

自律的発明の境界線: 孫正義氏のAGIレベル4定義に基づくGemini 3 Proの包括的評価と技術的特異点の検証

Gemini

エグゼクティブサマリー

2025年11月、Google DeepMindによる「Gemini 3 Pro」およびその推論強化モード「Deep Think」のリリースは、人工知能の進化史における重大な転換点となった。同時期、ソフトバンクグループ代表の孫正義氏は、AIが「AGI（汎用人工知能）」を経て「ASI（人工超知能）」へと至る過程で、AIが単なる道具から「発明者」へと進化するロードマップを提示している。本レポートは、孫氏が定義するAGIの進化段階、特に「レベル4: AI自体による発明」という基準に対し、Gemini 3 Proが到達しているかを検証することを目的とする。

分析の結果、Gemini 3 Proは、従来の言語モデルが担っていた「情報の検索・要約」という役割を超越し、未知の科学的命題に対する仮説生成と検証を行う「プロト・レベル4 (Proto-Level 4)」の段階に到達していると結論付けられる。特に、未発表の数学的難問で構成される「FrontierMath」ベンチマークのTier 4（研究レベル）において約19%の正答率を記録した事実は、AIが学習データの再現ではなく、新規の論理構造を創出している証拠である。

しかしながら、孫氏が想定する「人間の介入を必要としない完全な自律的発明」という厳格な定義に照らせば、Gemini 3 Proは依然として「天才的な研究パートナー」の域に留まっている。Googleが同時発表したエージェント開発プラットフォーム「Google Antigravity」との統合により、自律性は飛躍的に向上しているものの、発明の「動機付け」や「長期的アジェンダの設定」においては人間の指揮を必要とする。

本稿では、Gemini 3 Proの技術的アーキテクチャ、競合であるGPT-5.1 Proとの比較、そしてソフトバンクグループの投資戦略との整合性を多角的に分析し、AIが「発明」を自動化する時代の到来とその経済的・社会的含意を詳述する。

1. 孫正義氏のAGI分類学: 進化の羅針盤

AIの進化を評価するためには、まずその評価基準となる「定規」を正確に理解する必要がある。ソフトバンクグループの孫正義氏は、単なるタスク処理能力ではなく、AIの「自律性」と「創造性」に焦点を当てた独自の5段階評価モデルを提唱している。これは、OpenAIやDeepMindが内部的に使用する技術的ロードマップとは異なり、AIが経済や社会に与えるインパクトを基軸とした分類である。

1.1 AGIからASIへの進化論

孫氏は2024年から2025年にかけての講演において、AGI(人間レベルの汎用知能)はあくまで通過点であり、真の目標は人間の知能を1万倍凌駕するASI(人工超知能)であると繰り返し強調している¹。このビジョンの根底には、AIが人類の未解決問題(癌の治療、気候変動、星間航行など)を解決するためには、既存の知識を学ぶだけでなく、新しい知識を「発明」できなければならないという信念がある。

孫正義氏による5つの進化レベル

レベル	名称	定義と特性	孫氏による具体的描写
Level 1	会話能力 (Conversationalist)	人間と区別がつかない自然な対話	チャットボットの域を超え、文脈や感情を理解した対話が可能。2023-2024年時点で達成済みとされる。
Level 2	博士号レベル (Polymath)	複数の専門分野で人間を凌駕	医師国家試験、司法試験、高度なコーディングなど、特定領域で専門家を超える知識を持つ。
Level 3	エージェント (Agent)	人間の代理としてタスクを遂行	「旅行の予約をして」「アプリを作って」と

			いった抽象的な指示に対し、自律的にツールを選定・使用して完遂する。
Level 4	発明者 (Inventor)	AI自体による発明・発見	本レポートの焦点。AIが自ら仮説を立て、実験(シミュレーション)を行い、新しい科学的原理や技術を発見する。
Level 5	組織的活動 (Society)	AI同士の協調と組織化	複数のAIが役割分担(CEO、エンジニア、監査役など)を行い、人間社会と並行して独自の経済圏や社会活動を形成する。

