5.1 や Claude Sonnet 4.5 など他社モデルも独自の強みを持ち、例えば GPT-5.1 はコスト効率と高速性に優れ⁴²、Claude シリーズは安全性や長時間作業での安定性を重視しています⁴³。用途によっては競合モデルが適しているケースもあります。

結論

Gemini 3 は Google が推進する次世代 AI の象徴であり、ネイティブなマルチモーダル理解・深い推論能力・長期計画機能を備えた高性能モデルです。LMArena や GPQA Diamond など多くのベンチマークで過去最高スコアを記録し¹⁵、Generative UI や Antigravity を通じて AI の利用体験を刷新しています。一方で、料金の高さや幻覚率の問題、ベンチマーク偏重への懸念といった課題も存在します。ユーザーは用途やコストに応じて Gemini 3 と競合モデルを比較し、必要に応じて安全性評価を行うことが重要です。今後 Deep Think の一般提供や追加モデルのリリースが予定されており⁴⁴、Gemini 3 シリーズがどのように発展するか注目されます。

¹ ² ³ ⁷ ⁸ ¹⁰ ¹¹ ¹² ¹³ ¹⁴ ¹⁵ ¹⁶ ¹⁷ ¹⁹ ²⁰ ²⁴ ²⁶ ³⁰ ³¹ ⁴⁴ Gemini 3: Introducing the latest Gemini AI model from Google

<https://blog.google/intl/ja-jp/company-news/technology/gemini-3/>

- 4 18 21 25 33 34 40 41 **Google Gemini 3 Is the Best Model Ever. One Score Stands Out Above the Rest**
<https://www.thealgorithmicbridge.com/p/google-gemini-3-just-killed-every>
- 5 42 43 **Gemini 3 Pro: First Reviews. Launched on November 18, 2025 and... | by Barnacle Goose | Nov, 2025 | Medium**
<https://medium.com/@leucopsis/gemini-3-pro-first-reviews-527120cebe84>
- 6 **Gemini 3.0 Released TODAY – First Hands-On: This Thing Is Scary Good**
<https://apidog.com/blog/gemini-3-0/>
- 9 **Google Search AI Mode gets a substantial upgrade with Gemini 3**
<https://chromeunboxed.com/google-search-ai-mode-gets-a-massive-upgrade-with-gemini-3/>
- 22 **Google Gemini 3 Released: Features & Benchmarks 2025**
<https://max-productive.ai/blog/google-gemini-3-release/>
- 23 27 28 29 32 37 38 39 **Analysts say Google now leads the AI performance race with Gemini 3 Pro**
<https://the-decoder.com/analysts-say-google-now-leads-the-ai-performance-race-with-gemini-3-pro/>
- 35 **Gemini 3 Is Here—and Google Says It Will Make Search Smarter | WIRED**
<https://www.wired.com/story/google-launches-gemini-3-ai-bubble-search/>
- 36 **Google's Gemini 3 is here — and it's already beating ChatGPT in key benchmarks | Tom's Guide**
<https://www.tomsguide.com/ai/google-gemini/gemini-3-is-here-googles-most-powerful-ai-model-yet-is-crushing-benchmarks-improving-search-and-outperforming-chatgpt>