AI による科学研究の加速と数学論文急増の実態

現状の検証と今後の見通し 参考文献付き調査報告

GPT-6 Astra

調査基準日 2026 年 11 月 2 日

本報告は、動画「【数学論文が 1 年で 2 倍】AI が科学研究を加速させている」の提供文字起こしを起点に、元論文、arXiv の公式統計と投稿規則、関連する一次資料を照合したものである。動画の数字の意味、研究を実際に前進させた証拠、検証体制の課題を整理し、研究開発と知財実務への影響を検討する。[1]

AI が科学研究の一部を加速していることは、投稿統計だけでなく、専門家による数学的成果の検証や研究工程の実験からも裏付けられる。ただし、論文数の増加と、正しく重要な発見の増加は区別しなければならない。現状は、研究成果の候補を生み出す能力が拡大し、その正しさ・新しさ・意義を確かめる能力が追いつこうとしている局面と評価できる。[8][13]

本報告では、確認した事実、資料の著者による主張、本報告の解釈と予測を区別する。本文の角括弧内の番号は文末の参考文献に対応する。予測の時期は分析上の目安であり、確定した将来像ではない。

1 数学論文が約 2 倍という主張の検証

動画の主な根拠は、Carlo R. Contaldi が 2026 年 10 月 25 日に公開した分析論文である。公開ページには PNAS への投稿が記載されているが、これは採択や査読完了を意味しない。[2]

表 1 動画と文字起こしの主要数値の照合

主張	一次資料に基づく整理
2026 年 9 月の投稿数は 4,363 件	公式発表は全分野合計 40,363 件。提供文字起こしの数字には脱落がある。[3]
数学論文が 1 年で約 2 倍	CS 分類を併記しない数学論文の 9 月の前年同月比較。3,945 件から 8,203 件、約 2.08 倍である。[2]
年間の数学研究成果も 2 倍	この解釈はできない。同条件の 1〜9 月累計は前年同期比 43.6%増である。[2]
数学の単著論文が 70%増	1〜9 月累計の前年同期比較。9 月単月比較では約 2.67 倍で、比較期間が異なる。[2]
非 CS 分野が 17%増	過去の傾向から予測した投稿数を平均約 17%上回ったという意味。単純な前年比ではない。[2]

ここで CS はコンピューターサイエンスを意味する。数学の件数は元論文の集計値を照合したもので、今回 API から全件を再取得して解析を再現したわけではない。世界のすべての数学論文、査読済み論文、確立した数学的成果の総数を測った統計でもない。

arXiv はプレプリントを公開する基盤である。掲載前のモデレーションは、学術誌の査読と同じ手続ではない。投稿数は公表された研究原稿の量を把握するうえで有用だが、正しい定理、再現可能な実験結果、重要な発見の件数を直接表す指標ではない。[5]

2 10 月の投稿動向と投稿制限

arXiv 公式の月次 CSV を 2026 年 11 月 2 日に取得し、前年同月比を計算した。投稿の高水準は 10 月にも続いている。表 2 は全分野の合計であり、数学だけの値ではない。[4]

表 2 arXiv 全体の新規投稿数

対象月	2025 年	2026 年	前年同月比
9 月	26,646 件	40,363 件	＋51.5%
10 月	27,692 件	40,657 件	＋46.8%

出典は arXiv 月次統計の公式 CSV。増減率は取得値から計算した。10 月の件数は 9 月を約 0.7%上回る。ただし、この前月比だけで投稿制限が無効だったとは判断できない。制限がなければさらに増えていた可能性があり、季節性や投稿処理のタイミングも影響する。

arXiv は 2026 年 10 月 1 日から、投稿を行う人について暦月あたり新規投稿 2 件、同時に処理中の投稿 3 件という上限を適用している。共著者全員の枠が消費される仕組みではなく、掲載を断られた投稿も上限に数えられる。[3][5]

公式説明では、内容の薄い原稿、研究を細切れにした原稿、AI で書かれた大量の原稿が運営負担を増やしている。AI 利用研究を一律に禁止する措置ではなく、投稿処理能力を守るための暫定対応と位置づけられる。[3]

本報告では、投稿増加の原因と、審査負担の原因を分けて評価すべきだと考える。少数の投稿者が難しい対応を大量に発生させる一方、多数の研究者が少しずつ投稿を増やすこともあり得る。制度の効果を判断するには、総件数に加え、処理時間、異議申立て件数、低品質投稿の比率などを確認する必要がある。

