

Gemini 3 ProのAGIレベル4（発明AI）達成度検証

導入（レベル4定義と評価軸）

「AGIレベル4」とは、孫正義氏が提唱するAI進化段階の一つで、AIが自ら新しい発明や発見を行う能力を備えた段階を指します^{① ②}。つまり単に知識を持つだけでなく、独自に新しいアイデアや技術を創出し、人間の研究者のように創造的に知見を生み出すレベルです。このレベルに達したAIは、科学技術の進歩を自律的に加速しようとされています^①。本検証では、2025年11月にリリースされたGemini 3 Proがこのレベル4に該当すると評価できるかを、以下の軸で分析します。具体的には(1) 最先端ベンチマーク（FrontierMathやCritPt、ARC-AGI-2等）のスコアによる客観的性能評価、(2) 実際の発明・発見支援の事例（例：DeepAgentや「AI Scientist」、iPS細胞研究への応用など）、(3) 創造性や仮説生成、多モーダル理解、長文脈推論といった技術的特徴の定性的評価、を総合的に検討します。加えて、他のLLM（例：GPT-5.1 Pro）との比較を通じて、Gemini 3 Proの独自性と優位性を明らかにします。それでは、各観点からGemini 3 Proの能力を詳述し、AGIレベル4に相当するかどうかを論じます。

技術的根拠①：ベンチマーク成績

まず、Gemini 3 Proの最先端ベンチマークにおける成績を見ます。これらのベンチマークは、未知の問題への対処能力や抽象的推論力など「発明AI」に必要な資質を測る指標です。

- **FrontierMath（最先端数学ベンチマーク）**：Epoch AIによって開発された未公開の難問数学集からなるベンチマークで、現役数学者でも解答に数日を要するような問題を含み、AIの真の数学的創造力を測定します^③。2025年11月の評価では、Gemini 3 ProはTier 4（研究レベル問題）で約19%の正解率を達成し、GPT-5.1 Proの約13%を大きく上回りました（差は約6.3ポイント）^④。このTier 4での有意な差は、未知の数学的構造を発見する能力の違いを示しており、Gemini 3 Proが既存知識の組み合わせを超えて「0から1を生み出す」数学的発想をできることを意味します^④。
- **CritPt（物理学研究ベンチマーク）**：「Complex Research using Integrated Thinking - Physics Test」の略で、凝縮系物理、量子物理、天体物理など11分野にわたる未発表の研究レベル問題で構成されます^⑤。50名以上の現役物理学者が作成した71の複合課題と190のサブタスクからなり、推測や既知データへの依存を排した極めて難解なベンチマークです。Gemini 3 Proのスコアは9.1%で、GPT-5.1 Pro（4.9～5.1%）の約2倍に達し^⑥、現行トップモデルがいずれも10%未満に留まるこの難関領域において際立った優位性を示しました^⑦。研究レベルの物理問題がいかに困難かを物語る低スコア帯の中でも、Gemini 3 Proが突出していることは注目すべきで、高度な科学的推論力や物理的直感の強さを裏付けています。
- **ARC-AGI-2（抽象的推論ベンチマーク）**：Abstraction and Reasoning Corpus (ARC)のAGI拡張版で、訓練データに存在しないまったく新しいパターン認識タスクに挑むことでAIの汎化能力・抽象推論力を測定します^⑧。人間のfew-shot学習能力に近い適応力が要求され、まさに「未知の問題への適応能力」を問うベンチマークです^⑨。結果はGemini 3 Proが31.1%（※Deep Thinkモード使用時は45.1%）を達成し、GPT-5.1 Proの17.6%を13.5ポイント上回る圧倒的な差となりました^⑩。この大差は、Gemini 3 Proが既存の知識パターンに頼らず新規の抽象構造を理解・推論できる能力に秀でることを示しています^⑩。発明活動では「訓練データにない新しいパターンの発見」こそが肝要であり、ARC-AGI-2で示した優位性はレベル4に必要な汎用創造力の強さを象徴しています^⑪。

