

Target Event: 2026-10-06 AI Mathematical Manuscript Release
Document Type: Independent Audit & Institutional Blueprint
Status: Critical Assessment required for Academic Stakeholders

科学的ディスラプションの解剖学

OpenAI 「722件の数学原稿」 公開がもたらす制度的ストレステストと監査報告

**READ-AND-SHARE
REPORT**

これは単なる「AIのブレイクスルー」ではなく、
学術界への不可逆的なストレステストである

未公開の社内 frontier model
(名称未指定・パラメータ不明)

約4,000の研究問題の投入と集約
(1結果あたりChatGPT Pro思考約3時間相当)

372 result families



722 manuscripts (GitHub / Apache-2.0公開)



OpenAIによるこの一括公開は、「AIが論文を書いた」という事実以上に、数学研究における
【発見・検証・著者性・プレプリント流通】の既存インフラを一度に問う巨大な実験である。

「成果物の公開」と「科学的再現性」の間に横たわる巨大なギャップ

公開されたもの (Artifact Openness: 高)

- ✓ ・ PDF原稿、LaTeXソース、BibTeX
- ✓ ・ 多数のLean形式化 (many, not all)
- ✓ ・ ビルド情報と引用設定
- ✓ ・ Apache-2.0ライセンス付与

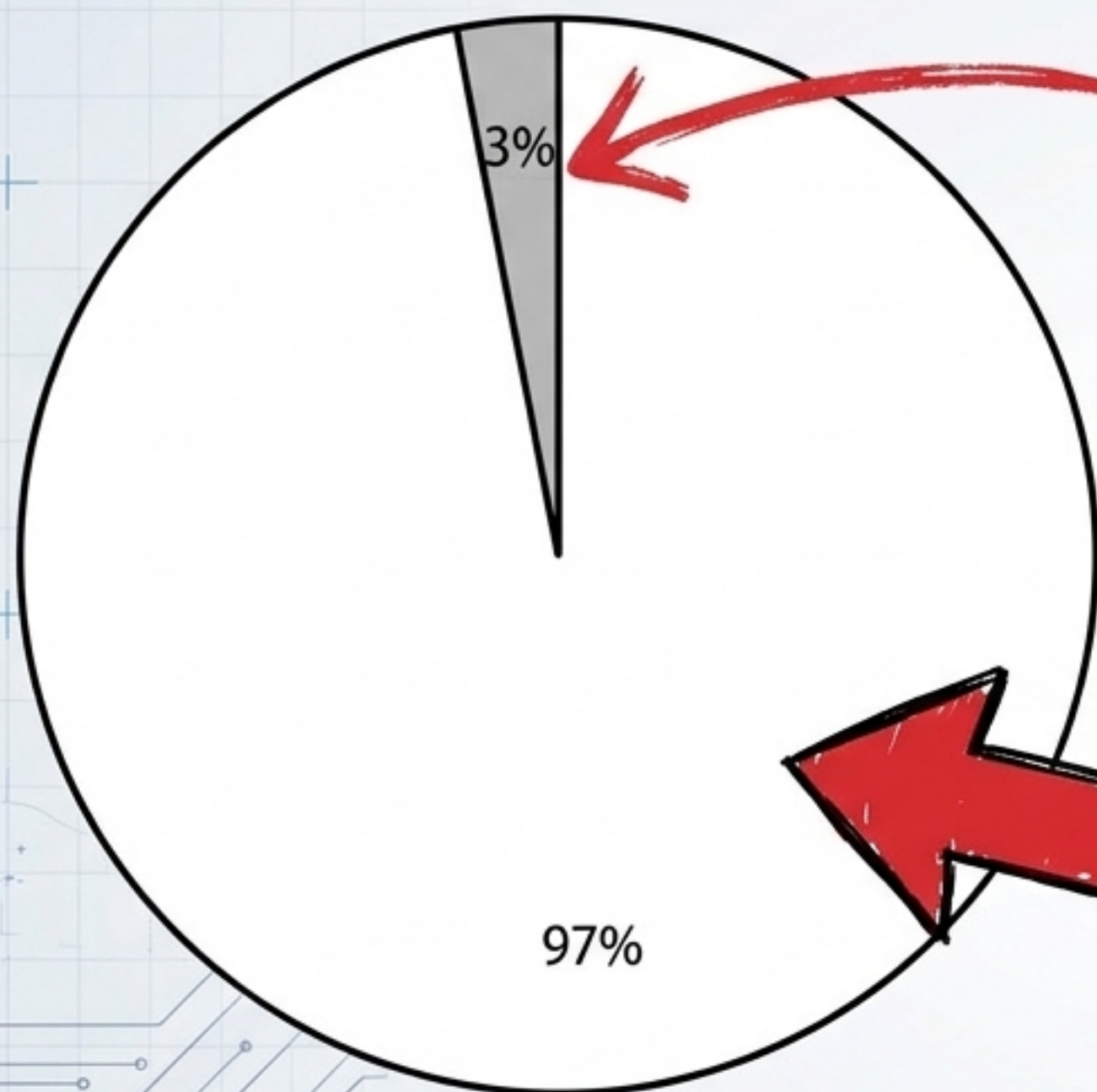
隠蔽されたもの (Scientific Reproducibility: 事実上不可能)