3 AI と投稿増加の因果関係

元論文は増加時期と分野差を分析しているが、各論文で AI が実際にどう使われたか、成果が正しいかまでは調べていない。また、著者数が増えたことは、その人たちが新たに研究を始めた初心者だったことを意味しない。[2]

AI による時間短縮のほか、従来は投稿されなかった結果の公開、既存原稿の完成、投稿先の変化、論文の細分化なども件数を動かし得る。これらは本報告が考える代替説明であり、現状の集計だけで寄与を切り分けることはできない。

AI の生産性研究には、導入時点の定義に由来する統計上の問題もある。2026 年の PNAS 論文は、最初に AI 利用が検出された論文を導入時点とする方法が、実際の因果効果がなくても導入後の生産性上昇を示し得ると指摘した。AI 利用者の投稿数が多いという相関だけで、AI による増加率を断定することは適切でない。[6]

動画で紹介される、約 3 時間 45 分で初稿が完成し従来なら少なくとも 1 週間を要するという話は、著者自身による単一事例と推計である。工程短縮の可能性を示すが、研究全体の所要時間や一般的な生産性倍率を測る比較実験ではない。[1][2]

したがって、現時点での判断は、AI による加速を示す証拠はあるものの、投稿増加の全体を AI に帰属させることも、科学的発見の増加率を投稿数と同一視することもできない、というものである。

4 数学研究を前進させた証拠

4 1 AI 利用の開示に基づく調査

数学での AI 利用を調べた別の研究は、2026 年 3 月 1 日〜8 月 20 日の 32,944 件を調査し、AI の実質的な数学的貢献が記載された論文を 1,712 件と分類した。その割合は 3 月の 1.39%から、8 月 20 日までの期間には 14.09%に上昇している。[7]

この調査は著者の利用開示を基にしており、未開示の利用は把握できない。分類には LLM によるレビューも使われている。また、未解決問題を解決したと分類された事例は、原稿中の主張を整理した結果であり、すべての証明を独立に検証した結果ではない。利用の拡大と成果の確立は分けて解釈する必要がある。[7]

4 2 First Proof による専門家評価

2026 年 10 月 14 日公表の First Proof 第 3 回評価では、研究数学者が解決済みだが解答を未公開だった 10 問を用い、約 30 人の専門家が AI の解答を評価した。数学的正しさ、説明の明瞭さ、出典表示を別々に評価した点が重要である。[8]

表 3 First Proof 第 3 回評価の数学的正しさに関する結果

対象	結果
ProofCouncil 複数モデルを用いる研究支援システム	10 問中 9 問
GPT-6 Astra	10 問中 7 問
Kimi K3 を用いる参加システム	10 問中 0 問
全参加システムの解答を合わせた場合	全 10 問に少なくとも一つの正答

出典は First Proof 第 3 回報告書。正しさとは別に、説明や出典表示には課題が残った。10 問の特定ベンチマークであり、数学一般の正答率でも、すべての公開モデルと非公開モデルの優劣を示す結果でもない。[8]

この結果は研究レベルの問題への有用性を示す一方、正しい答えを得ることと、他の研究者が理解・検証・利用できる成果にすることの間に差があることを示している。評価にはモデルだけでなく、問題の分解、探索、検査を行う周辺システムと計算予算も含める必要がある。

4 3 人間が検証した新しい数学的成果

2026 年 5 月には、Alon、Gowers らの数学者が、AI の生成した Erdős の単位距離予想への反例について、人間が検証し整理した解説を公開した。単なる文章生成を超え、数学的成果に AI が寄与した具体例である。ただし、この事例から、AI が生成する数学的主張一般の正しさを推定することはできない。[9]

5 数学以外で進む研究工程の自動化

科学研究への影響は、研究工程や検証方法によって異なる。計算機上で候補の性能を評価できる課題と、物理的な実験を必要とする課題では、加速を制約する条件が異なる。

表 4 研究工程における AI 活用の実証例

領域	確認された進展	評価の範囲
アルゴリズム探索	AlphaEvolve が候補プログラムを生成し、自動評価を使って改良する。[10]	明確な評価関数を設計できる課題に強み。あらゆる研究への一般化は別問題。
実験装置の操作	X 線ナノプローブや材料研究用ロボットを LLM エージェントで操作する研究が報告された。[11]	人間の入力と監督を組み込んだ実証であり、完全自律研究とは異なる。
研究工程の自動化	The AI Scientist の生成原稿 3 本のうち 1 本が、ICLR 関連ワークショップの査読を通過した。[12]	限定的な実験。主要会議への採択や、広範な科学研究の自動化を意味しない。

