

AlphaProof Nexus 調査報告

Google DeepMind の形式証明 AI — 確認できた事実、評価、影響

2026 年 10 月 11 日

Claude Opus 5.5

凡例：本文中の上付き番号は文末「参考文献」の番号に対応する。第 1～3 章は一次資料等で確認した事実の整理、第 4・5 章および「分析」「推測」と記した箇所は筆者による分析・推測を含む。確認できなかった情報はその旨を明記した。対象記事：Science Media Centre Spain「AlphaProof Nexus: A new AI tool designed to tackle mathematical proofs」。

要旨

- AlphaProof Nexus は、Google DeepMind が開発した数学の証明探索の仕組みである。LLM（Gemini 3.1 Pro）が証明を書き、形式証明言語 Lean が検証する。2026 年 5 月に arXiv で公開され^[1]、10 月 8 日に Science 誌に掲載された^[2]。
- 未解決の Erdős 問題 353 件中 9 件、OEIS 予想 492 件中 44 件を解決したと報告している^[1]。成果は本物だが、解かれた問題の多くは比較的無名で、規模は控えめである^[3]。
- **最も重要な知見は、最も単純な「LLM+コンパイラ」のループだけでも同じ 9 問をすべて解けた点である^[1]。**独立評価では汎用 LLM が同じ OEIS 予想で Nexus を上回っており^[4]、専用システムの優位は縮んでいる。
- 評価は分かれる。Science 誌の論考は「研究数学向け AI の到来」と肯定的である^[5]。一方、スペインの専門家からは過剰設計、コスト開示の不透明さ、「誰も理解していない正しい結果」が増えることへの懸念が示された^[6]。
- 影響（分析）：検証の焦点は「証明が正しいか」から「形式化された主張が元の問題に忠実か」へ移る。「強力な LLM+厳格な検証器」という構図は、ソフトウェア検証や AI for Science 全般に波及する。

1. AlphaProof Nexus の概要（事実）

1.1 基本情報

項目	内容
開発元	Google DeepMind。論文著者は 21 名で、筆頭は George Tsoukalas ら、最終著者は

項目	内容
	Swarat Chaudhuri ^[1]
論文	Tsoukalas ほか「Advancing mathematics research with AI-driven formal proof search」Science 394(6820):234–239 ^[2]
経緯	arXiv 版 v1 が 2026 年 5 月 21 日、v2 が 6 月 8 日 ^[1] 。Science 掲載は 10 月 8 日 ^[2] 。同号に Avigad 氏（カーネギーメロン大学）と Ballard 氏（サウスカロライナ大学）の論考「AI for research mathematics has arrived」が掲載され ^[5] 、AAAS がプレスリリースを配信した ^[7]
成果物の公開	Lean 証明と一部の自然言語証明を GitHub で公開（Apache-2.0／CC-BY 4.0） ^[8]
未確認事項	DeepMind 公式ブログ記事は今回確認できなかった

1.2 仕組み

入力は、定理の主張だけを書き、証明部分を未完（Lean の「sorry」）にした Lean ファイルである。エージェントが証明を書き、Lean コンパイラが毎回検証し、エラーを次の試行に反映する。論文は次の 4 つの構成を比較した^[1]。

構成	内容
(A) 基本エージェント	Gemini 3.1 Pro が推論・編集し、Lean のエラーを受けて修正を繰り返す単純なループ。サブエージェントを並列に実行する
(B) AlphaProof 併用	部分目標を、強化学習ベースの Lean 証明器 AlphaProof（2024 年 IMO 銀メダル水準 ^[9] 、2025 年 11 月に Nature 掲載 ^[10] ）に渡し、証明・反証させる
(C) 進化型	AlphaEvolve に着想を得た構成。証明スケッチをデータベースで共有し、Gemini 3.0 Flash が有望度を相対評価して次の探索対象を選ぶ
(D) フル機能版	(B) + (C)。実際の未解決問題の探索にはこれを使用した