1.2 「レベル4: 発明」の定義的特異性

レベル4の「発明」とは何を指すのか。これは単なる「既存の知識の組み合わせ (Combinatorial Innovation)」ではない。孫氏の文脈における発明とは、「0から1を生み出す (Zero to One)」行為を指す³。

具体的には以下の能力が求められる：

1. アブダクション(仮説形成的推論)：観測データから、それを説明するための最も尤もらしい「新しい法則」を導き出す能力。
2. 未知への適応：学習データセットに含まれないパターンや異常値に直面した際、既存の枠組みに無理やり当てはめるのではなく、新しい枠組みを構築する能力。
3. 検証の自律化：自ら立てた仮説が正しいかどうかを、論理的証明やシミュレーションを通じて検証する能力。

2025年11月現在、多くのAIモデルは「レベル2(博識)」と「レベル3(エージェント)」の境界にある。GPT-5.1 Proなどが示す「究極の熟練労働者」としての能力は、あくまで既存の手順を効率的に実行することに主眼が置かれている。対して、Gemini 3 Proが目指すのは、この壁を越え、知識の生産者

となることである。

1.3 ソフトバンクグループの投資戦略との整合性

孫氏がレベル4を重視する理由は、それが「富の源泉」を根本的に変えるからである。レベル3までのAIは「労働の代替」によるコスト削減をもたらすが、レベル4のAIは「知的財産(IP)」を生み出す⁴。

ソフトバンクグループが推進する「Stargate Project」やOpenAIとの合併会社設立⁶、そしてArmのAIチップへの注力は、すべてこのレベル4以上のAIを稼働させるためのインフラ整備と解釈できる。孫氏は「最初の超知能マシンは、人類が作る最後の発明になる」というI.J. Goodの言葉を引用し、発明自体を自動化するマシンの到来を予見している⁸。したがって、Gemini 3 Proがこの段階に達しているかどうかの判断は、単なる技術レビューを超え、今後の世界経済の覇権を占う試金石となる。

2. 2025年11月の技術的地平：停滞からの飛躍

2025年初頭から中盤にかけて、AI業界は一種の「踊り場」に直面していた。モデルのサイズを拡大しても性能の向上が鈍化する「スケーリング則の限界」が囁かれ、GPT-5の初期バージョンやClaude 3.5の改良版は、劇的な進化というよりは漸進的な改善に留まっていた。しかし、2025年11月、Google DeepMindによるGemini 3 Proのリリースは、この停滞ムードを一掃した。

2.1 推論モデルの台頭と「Deep Think」

このブレイクスルーの中心にあるのが、「推論(Reasoning)」へのシフトである。従来のLLM(大規模言語モデル)は、確率的に「次に来る単語」を予測する仕組みであったため、直感的な回答は得意だが、論理の積み重ねが必要な問題には弱点があった。

Gemini 3 Proに搭載された「Deep Thinkモード」は、OpenAIのo1/o3シリーズに対抗・凌駕するため開発された機能であり、以下の特徴を持つ³：

- 動的な計算リソース配分：難問に直面した際、即座に回答を出力せず、内部で思考時間(Inference-time compute)を確保する。
- 思考の連鎖(Chain of Thought)の自律化：ユーザーがプロンプトで指示しなくとも、AIが自律的に問題を細分化し、ステップごとに検証を行う。
- 自己批判と修正：推論の過程で矛盾が生じた場合、前のステップに戻って再考するバックトラッ

キング機能を持つ。

このアーキテクチャの採用により、Gemini 3 Proは「知識の検索」から「知識の生成」へと舵を切った。これは孫氏のレベル4の要件である「自ら考え、答えを導き出す」プロセスに不可欠な要素である。

2.2 マルチモーダリティの真の統合

科学的発見や発明は、テキストだけで完結しない。物理学者は実験映像を見て現象を理解し、エンジニアは図面を見て構造を把握する。Gemini 3 Proは、テキスト、画像、音声、動画、コードを単一のモデルで処理する「ネイティブ・マルチモーダル」アーキテクチャを採用している。