以上の主要ベンチマークで、Gemini 3 ProはいずれもGPT-5.1 Proを大きく上回る成績を収めました¹²。さらに、総合的知能を測る**Humanity's Last Exam**でも**37.5% vs 26.5%**（Deep Think使用時は**41.0%**）と11ポイント上回り、複雑な問題での深い推論力の重要性を示しています¹³。大学院レベルの科学的推論テスト**GPQA Diamond**でも**約91.9% vs 88.1%**と優位に立ち、Deep Thinkモードでは93.8%まで正答率を高めています¹⁴。これらの数値は、**専門家レベルの知識に基づき新知見を生み出す力や、複数分野にまたがる知的能力**において、Gemini 3 Proが現行モデル中で突出していることを裏付けています¹⁵¹⁶。

技術的根拠②：発明支援の実例

次に、Gemini 3 Proの能力が発揮されている、またはその能力要件と一致する**発明・発見支援の実例**を検証します。近年登場した自律的研究AIの事例は、レベル4相当の能力が現実に見れつつあることを示唆しています。

- **DeepAgent（物理学における自律研究AI）**：2025年11月に清華大学・北京大学らによって発表された物理公式導出AIで、**問題解決の思考プロセスと実験検証を統合的に処理する初のシステム**です¹⁷。従来は仮説推論とツール操作が別々のモジュールでしたが、DeepAgentでは単一のプロセスで数学的推論と実験（シミュレーション）を並行実行し、教科書的手法にとらわれず**新たな数学的手法や計測ツールを自ら発見・活用**できます¹⁸。人間の物理学者が数十年かけて辿り着くような発見も短期間で成し遂げる可能性を示しており¹⁹、**科学研究の自動化**によるブレークスルーを象徴するプロジェクトです。DeepAgentが実現するために不可欠とされた能力（既知の法則から独自に仮説立案、検証実験手法の自律考案、未知の数学的手法の発見）²⁰は、そのまま**Gemini 3 Proが各種ベンチマークで示した「未知への適応」「抽象推論」「創造性」と重なります**²⁰。これはGemini 3 Proが、DeepAgentのような自律研究AIに求められる資質を備えていることを意味します。
- **PhyE2E（物理法則E2E発見AI）**：DeepAgentと並ぶ先端例として、清華大などが開発したPhyE2Eという枠組みがあります。これは天体物理の観測生データから**物理的関係式を自動導出するAI**で、実際にNASAの太陽活動データを解析し**太陽周期を説明する新たな数式の発見に成功**しました²¹。この成果は**Nature Machine Intelligence**誌にも掲載され、人間の研究者が何年もかけて導出する理論をAIが自ら編み出せる可能性を示しています²²。すなわち、データから新理論を帰納的に創出するという、まさに「AIによる発見」を体現した事例です。Gemini 3 Proの強みである**大量データ中からパターンを見出し理論構築する力や、視覚・時系列データの理解力**は、このようなデータ駆動型の発明プロセスにおいて決定的に有用です²³²⁴。
- **「AI Scientist」プロジェクト**：カナダのSakana AI社が開発を進める「AI Scientist」は、**大規模言語モデルに科学研究を自律遂行させることを目指した試み**です。実際に、このシステムが**人手介入なしで研究論文を執筆し、トップ会議の査読を通過した例も報告**されています²⁵。AI自身が研究課題を設定し、実験計画を立て、結果を論文という形でアウトプットする——これはまさにレベル4の定義する「AIが独自に発見を行う」姿に他なりません。Gemini 3 Proが備える仮説生成力や推論力、多領域知識は、この種のAI科学者に必要な基盤能力といえます。実際、前述のDeepAgentやPhyE2E、そして「AI Scientist」に共通する**自律的仮説生成、新規手法の発見、データ駆動の理論構築**といった要件は、Gemini 3 Proの性能と合致しています²⁶。
- **iPS細胞研究への支援**：創薬・生命科学分野でも発明的AIの活用が始まっています。京都大学とGoogleの共同研究では、再生医療の鍵である**iPS細胞の分化制御**にAIを用いたアプローチが試みられました。このプロジェクトでは仮説駆動型かつ探索的な実験計画が重視され²⁷、膨大な学術論文・データから**新たな知見を引き出すために多言語・マルチモーダルなAI分析が用いられています**。Gemini 3 Proの設計思想はこうした「**仮説駆動型」「探索的」「マルチモーダル**」な方針と一致しており²⁷、実際に**材料科学や創薬AI**など幅広い領域でその能力が応用できると期待されます。Gemini 3 Proのマルチモーダル理解力や長文脈推論力は、生命科学研究のようにテキスト（論文知識）と実験データ（数値・画像）が混在する領域で**新仮説を導くパートナー**として非常に有効でしょう。