1.3 主な成果（論文の主張）

- **Erdős 問題**：形式化済みの 353 件に適用し、9 件（約 2.5%）を解決した。うち#12 の(i)(ii)は 1970 年に提起され、56 年間未解決だった。解決後、チームの専門家が Lean で書かれた主張が元の予想に忠実かを確認した^[1]。
- **OEIS 予想**：492 件を Gemini で自動形式化し、44 件を証明した。数列の最初の数項を確認する「テスト補題」で、誤った形式化を防いだ^[1]。

- **研究現場での利用**：最適化（アンカー付き勾配降下上昇法で $O(1/t)$ の収束率を証明）、代数幾何（Zanella の予想の特定ケース）、Ben Green の問題リスト#57 の解決支援などを報告している^[1]。
- **コスト**：Erdős 問題について「1 問あたり数百ドル」と記載している^[1]。ただし問題ごとの正確な金額は示されていない^[11]。

1.4 論文自身が認める限界

- 解けたのは主に組合せ論・凸最適化・数論など、Lean の数学ライブラリ（Mathlib）が充実し、部分目標に分けやすい分野である。新しい理論を要する問題は射程外とされる。
- エージェントは問題の核心を未証明の補題に押し込む傾向があり、禁止しても防げなかった。「既知の文献結果」とした補題が、実は幻覚だった例もある。
- #125 と #741(i) では「密度」の解釈を後から修正した。誤った形式化の発見とも、人手による再定式化とも読める。
- 事後分析では、最も単純な(A)でも 9 問すべてを解けた（難問ではコストが増える）。著者らはこれを、専用の訓練済みシステムから、LLM の進歩に伴う単純なループへの移行の兆しとみている^[1]。

2. Science Media Centre Spain 記事の内容（事実）

SMC Spain は 2026 年 10 月 8 日（中央ヨーロッパ夏時間 20:00）、Science 論文の概要と、スペインの専門家 4 人のコメントを掲載した。4 人とも利益相反の申告依頼には回答していないと注記されている^[6]。

専門家（所属）	評価	要点
Pablo Haya Coll（マドリード自治大学 計算言語学研究室／IIC）	肯定＋留保	価値は個々の成果より、数学的発見を体系的に生む方法論にある。一方、Terence Tao の指摘を引き、多くの問題を解いて技術を示したい AI 企業と、理解を深めたい数学界とでは目的が一致しないと警告した
Josep Curto（カタルーニャ公開大学／IE Business School）	批判的	「方法論上の不整合と過剰設計」と評した。①単純な(A)が(D)と同じ 9 問を解いた、②コストは成功した実行分のみで、AlphaProof の推論費用も除外されており不透明、③Lean は書かれた命題の正しさは保証するが、元の問題を忠実に翻訳しているかは保証しない、と指摘した
Javier Aramayona（数学	懸念	研究評価、LLM へのデータ利用、若手育成に重大な問いを投げ

専門家（所属）	評価	要点
科学研究所 ICMAT 所長)		かけると指摘。「正しさの証明書はあるが、誰も十分に理解していない結果」が大量に生まれる逆説を懸念した
Teodoro Calonge（パリ ヤドリード大学 計算機科学科）	概ね好意的	質の高い論文で、初期の試みとして有用だが、多くの課題で失敗しており改善の余地は大きいと評価した。※同氏は手法を「RAG（検索拡張生成）」と説明しているが、論文の手法（Lean のフイードバックループ等）とは一致しないと判断する

3. その他の反響（事実）

3.1 肯定的評価

- **Science 誌の論考（Avigad 氏、Ballard 氏）**：プレスリリースによれば、失敗した証明の試みも問題の理解に役立ったという著者らの報告に注目した。そのうえで、数学向け AI の本質的な目標は、理解を深める探求において数学者を支援することだと強調した^[5,7]。
- **The Decoder（独テックメディア）**：OpenAI の自然言語アプローチより体系的で拡張しやすく、日常的な研究の道具を目指すものと評価した。一方、9/353 という成功率は「AI の Erdős 問題の成功率は 1～2%程度で、易しい問題に偏る」という Tao 氏の見立てとほぼ一致するとも指摘した^[11]。
- **「FrontierMath Erdős」論文（Epoch AI の Adamczewski 氏、erdosproblems.com 運営者の Bloom 氏）**：Nexus の主張のうち#152、#846、#125、#26 を解決と認め、#741 も解決扱いとした（密度の解釈が必要と注記）。一方、AI が解いた Erdős 問題の多くは無名で、既知の手法を素直に改変すれば解けたとも述べている^[3]。