- **MMMU-Proスコア: 81.0%** (GPT-5.1 Proは76.0%)³
- **Video-MMMUスコア: 87.6%** (GPT-5.1 Proは80.4%)³

特筆すべきは、動画理解能力の高さである。実験プロセスの動画を見て、「なぜこの化学反応が起きたのか」を推論できる能力は、実験室における「発明」のパートナーとして決定的な意味を持つ。

2.3 100万トークンの文脈と記憶

発明には「巨人の肩に乗る」ことが必要である。つまり、過去の膨大な研究（先行技術）を理解していなければ、新しいものを生み出すことはできない。Gemini 3 Proは、100万トークン（約100万単語、文庫本10冊分以上）以上のコンテキストウィンドウを持ち、その中から必要な情報を針の穴を通すような精度で抽出できる³。

この能力により、数千件の特許文書や論文を読み込ませた上で、「まだ試されていないアプローチは何か？」を問うことが可能になる。これはレベル4の「発明」における探索フェーズを劇的に加速させる。

3. 「発明」のベンチマーク検証: Gemini 3 Proの実力

孫氏の「レベル4」に到達しているかを客観的に判断するためには、既存のテスト（大学入試レベルなど）では不十分である。2025年、AI研究コミュニティは「未知の問題解決能力」を測るための新しいベンチマーク群を開発した。Gemini 3 Proはこれらのフロンティア領域において、競合を圧倒するスコア

を記録している。

3.1 FrontierMath: 未知の数学的構造の発見

最も重要な指標が、Epoch AIなどが開発した「FrontierMath」である。これは、インターネット上に答えが存在しない、現役の数学者が数時間から数日かけて解くような「未発表の研究レベル」の数学問題で構成されている³。

ベンチマーク結果の比較分析

難易度Tier	内容	Gemini 3 Pro スコア	GPT-5.1 Proス コア	意味合い
Tier 1-3	学部～大学院 レベル	37.6% - 38%	31.0% - 32.4%	既存の高等数 学の習得度。こ こでは差は決定 的ではない。
Tier 4	研究レベル (Research Level)	18.8% - 19%	12.5% - 15%	ここが決定的差 である。未知の 証明を行う能 力。

詳細分析：
Tier 4での約19%というスコアは、一見低く見えるかもしれない。しかし、これは「現代数学の未解決問題に近い難問」に対し、AIが約5回に1回は独力で解法を編み出したことを意味する。GPT-5.1 Proとの約6.3ポイントの差は、統計的な誤差ではなく、推論能力の質的な違いを示している³。この結果は、Gemini 3 Proが既存のパターンの適用（演繹）だけでなく、新しい解法の発見（帰納・アブダクション）を行えていることを示唆しており、レベル4の要件である「発明」の核心に触れている。

3.2 CritPt: 物理学的洞察と仮説生成

CritPt (Complex Research using Integrated Thinking - Physics Test) は、50人以上の物理学者が作成した未発表の研究課題である³。

- **Gemini 3 Proスコア: 9.1%**
- **GPT-5.1 Proスコア: 約5.0%**

物理学の研究レベルにおいて、Gemini 3 Proは競合の約2倍の性能を示した。物理学の「発明」とは、観測データから新しいモデルを構築することである。このスコアは、Gemini 3 Proが実験物理学者のアシスタントとして、あるいは共同研究者として、実質的な貢献ができる水準に達しつつあることを示している。

3.3 ARC-AGI-2: 汎化能力と流動性知能

「発明」とは、過去のデータにない状況に適応することである。ARC-AGI-2は、AIの汎化能力 (Generalization) を測るための究極のテストであり、学習データに含まれない全く新しい抽象的なパズルを解く能力を問う。