これらは、開発元による技術発表、装置運用の研究論文、Nature 掲載の研究報告という異なる種類の資料である。開発元の成果報告は、第三者による独立検証と同じ証拠水準として扱わない。また、AI で作成された個別原稿の査読通過と、AI システムを評価した研究論文の学術誌掲載は別の事実である。[10][11][12]

6 検証能力と研究の多様性が制約になる

6 1 未検証の仮説と確認作業の増加

Google、Google DeepMind、MIT FutureTech などによる研究は、600 人超の科学者への調査で、約 41%が未検証の仮説の増加を報告したとする。AI で時間を節約できた回答者の約 46%は、節約時間

の 4 分の 1 超を AI 出力の確認に使っている。自己申告調査であり、因果効果の推定ではないが、現場の制約を示す証拠である。[13]

仮説の候補を 10 倍作れるようになっても、実験設備が処理できる量や、専門家が検証できる量が変わらなければ、確立した知識が 10 倍増えるとは限らない。研究開発の投資判断では、生成 AI の導入と、検証設備、人材、データ整備への投資を一体で考える必要がある。

6 2 形式検証と専門家による理解

数学では Lean などの証明支援系が検証の制約を緩和する可能性がある。ただし、形式検証で確認できるのは、指定された定義と仮定の下で、形式化された命題が導かれることである。その命題が本来の問いを正しく表しているか、重要な問題か、先行研究とどう異なるかは別途確認が必要である。Lean 公式文書も、証明の妥当性と命題の意味を区別している。[14]

OpenAI は 2026 年 10 月 26 日の発表で、数学・計算機科学・物理の成果について、AI の出力を研究素材として位置づけ、外部専門家による理解・整理・評価を進める方式を示した。発表された成果候補と、学術共同体で理解・検証された成果の間に工程が存在することを示す対応である。[15]

6 3 研究テーマの集中と評価の偏り

Nature 掲載の研究は、4,130 万件の自然科学論文を分析し、AI を用いる研究者の成果・被引用への優位と、科学全体で扱うテーマの集中が併存すると報告した。過去の機械学習などを含む分析であり、最新の生成 AI だけの効果として解釈することはできない。[16]

6,749 人の科学者による AI 生成仮説の評価でも、モデルが類似の発想に偏る傾向や、自動評価と専門家判断の一致が弱いことが報告されている。人間の評価者にも自分の考えに近い案を好む傾向があるため、最後に人が見るだけで十分とは言えない。[17]

評価には、正しさ、新規性、反証への耐性、説明の明瞭さ、再利用可能性、後続研究への貢献など複数の軸が必要である。研究者が AI の案を選ぶ際も、既存の主流に近い案だけを優先しない仕組みを検討する価値がある。

6 4 AI による評価支援の効果と限界

31,000 件超のプレプリントを対象とする無作為化実験では、AI のフィードバックを受け取った著者が原稿を改訂する確率は、基準となる改訂率に対して 12.55%高まった。これは改訂行動への効果であり、研究の正しさや価値が同率で向上したことを意味しない。[18]

AI による評価支援は利用できるが、単独で真偽や学術的価値を確定する仕組みとして扱わず、文献照合、再計算、実験、専門家による判断と組み合わせるのが妥当である。

7 今後の見通し

以下は、前節までの事実に基づく本報告の予測である。分野ごとに異なる速度で変化し、研究成果の候補が増える速度に検証能力が追いつこうとする状態が数年間続く、という見方を中心シナリオとする。表 5 の時期は目安であり、発生時期を実証した予測モデルに基づくものではない。

表 5 時期と領域別の変化の見通し

時期と領域	起こりやすい変化	実現を左右する条件
2026 年末〜2027 年 数学と計算研究	少人数で探索できる範囲が広がり、証明候補、反例、分析コードが増える。	正しさの検証、文献照合、説明の改善。
2027〜2029 年 研究機関と出版	投稿前チェック、再計算、引用確認が広がり、コード・データ・検証履歴の重要性が増す。	共通形式、独立した検証、運用費用。
2027〜2030 年 材料 化学 生物の一部	自動実験設備と AI を結びつけた探索が拡大する。	装置接続、測定の信頼性、実験速度、再現性。
中長期 研究者評価と育成	論文数に加えて、問題設定、検証、概念整理への貢献が重視される。	採用、昇進、研究費配分の制度変更。

楽観シナリオでは、形式検証、自動実験、再現性確認が整備され、候補の増加が確立した知識の増加につながる。検証への投資が不足する場合は、投稿量と査読負担が増える一方で、信頼できる知識を探す費用が上昇する可能性がある。