3.2 懐疑的・批判的評価とその検証

- **Anatol Wegner 氏（Substack、2026 年 5 月 31 日、Gemini 3.1 との対話形式）**：#12、#26、#152、#846 は既存の文献で解決済み、#138 は部分的な結果にすぎず、証明は大学院の演習レベルだと批判した^[12]。
 - － 検証（分析）：Tao の AI 貢献 wiki^[13]では、#12・#152・#846 について引用された文献はいずれも「部分的」扱いで、#846 の文献は関連問題#847 の解決である。したがって「すでに解かれていた」という主張は大部分が当たらない。ただし、#26 で DeepMind が解いたのは弱い変種（論文自身が「Erdős 本人の提起ではない」と明記）、#138 は論文も変種と認める部分結果であり^[1]、この 2 点の批判は正当である。

- **Epoch AI の独立ベンチマーク「OEIS Open」**（Tom Adamczewski 氏）：同じ 492 件の予想で、最小限のツールを持つ Claude Opus 4.8 が 1 件 50 ドルの予算で 147 件（30%。再検証では 144 件・29%）を解いた。GPT-5.5 は 26%、Gemini 3.5 Flash は 22%で、いずれも Nexus の 44 件（9%）を上回る^[4]。

－ 同論文によれば、Nexus の著者は私信で、OEIS 予想 1 件あたりのコストを平均約 10 ドル、最難関の数件で最大約 50 ドルと見積もっている。ほぼ同じコスト条件での比較といえる。また、論文の 44 件に対し、公開リポジトリの OEIS 証明は 38 件だと指摘した。一方で、これらの予想の多くは数学的な重要性が不確かだとも述べている^[4]。

- **Hassabis 氏の発言**：DeepMind の Hassabis CEO が「まだ AGI ではない」と期待を抑えたとの報道があるが^[14]、本人の発言の一次ソースは確認できなかった。

報道状況：スペイン語圏では EFE 通信の配信記事が掲載された^[15]。Nature や NYT など英語圏の主要メディアによる Nexus 単独の報道は、今回の調査範囲では確認できなかった（存在しないという意味ではない）。

4. 影響の解説（分析）

4.1 数学研究への影響

- **検証の焦点が移る**：Lean が論理の正しさを機械的に保証するため、査読の重心は「形式化された主張は本当に元の問題か」に移る。#125・#741 での密度の修正や、OEIS の自動形式化での人手確認がその例である^[1,6]。人間の専門家は不要にはならず、役割が変わる。
- **「解く」から「問う・選ぶ・理解する」へ**：Mathlib の既存の道具で届く中程度の難度の未解決問題は、今後急速に片付けられる可能性が高い。数学者の価値は、問題設定、形式化、結果の解釈・理論化に移る。Aramayona 氏の懸念^[6]は現実的である。
- **失敗した試みも資産になる**：形式的なスケッチでは未解決の部分目標が明示されるため、専門家は残りの部分に集中できる。共同研究の進め方を変えうる点で、Science 誌の論考が最も評価した部分である^[5]。
- **規範づくりが進む**：2026 年 6 月に公表された「AI と数学に関するライデン宣言」は、AI 利用の開示、独立した検証可能性、正しさと帰属への人間の責任を求めている^[16]。Nature も社説でこれを支持し、他分野も続くべきだとした^[17]。今後は学術誌や助成機関の方針に反映されると見込まれる。