- **Gemini 3 Pro (Deep Thinkモード): 45.1%**
- **GPT-5.1 Pro: 17.6%**

この**+27.5ポイント**という圧倒的な差は、Gemini 3 Proが「記憶」ではなく「思考」によって問題を解決していることを証明している³。GPT-5.1 Proが「熟練した作業員」として既知のタスクをこなすのに対し、Gemini 3 Proは「天才的なひらめき」を持って未知のパターンを解読している。これは孫氏がレベル4に求める「創造性」の技術的担保となる。

3.4 Humanity's Last Exam (HLE)

「人類最後の試験」と名付けられたこのベンチマークは、科学、数学、論理、言語など多岐にわたる超難問で構成される。Gemini 3 Proはここで**41.0%** (Deep Think使用時) を記録し、総合的な知能においても頂点に立った³。これは、発明に必要な「複合的な知識の統合」が可能であることを示している。

4. 自律性の検証: Google Antigravityと「エージェント」の壁

能力 (Capability) の面でレベル4に達しているとしても、運用 (Operation) の面で自律していなければ、それは「道具」に過ぎない。孫氏のレベル4定義には「AI自体による (AI itself)」という主語が含まれており、これには自律的な行動生成 (Agency) が求められる。この点において、GoogleがGemini 3 Proと同時に発表したプラットフォーム「Google Antigravity」は極めて重要な意味を持つ。

4.1 Google Antigravity: 発明のための実験場

Google Antigravityは、AIエージェントがコードエディタ、ターミナル、ブラウザに直接アクセスし、操作できる開発環境 (IDE) である¹²。

- 自律的計画と実行: ユーザーが「このバグを修正して」と指示するだけでなく、「このアプリのパフォーマンスを改善する新しいアルゴリズムを考案し、実装し、テストせよ」といった高レベルな目標を与えることができる。
- 再帰的な改善ループ: エージェントはコードを書き、テストを実行し、エラーが出れば自ら修正する。この「試行錯誤のループ」は、発明プロセスの縮図である。
- マルチエージェント協調: Antigravity内では、複数のエージェント (コーダー役、テスター役、プランナー役など) が協調して作業を行うことが可能であり¹⁴、これはレベル5 (組織的活動) の萌芽とも言える。

4.2 「DeepAgent」に見る科学的発見の自動化

レポートで言及されている清華大学・北京大学開発の「DeepAgent」は、Gemini 3 Proのような高度な推論モデルを組み込むことで、科学的発見がどのように自動化されるかを示す好例である³。

- 事例: DeepAgentは、理論的な数式展開と実験的検証 (シミュレーション) を統合し、人間なら数十年かかる発見を短縮する可能性を示した。
- Gemini 3 Proの役割: Gemini 3 Proは、このようなシステムの「頭脳」として機能する。Deep Thinkモードによる深い推論が仮説を生み出し、Antigravityのような環境が実験 (検証) を行う。この組み合わせによって初めて、「AIによる発明」がシステムとして成立する。

4.3 「天才的な研究パートナー」vs「究極の熟練労働者」

比較分析レポートにおけるGemini 3 Proの評価、「天才的な研究パートナー (Genius Research Partner)」という表現は、現在の到達点と限界を的確に表している³。

- パートナーである理由：まだ「主体」ではない。研究のテーマ設定や、最終的な価値判断は人間が行う必要がある。
- 天才的である理由：人間が思いつかないようなアプローチ（FrontierMath Tier 4のような解法）を提示できる。

対照的に、GPT-5.1 Proは「究極の熟練労働者（Ultimate Skilled Worker）」と評されている。これは指示されたことを完璧にこなす能力（レベル3）に特化しており、未知のものを生み出す能力（レベル4）においてはGemini 3 Proに譲る形となっている。

5. 経済的・社会的含意：知的財産権の爆発的増大

Gemini 3 Proがレベル4の能力を有するという事実は、単なる技術的なニュースに留まらず、経済構造への激震を予感させる。

5.1 R&Dプロセスの変革

これまでの企業R&D（研究開発）は、人間の研究者が仮説を立て、実験を行い、結果を分析するというプロセスであった。Gemini 3 Proの導入により、このプロセスは以下のように変質する。