科学の民主化も条件付きである。文章作成やプログラミングへのアクセスが広がっても、高度なモデル、計算予算、実験設備、検証できる専門家へのアクセスは均等ではない。First Proof の報告も、高い成績を得るための計算資源と費用を課題として挙げる。参入障壁の低下と最前線の競争条件の平等化は、別々に評価する必要がある。[8]

問いを設定する力は重要になると考えられるが、計算・実験・証明を自分で理解する力の重要性が下がるとは考えにくい。AI の誤りを発見し、研究の意義を判断するためにも、その基礎が必要である。若手が実作業を通じて学ぶ機会をどう維持するかが、研究機関の課題になる。

8 研究開発と知財実務への示唆

以下は本報告から導く実務上の提案である。AI による候補生成の拡大に合わせ、検証状況、人的貢献、総費用を記録し、成果の評価方法を整えることが有効である。個別の特許要件や発明者認定について法的結論を示すものではない。

- 調査では検証状況を記録する。プレプリント、査読済み、独立再現済み、形式検証済みを区別し、重要な結論は根拠資料まで追跡する。
- 発明・技術候補では生成件数と実証件数を分ける。AI が多数の案を提示しても、実現可能性や効果が確認された案が同じ割合で増えるとは限らない。
- 人と AI の寄与を残す。誰が課題を設定し、どの案を選択・修正し、どの結果を検証したかを記録する。
- 研究力を複数指標で評価する。論文数や出願数に、再現性、独自データ、実装実績、後続研究への貢献を組み合わせる。

- AI 導入の効果に確認費用を含める。下書き時間だけでなく、誤りの修正、実験、レビューまで含めた総工数で評価する。

今後の科学的加速を判定するためには、投稿数に加え、独立に検証された成果数、重要な誤りの発生率、検証に要する時間、実験で確認された候補の割合、新しい研究領域の広がりを追う必要がある。研究機関や企業は、これらの成果をどれだけ少ない時間と費用で得られるようになったかを継続して評価することが望ましい。

調査方法と読み方

動画内容の確認には利用者から提供された文字起こしを用いた。数値と制度は元論文および公式資料と照合し、arXiv 全体の月次投稿数は公式 CSV から取得した。各研究の解析コードを実行した独立再現や、個々の数学的証明の再検証は実施していない。引用したプレプリント、学術誌掲載論文、公式方針、開発企業の発表は、それぞれの資料種別に応じて解釈する。参考文献の閲覧日はすべて 2026 年 11 月 2 日である。

参考文献

本文の番号順に掲載する。閲覧日はすべて 2026 年 11 月 2 日。プレプリントの公開や投稿先の記載は、査読完了や採択を意味しない。

[1] 動画「【数学論文が 1 年で 2 倍】AI が科学研究を加速させている」

YouTube。利用者提供の文字起こし transcript (56).txt を内容確認に使用。投稿者名と公開日は本調査では未確認。

https://www.youtube.com/watch?v=CmH_iwRN3jo

[2] Contaldi, Carlo R. Is the arXiv Apocalypse upon us? A break in preprint growth coincident with agentic AI.

arXiv:2610.34449v1, 2026 年 10 月 25 日。プレプリント。公開ページに PNAS 投稿との記載。主な参照箇所は II、IV、V および表 2・表 3。DOI: 10.48550/arXiv.2610.34449.

https://arxiv.org/html/2610.34449v1

[3] Boboris, Kat. Fair Moderation, Equitable Access, and AI: arXiv's Updated Rate Limit Policy.

arXiv 公式ブログ、2026 年 10 月 1 日。投稿制限、9 月の総投稿数、共著者と投稿者の扱いなどを参照。

https://blog.arxiv.org/2026/10/01/updated-rate-limit-policy/

[4] arXiv. Monthly Submissions.

公式月次統計。2026 年 11 月 2 日取得。2025 年と 2026 年の 9 月・10 月の値を使用し、増減率を計算。

https://arxiv.org/stats/monthly_submissions

CSV 直接取得先

https://arxiv.org/stats/get_monthly_submissions

[5] arXiv. Content Moderation.

公式投稿方針。モデレーションと査読の相違、生成 AI 利用、投稿上限などを参照。ページの公表日表示なし。

https://info.arxiv.org/help/moderation/index.html

[6] Renault, Thomas; Bergeaud, Antonin; Bosquet, Clément. Scientific production in the era of large language models: Outcome-triggered treatment timing and spurious event-study dynamics.