4.2 AI 研究・科学研究全般への影響

- 「専用システム」から「汎用 LLM+検証器」へ：本論文の最大の教訓は、強化学習の専用器や進化探索よりも、強力な LLM と厳格なフィードバック（コンパイラ）の組み合わせが効くという点にある^[1]。Epoch AI の結果もこれを裏付ける^[4]。コンパイラ、シミュレータ、実験など「検証可能な環境」を持つ分野から自動化が進むという見通しを強める。
- 形式検証・ソフトウェア検証への波及：「LLM が書き、検証器が判定する」構造は、ソフトウェアの正しさの証明に直結する。Amazon は 2026 年 7 月、Lean の開発組織 Lean FRO への長期的な資金支援（同組織史上最大の寄付）を発表し、AI エージェントの行動範囲を定めるポリシー言語の正しさの証明に Lean を使っていると説明した^[18]。数学での成功は、検証済みコードや AI エージェントの安全性保証への投資を後押しすると考えられる。

4.3 競合との比較・位置づけ

取り組み	時期	特徴	検証状況
AlphaProof Nexus (DeepMind)	2026 年 5 月公開、10 月 Science 掲載	形式化済みの多数の問題に一括適用。Lean で検証。成功率約 2.5%	査読済み・証明公開 ^[2,8]
OpenAI：離散幾何の予想の反証	2026 年 5 月 20 日	内部モデルが単位距離予想を反証。Gowers 氏は「AI 数学のマイルストーン」と評価	自社発表 ^[19]
OpenAI：数学・理論 CS の 10 の進展	2026 年 8 月 1 日	内部モデル「Astra」による未解決問題の進展 10 件	自社発表・査読前 ^[20]
OpenAI：Navier–Stokes 問題の解決主張	2026 年 9 月 8 日	ミレニアム懸賞問題の解決を主張	先取権をめぐる指摘あり ^[21] 。クレイ数学研究所は賞を授与しておらず評価継続中とされる（一次資料は未確認） ^[22]
OpenAI：数学文書の一括公開	2026 年 10 月 6 日	内部モデルの多数の成果を GitHub で公開 ^[23]	数学者団体 AHM が「700 超のファイルの一括公開は学術ではなく力の誇示」と批判 ^[24]
Harmonic「Aristotle」	2026 年 1 月	Erdős #728 を Lean 証明で解決	解説論文あり ^[25]

OpenAI は難しく名の知れた問題を狙い、Nexus は検証可能な成果を体系的かつ安価に、公開手続きを踏んで出す路線である。難易度の印象では OpenAI が先行するが、検証や帰属をめぐる論

争を抱えている。Nexus は査読付き論文、証明の公開、成功率の開示という点で透明性が高い。ただし、仕組みそのものの優位は、汎用 LLM の進歩によって急速に薄れつつある。

4.4 研究開発・知財実務への示唆

- **研究開発**：課題を形式仕様・テスト・シミュレーションなど機械で検証できる形に落とせる部門ほど、AI による探索の恩恵を受けやすい。投資先としては、複雑なエージェント設計よりも、①最新 LLM へのアクセス、②検証環境の整備、③形式化できる専門人材、のほうが費用対効果が高いと考えられる（Curto 氏の結論とも一致する^[6]）。
- **知財**：数学的発見そのものは一般に特許の対象外だが、アルゴリズム（例：最適化の新しい学習率スケジュール）や検証済みソフトウェアは実務上の価値を持ちうる。AI が主に生み出した成果では、①発明者・著作者の認定（多くの法域で人間の関与が必要）、②先行技術調査（本件でも「既に解かれていた」との批判が出たように^[12]、AI の成果の新規性確認は難しい）、③プロンプトや社内データの利用記録、が論点になる。

4.5 情報の受け手としての確認ポイント

「AI が未解決問題を解いた」という見出しに接したときは、次の 4 点を確認するとよい。①対象問題の重要度、②成功率（本件は約 2.5%）、③解いたのは変種か原問題か、④独立した検証の有無。