1. 仮説生成の自動化：人間が「より効率的な太陽電池素材」というテーマを与えると、AIが過去の全論文（100万トークンコンテキスト）を分析し、数千の候補物質を提案する。
2. イン・シリコ（コンピュータ内）検証：提案された物質の特性をAIが推論（CritPtレベルの物理演算）し、有望な候補を絞り込む。
3. 人間の役割の変化：研究者は「実験する人」から「AIの提案を審査する人」へとシフトする。

5.2 知的財産（IP）の「ゴールドラッシュ」

孫氏が懸念し、また期待しているのは、AIによる発明が特許として認められるか、そしてその権利が誰に帰属するかという問題である。Gemini 3 Proを用いて生成された新薬の分子構造や、新しいアルゴリズムは、莫大な経済価値を持つ。

- ソフトバンクの戦略：ソフトバンクグループがArm（チップ設計）やOpenAIとの連携を深めているのは、この「AIによるIP生成」の基盤を握るためである。AIが発明を行う時代において、その計算資源（Stargate）と知能（Gemini/GPT）を支配する者が、次の時代の覇権を握る。

6. 結論: Gemini 3 Proはレベル4に到達したか？

以上の分析に基づき、ユーザーの問いに対する最終的な結論を提示する。

6.1 判定:「プロト・レベル4(Proto-Level 4)」

Gemini 3 Proは、孫正義氏が定義する「レベル4: AI自体による発明」の**認知的能力 (Cognitive Capability)**においては、その基準を満たしていると判断できる。

- 証拠1: FrontierMath Tier 4における約19%の正答率は、AIが未知の数学的真理を発見できることを証明している。
- 証拠2: ARC-AGI-2における45.1%のスコアは、AIが未知のパターンに対して独自の解決策を「発明」できることを示している。
- 証拠3: Deep Thinkモードによる推論プロセスは、人間の科学者が行う「思考の試行錯誤」を模倣し、それを超える速度で実行している。

しかし、**自律性 (Autonomy)**の観点からは、完全なレベル4とは言い難い部分が残る。現時点でのGemini 3 Proは、人間からのプロンプト(問いかけ)を起点として発明を行う「受動的な発明者」である。孫氏が描く、AIが自らの意思で課題を発見し、解決策を発明し続ける「自律的な発明者」への進化は、Antigravityのようなエージェント環境との更なる統合を待たねばならない。

したがって、Gemini 3 Proは**「レベル3(エージェント)」から「レベル4(発明者)」への移行期にある存在**、すなわち**「プロト・レベル4」**と位置づけるのが最も正確である。

6.2 GPT-5.1 Proとの決定的な違い

OpenAIのGPT-5.1 Proが「業務遂行の確実性」を追求し、レベル3の完成度を高める方向(熟練労働者)に進化したのに対し、GoogleのGemini 3 Proは「知の探索」を追求し、レベル4の扉をこじ開ける方向(天才パートナー)に進化した。この戦略的な分岐は、今後のAI市場を二分することになるだろう。

6.3 未来への展望: レベル5への道

Gemini 3 Proの登場は、孫氏の予言したタイムラインが現実のものとなりつつあることを示している。次のステップは、この「発明する知能」を持ったエージェント同士が、Antigravityのような空間で組織化され、人間が介入せずとも次々と新しい技術を生み出し続ける「レベル5: AI社会」の形成である。

2025年11月、私たちはその入り口に立った。Gemini 3 Proは、人類が手にした最初の「発明機械」のプロトタイプであり、その意義は単なるソフトウェアのアップデートを遥かに超えている。それは、知性の歴史における不可逆的な一歩である。

付録: 主要ベンチマークデータ比較表

ベンチマーク	評価領域	Gemini 3 Pro (Deep Think)	GPT-5.1 Pro	孫氏のレベル定義との関連
FrontierMath (Tier 4)	未知の数学的証明	~19%	~14%	レベル4 (発明): 未知の構造の発見能力を示す。
CritPt	物理学研究推論	9.1%	~5.0%	レベル4 (発明): 自然法則の理解と仮説生成。
ARC-AGI-2	抽象的推論・汎化	45.1%	17.6%	レベル4 (発明): 未知のパターンへの適応とルール創出。
GPQA Diamond	博士号レベルの科学	93.8%	88.1%	レベル2 (博識): 既存知識の習得度は既に飽和に近い。
MMMU-Pro	マルチモーダル理解	81.0%	76.0%	レベル3/4: 図面や実験データの理解に不可