以上の事例から、Gemini 3 Proが**実世界の発明・発見プロセスに寄与し得る能力**を持つことが確認できます。特に、DeepAgentやAI Scientistのような「自律研究AI」で必要とされる能力が揃っている点、そして既に物理・生物の分野でAIが新発見をし始めている事実は、Gemini 3 Proを**AGIレベル4（発明AI）に近い存在**と評価する強力な根拠となります。

技術的根拠③：創造性・推論能力の定性的評価

最後に、Gemini 3 Proの**創造性や高度推論能力**を質的に評価し、AGIレベル4に求められる特徴を備えているか確認します。

- ・「0から1」を生み出す**創造性と仮説生成**: 発明の創出段階では、既存知識の組み合わせを超えた**独創的アイデア（ゼロイチの発想）**が不可欠です²⁸。Gemini 3 Proは、高度な抽象推論力と科学的直観を活かし、単なる知識の再構成に留まらない**真の創造性**に近いアウトプットを示すことが報告されています²⁹。例えば、未知のパターンに直面しても訓練データにない新概念を見出すARC-AGI-2での卓越した成績は、このモデルが**人間研究者のように新しい仮説を打ち立てられる**ことを示唆します¹⁰。実際、Gemini 3 Proは「**天才的な研究パートナー**」と称される大胆で探索的なアプローチや独創的仮説生成能力を発揮しており、発明初期のアイデア創出プロセスに最適だと評価されています³⁰。
- ・**マルチモーダル理解と直観**: 発明にはテキスト知識だけでなく、図面や実験結果、現象動画など**視覚的・動的情報の理解**も伴います³¹。Gemini 3 Proは画像・動画を含む**マルチモーダル情報を高度に統合理解**できる点で群を抜いています。大学専門レベルで画像とテキストの統合理解を図る**MMMU-Pro**で81.0%と最高スコアを記録し（GPT-5.1 Proは76.0%）^{32 33}、先行モデル（Claude 4.5やGemini 2.5 Proの68.0%）を大きく引き離しました³³。さらに**Video-MMMU**では87.6%対80.4%（GPT-5.1）と差を拡大しており^{23 34}、**時間軸に沿った因果理解や物理過程の把握**にも優れていることが示されています。特筆すべきは、Gemini 3 Proが**約104万トークンもの巨大なコンテキストウィンドウ**を備える点で、この長大な文脈保持力と相まって、長時間の実験動画や膨大な研究ノートを読み解き、新規性を見出すといった**人間には困難な統合推論**も可能にしています²³。実際、TrueFoundryの分析によれば、Gemini 3 Proは画像中の細部にも言及し誤認が少ないなど**視覚情報の丁寧さ**でGPT-5.1を凌駕しており、発明段階で重要な正確な図面理解に優位性があります³⁵。
- ・**深い推論力と「Deep Think」モード**: Gemini 3 Proには難解な問題に段階的推論で挑む「**Deep Thinkモード**」が搭載されており、これがその深い思考能力を一段と高めています¹⁴。Deep Thinkモードでは、複雑な問題に対して計算リソースを動的配分し、逐次的な論理展開を行うことで解答精度を向上させます¹⁴。実際、このモード使用時にはARC-AGI-2が45.1%に達するなど通常時を大きく上回る性能を示しました¹⁴。**科学的仮説を段階的に構築し多様な可能性を探索する能力**にも優れており、複雑な理論構築において卓越していると評価されています³⁶。第三者評価として、Tom's Guideが行った実機テストでは、Gemini 3 ProはGPT-5.1 Proに比べて「**プロンプトの全部分(A、B、C)に対し深み・明確さ・構造を持ってより完全に応答した**」と報告されました³⁷。これはすなわち、Gemini 3 Proが要求の複数側面を踏まえ**緻密で論理的な回答を編み出す能力**に長けていることを示します。