5. 未確定の論点と今後の見通し

5.1 未確定の論点

- **成果の件数**：OEIS は論文では 44 件、公開リポジトリでは 38 件である^[4]。差の理由は確認できていない。
- **コストの全体像**：「1 問あたり数百ドル」は成功した実行の推論費用のみで、353 問全体への適用費用や AlphaProof の推論費用は含まれない^[1,6]。総コストは不明である。
- **数学的な重要性**：解かれた問題の多くは比較的無名である^[3,4]。「数十年未解決」は「難問」と同じ意味ではない。
- **参照データの鮮度**：Tao の AI 貢献 wiki は 2026 年 6 月 30 日で更新を停止しており、Science 掲載後の状況は反映されていない。また同 wiki に「AlphaProof Nexus」の名称は見当たらない^[13]。

5.2 確認できなかった情報

- DeepMind 公式ブログ記事、Hassabis 氏の発言の一次ソース、Terence Tao の Nexus に関する直接の発言。
- Science 論文・論考の本文（アクセス制限のため、論文は arXiv v2 で代替。書誌事項は OpenAlex で確認）。論考が「Perspective」区分であることも一次資料では未確認。
- The Decoder 記事の著者・掲載日（ページ上に記載なし）。「FrontierMath Erdős」論文の Nexus に関する評価部分（検索結果経由の確認）。

5.3 今後の焦点（推測）

- (1) 汎用 LLM の能力向上により、形式化済みの中程度の難度の問題の「掃討」が加速する。
- (2) 自動形式化の信頼性（主張の忠実性）が次の主戦場になる。
- (3) OpenAI の Navier–Stokes 主張の検証結果が、分野全体の信頼性を左右する。
- (4) 学術誌・数学団体・助成機関による AI 利用規範の制度化が進む。

参考文献

-
- [1] Tsoukalas, G., Kovsharov, A., Shirobokov, S., et al. (21 名) . “Advancing Mathematics Research with AI-Driven Formal Proof Search.” arXiv:2605.22763, v1 2026 年 5 月 21 日／v2 2026 年 6 月 8 日.
<https://arxiv.org/abs/2605.22763>
 - [2] Tsoukalas, G., et al. “Advancing mathematics research with AI-driven formal proof search.” Science, 394(6820), 234–239, 2026 年 10 月 8 日. <https://doi.org/10.1126/science.aej2213>（本文はアクセス制限のため未閲覧。書誌事項は OpenAlex で確認。）
 - [3] Adamczewski, T., Bloom, T. F. “FrontierMath Erdős.” arXiv:2609.25050, 2026 年 9 月 6 日.
<https://arxiv.org/abs/2609.25050>（AlphaProof Nexus に関する記述は検索結果経由で確認。）
 - [4] Adamczewski, T. (Epoch AI). “OEIS Open: How many conjectures can language models turn into theorems?” arXiv:2608.11941, v2 2026 年 8 月 13 日. <https://arxiv.org/abs/2608.11941>
 - [5] Avigad, J. D., Ballard, M. “AI for research mathematics has arrived.” Science, 394(6820), 166–167, 2026 年 10 月 8 日. <https://doi.org/10.1126/science.ael6628>（本文未閲覧。書誌事項は OpenAlex で確認。）
 - [6] Science Media Centre España. “AlphaProof Nexus: A new AI tool designed to tackle mathematical proofs.” 2026 年 10 月 8 日. <https://sciencemediacentre.es/en/alphaproof-nexus-new-ai-tool-designed-tackle-mathematical-proofs>
 - [7] AAAS (EurekaAlert!). “Introducing AlphaProof Nexus: An AI tool for formal mathematical proof discovery.” 2026 年 10 月 8 日. <https://www.eurekaalert.org/news-releases/1146436>
 - [8] Google DeepMind. “AlphaProof Nexus Results” (GitHub リポジトリ) . 2026 年 10 月閲覧.
<https://github.com/google-deepmind/alphaproof-nexus-results>