Proceedings of the National Academy of Sciences, 123(33), e2618638123, 2026 年。オンライン公開 2026 年 8 月 11 日。DOI: 10.1073/pnas.2618638123.

https://pmc.ncbi.nlm.nih.gov/articles/PMC13486544/

[7] Jin, Jiashun; Ke, Zheng Tracy; Sui, Bingcheng. The Gold Rush in AI4Math: Where Are We Now?

arXiv:2608.24961v1, 2026 年 8 月 25 日。プレプリント。表 1、未解決問題の分類、研究上の限界を参照。DOI: 10.48550/arXiv.2608.24961.

https://arxiv.org/html/2608.24961v1

[8] Abouzaid, Mohammed; Srivastava, Nikhil; Ward, Rachel; Williams, Lauren. First Proof Third Batch.

Hexagon, hexagon:2610.00244v1, 2026 年 10 月 14 日。研究数学の評価報告書。正しさ、説明、出典表示を別々に評価。DOI 未付与。

https://hexagonmath.org/2610.00244

[9] Alon, Noga; Bloom, Thomas F.; Gowers, W. T.; Litt, Daniel; Sawin, Will; Shankar, Arul; Tsimerman, Jacob; Wang, Victor; Wood, Melanie Matchett. Remarks on the disproof of the unit distance conjecture.

arXiv:2605.20695v1, 2026 年 5 月 20 日。人間が検証し整理した数学的解説。DOI: 10.48550/arXiv.2605.20695.

https://arxiv.org/abs/2605.20695

[10] AlphaEvolve team. AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms.
Google DeepMind 公式技術発表、2025 年 5 月 14 日。開発元による報告として参照。
https://deepmind.google/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/

[11] Vriza, Aikaterini; Prince, Michael H.; Zhou, Tao; Chan, Henry; Cherukara, Mathew J. Operating advanced scientific instruments with AI agents that learn on the job.
npj Computational Materials, 12, Article 160, 2026 年 3 月 6 日。DOI: 10.1038/s41524-026-02005-0.
https://www.nature.com/articles/s41524-026-02005-0

[12] Lu, Chris; Lu, Cong; Lange, Robert Tjarko; Yamada, Yutaro; Hu, Shengran; Foerster, Jakob; Ha, David; Clune, Jeff. Towards end-to-end automation of AI research.
Nature, 651, 914–919, 2026 年 3 月 25 日。The AI Scientist の評価研究。DOI: 10.1038/s41586-026-10265-5.
https://www.nature.com/articles/s41586-026-10265-5

[13] Codreanu, Mihai; Imas, Alex; Mateos-Garcia, Juan, et al. AI in Science: Early Insights.
arXiv:2609.28504v2。初版 2026 年 9 月 15 日、参照版 2026 年 9 月 25 日。プレプリント。主に 5.3 節の未検証仮説と確認負担を参照。DOI: 10.48550/arXiv.2609.28504.
https://arxiv.org/html/2609.28504v2

[14] Lean. The Lean Language Reference: Validating a Lean Proof.
Lean 公式技術文書。形式的証明の確認範囲、命題の意味、仮定と検証手続を参照。随時更新。
https://lean-lang.org/doc/reference/latest/ValidatingProofs/

[15] OpenAI. Working with academic communities to build scientific understanding.
公式研究発表、2026 年 10 月 26 日。AI 生成成果と外部専門家による理解・検証の進め方を参照。
https://openai.com/index/building-scientific-understanding-with-academic-communities/

[16] Hao, Qianyue; Xu, Fengli; Li, Yong; Evans, James. Artificial intelligence tools expand scientists' impact but contract science's focus.
Nature, 649, 1237–1243, 2026 年 1 月 14 日。DOI: 10.1038/s41586-025-09922-y.
https://www.nature.com/articles/s41586-025-09922-y

[17] Bao, Honglin; Wu, Siyang; Liu, Xiao; Li, Sida; Cao, Shiyun; Evans, James A. Contemporary AI lacks the imagination to diverge or negate in science.
arXiv:2606.08251v3。初版 2026 年 6 月 6 日、参照版 2026 年 8 月 10 日。プレプリント。DOI: 10.48550/arXiv.2606.08251.
https://arxiv.org/html/2606.08251v3

[18] Wang, Binglu; Liang, Weixin; Xue, Jiahui; Zhang, Yuhui; Cao, Hancheng; Wang, Dashun; Yin, Yian. Human-AI Collaboration in Science at Scale: A Global Large-scale Randomized Field Experiment.
arXiv:2605.24180v1, 2026 年 5 月 22 日。プレプリント。AI フィードバックによる原稿改訂行動への効果を参照。DOI: 10.48550/arXiv.2605.24180.
https://arxiv.org/abs/2605.24180