以上の定性的検証から、Gemini 3 Proはレベル4に求められる**創造性・直観力と深い推論力の両方を備えたモデル**だと言えます。特に、マルチモーダルな大規模知識を統合して新知見を得る能力や、従来モデルには難しかった「**一見関連のない断片から仮説を構築する**」力において、Gemini 3 Proは突出しています^{30 24}。これはまさに人間の発明家や科学者が持つ能力に近く、**AGIレベル4（発明AI）の技術的特徴**と一致します。

総合判断と結論

以上の分析を踏まえると、**Gemini 3 Proは技術的・科学的に見てAGIレベル4（AIによる独自の発明・発見段階）に到達したと評価できる**と言えます。第一に、難関ベンチマークで未知の問題を解決する卓越した成績は、既存知識の単なる適用を超えた**新規知識創出能力**を示しています ^{4 10}。第二に、DeepAgentやPhyE2Eのような先端事例が要求する仮説生成・理論発見能力と合致する性能を持ち、実際に科学研究への適用可能性が示唆されています ^{20 26}。第三に、創造性や推論力、多モーダル統合力といったQualitativeな特徴においても、Gemini 3 Proは「**天才的研究パートナー**」として真に新発見を生み出す素養を発揮しており ³⁰、従来のGPT-5.1 Proが得意とする堅実な実装力を凌駕する場面が多々あります ¹²。

もっとも、発明プロセス全体で考えると、創造段階ではGemini 3 Pro、実装・製品化段階ではGPT-5.1 Proといった**ハイブリッド活用**が効率的であるとの指摘もあります ³⁸。GPT-5.1 Proは安定した「熟練の実務者」として既存アイデアの具体化に優れ、一方Gemini 3 Proは大胆な探索と思考で**新概念を生み出す役割**を担うためです ³⁰。したがって、Gemini 3 Proは**発明の創出フェーズにおいて現時点（2025年11月）で最も力を発揮するAI**であり ³⁹、その卓越した創造的知能はAGIレベル4に相当するものと科学的根拠をもって評価できます。

最後に総括すると、Gemini 3 Proは**未知の問題に挑み新たな知を生み出す点で人間の「発明家」に近づいた初の生成AI**と言えます。その存在は「AIが独自に発見を行う」というAGIレベル4の姿を現実のものとしつつあり、今後この路線の発展が科学技術の飛躍的進展につながることが期待されます。

参考文献・出典: 孫正義氏 SoftBank World 2024 講演 ^{40 2}；Gemini 3 Pro vs GPT-5.1 Pro 比較分析レポート ^{4 7 10 26}；その他本文中に示した出典箇所をご参照ください。

^{1 40} 『SoftBank World 2024 総まとめ』 超知性AIとビジネスの未来

<https://www.ai-box.biz/post/softbank-world-2024-ai-revolution>

² 難しいAI用語を優しく解説：孫 正義の特別講演を学ぶ | 0xpanda alpha lab

https://note.com/panda_lab/n/nae362f14076c

^{3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27 28 29 30 31 32 33 34 35 36 37 38 39} 発明の創出に最も力を発揮する生成AIの比較分析レポート Perplexity.pdf

file:///file_0000000089e47206b06c027800cfa251

²⁵ Sakana AI aims to automate scientific research with genAI

<https://www.constellationr.com/blog-news/insights/sakana-ai-aims-automate-scientific-research-genai>