- [9] AlphaProof and AlphaGeometry teams (Google DeepMind). “AI achieves silver-medal standard solving International Mathematical Olympiad problems.” 2024 年 7 月 25 日. <https://deepmind.google/blog/ai-solves-imo-problems-at-silver-medal-level/>
- [10] Ringer, T. “Mathematicians put AI model AlphaProof to the test.” Nature (News & Views), 2025 年 11 月 12 日オンライン公開. <https://www.nature.com/articles/d41586-025-03585-5>
- [11] The Decoder. “Google Deepmind's AlphaProof Nexus solves decades-old math problems for a few hundred dollars.” <https://the-decoder.com/google-deepminds-alphaproof-nexus-solves-decades-old-math-problems-for-a-few-hundred-dollars/> (著者・掲載日はページ上で確認できず。)
- [12] Wegner, A. “DeepMind's AlphaProof Nexus.” Substack, 2026 年 5 月 31 日. <https://aichats.substack.com/p/deepminds-alphaproof-nexus>
- [13] Tao, T. (管理) . “AI contributions to Erdős problems.” GitHub wiki (teorth/erdosproblems) , 2026 年 6 月 30 日で更新停止. <https://github.com/teorth/erdosproblems/wiki/AI-contributions-to-Erd%C5%91s-problems>
- [14] Kasanmascheff, M. “Google DeepMind's AlphaProof Nexus Solves Erdős Problems as AI Math Race Moves Beyond Benchmarks.” WinBuzzer, 2026 年 5 月 26 日. <https://winbuzzer.com/2026/05/26/google-deepmind-says-alphaproof-nexus-is-still-not-agi-xcxwbn/>
- [15] EFE. “AlphaProof Nexus, la nueva IA capaz de resolver problemas matemáticos.” La Jornada Maya, 2026 年 10 月 9 日. <https://www.lajornadamaya.mx/internacional/269336/alphaproof-nexus-la-nueva-ia-capaz-de-resolver-problemas-matematicos-inteligencia-artificial>
- [16] Alper, J., et al. (16 名) . “Leiden Declaration on Artificial Intelligence and Mathematics.” Zenodo, 2026 年 6 月 2 日. <https://doi.org/10.5281/zenodo.20302944>
- [17] Nature (社説) . “Mathematicians are developing rules for AI use — other fields should follow.” Nature, 654, 571, 2026 年 6 月 16 日. <https://doi.org/10.1038/d41586-026-01881-2>
- [18] Cook, B., Bice, S. “Amazon is investing in the Lean Focused Research Organization.” Amazon Science, 2026 年 7 月 26 日. <https://www.amazon.science/news/amazon-is-investing-in-the-lean-focused-research-organization>
- [19] OpenAI. “An OpenAI model has disproved a central conjecture in discrete geometry.” 2026 年 5 月 20 日. <https://openai.com/index/model-disproves-discrete-geometry-conjecture/>
- [20] OpenAI. “Ten advances in mathematics and theoretical computer science.” 2026 年 8 月 1 日. <https://openai.com/index/ten-advances-in-mathematics/>
- [21] IJR. “OpenAI Claims AI Solved Navier-Stokes Millennium Problem Amid Credit Allegations.” 2026 年 9 月 9 日更新. <https://ijr.com/discover/tech-2026-09-08-openai-claims-ai-solved-millennium-math-problem-dfca5bab>
- [22] Wikipedia. “Navier–Stokes existence and smoothness.” 2026 年 10 月閲覧. https://en.wikipedia.org/wiki/Navier%E2%80%93Stokes_existence_and_smoothness
- [23] OpenAI. “Sharing AI progress in mathematics.” 2026 年 10 月 6 日. <https://openai.com/index/sharing-ai-progress-in-mathematics/>
- [24] AHM (Association for Human Mathematics) Communications Working Group. “AHM Statement on

OpenAI's October 6 Release of Mathematical Documents.” Terence Tao のブログ「What's new」掲載,
2026 年 10 月 7 日. <https://terrytao.wordpress.com/2026/10/07/ahm-statement-on-openais-october-6-release-of-mathematical-documents/>

- [25] Sothanaphan, N. “Resolution of Erdős Problem #728: a writeup of Aristotle's Lean proof.”
arXiv:2601.07421, 2026 年 1 月 12 日. <https://arxiv.org/abs/2601.07421>