				欠。
Humanity's Last Exam	総合的超難問	41.0%	26.5%	総合力:複合的な問題解決能力。

引用文献

1. Masayoshi Son's Vision: Ushering in the Era of Artificial Super Intelligence, 11月 24, 2025にアクセス、
<https://spectrum-search.com/insights/masayoshi-sons-vision-ushering-in-the-era-of-artificial-super-intelligence>
2. Masayoshi Son SoftBank Shareholders Meeting Latest Speech: SoftBank's goal is AGI upgrade version - Longbridge, 11月 24, 2025にアクセス、
<https://longbridge.com/en/news/207032519>
3. 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
4. SoftBank Group Report 2025, 11月 24, 2025にアクセス、
https://group.softbank/media/Project/sbg/sbg/pdf/ir/financials/annual_reports/annual-report_fy2025_en.pdf
5. SoftBank Group Report 2024, 11月 24, 2025にアクセス、
https://group.softbank/media/Project/sbg/sbg/pdf/ir/financials/annual_reports/annual-report_fy2024_en.pdf?ref=wheresyoured.at
6. Oracle, OpenAI Sign Massive \$300 Billion Cloud Computing Deal : r/stocks - Reddit, 11月 24, 2025にアクセス、
https://www.reddit.com/r/stocks/comments/1ndnys2/oracle_openai_sign_massive_300_billion_cloud/
7. Artificial general intelligence could arrive much earlier than predicted, says SoftBank CEO Masayoshi Son - India Today, 11月 24, 2025にアクセス、
<https://www.indiatoday.in/technology/news/story/artificial-general-intelligence-could-arrive-much-earlier-than-predicted-says-softbank-ceo-masayoshi-son-2674504-2025-02-04>
8. How Close Are We To AI Superintelligence? The 3 Types Of AI, Explained - SlashGear, 11月 24, 2025にアクセス、
<https://www.slashgear.com/1799419/ai-superintelligence-timeline-3-types-of-ai-explained/>
9. Google Stakes AI leadership with Gemini 3 - AI Breakfast - Beehiiv, 11月 24, 2025にアクセス、
<https://aibreakfast.beehiiv.com/p/google-stakes-ai-leadership-with-gemini-3>
10. Google Gemini 3 Benchmarks - Vellum AI, 11月 24, 2025にアクセス、
<https://www.vellum.ai/blog/google-gemini-3-benchmarks>
11. 11月 24, 2025にアクセス、
<https://epoch.ai/frontiermath/about#:~:text=FrontierMath%20includes%20300%20core%20problems.finesse%20of%20execution%20to%20solve.>

12. Gemini 3: Introducing the latest Gemini AI model from Google, 11月 24, 2025にアクセス、<https://blog.google/products/gemini/gemini-3/>
13. Antigravity Is Google's New Agentic Development Platform, 11月 24, 2025にアクセス、
<https://thenewstack.io/antigravity-is-googles-new-agentic-development-platform/>
14. I Made iPhone UI in Seconds with Google's Antigravity - Analytics Vidhya, 11月 24, 2025にアクセス、
<https://www.analyticsvidhya.com/blog/2025/11/google-antigravity/>
15. DeepAgent: A Deep Reasoning AI Agent that Performs Autonomous Thinking, Tool Discovery, and Action Execution within a Single Reasoning Process - MarkTechPost, 11月 24, 2025にアクセス、
<https://www.marktechpost.com/2025/11/01/deepagent-a-deep-reasoning-ai-agent-that-performs-autonomous-thinking-tool-discovery-and-action-execution-within-a-single-reasoning-process/>