- ✗ ・ 正確なモデルのチェックポイントと重み
- ✗ ・ 原稿ごとの完全なプロンプトと乱数シード (Seeds/Sampling)
- ✗ ・ 全試行ログと、失敗した約3,600件前後の生データ
- ✗ ・ 生の出力テキストと人間による編集の差分 (diff)



アーティファクトの透明性は通常のデモより圧倒的に高いが、
発見プロセスの再実行は現状不可能である。

推論の透明性：監査可能なプロセスは全体のわずか 2.7%



10件の短縮された推論サマリーのみが公開 (Family 007, 017, 362など)。推論量が均一ではなく、完全なログではない。

■ reasoning summaryあり
READMEでsummaryが列挙されていないfamily

残りの362 families (約97%) については、AIがいかにしてその結論に至ったかの軌跡 (Reasoning trace) が完全に欠落している。



結論：コレクション全体の「発見過程」を数学コミュニティが独立監査するには、データが決定的に不足している。

Apache-2.0が隠蔽する学術的・法的な錯覚 (The Legal Illusion)

Apache-2.0が許可すること (The License)	学術・法務的現実 (The Reality)
[許可] 複製、改変、派生物の作成、商用利用。	[現実] 純粋なAI生成物にはそもそも 米国法上著作権が発生しない可能性 。Apache-2.0は存在しない著作権を創出しない。
[許可] 最低限のライセンス通知・NOTICE保持。	[現実] 学術的な先行研究への適切な クレジット（引用）義務は免除されない 。第三者の権利侵害リスクは残存。
[許可] 一定の特許ライセンス付与、無保証での提供。	[現実] OpenAIは定理の正しさを保証していない。コードや応用アルゴリズムでの 特許報復条項のリスク 。



「オープンソースライセンス＝法的にクリア」ではない。学術的剽窃の回避や独立した権利分析は依然として利用者の責任である。

モデル生成テキストのフォレンジック：AI検出器の無力さ

商用AI検出器 / Stylometry
(DetectGPT等)

パラフレーズやWatermark偽装で容易に回避可能。「AI生成」の証拠には不適切。

証明構造のフィンガープリント
(Proof-Structure Fingerprints)

補題分解の反復、異常に共通する記法、全体依存関係の飛躍など、数学特有の構造的痕跡。

真のプロヴェナンス要件
(Provenance Logs)

これ以上の推測を排除するためには、
Prompt hash、Raw output hash、
Tool-use log、Human edit diffが不可欠。

※今回はOpenAI自身が「AI生成」を宣言しているため、「AIが書いたか」を探る通常のAI検出技術は意味を持たない。真の問いは「どの部分が直接モデル由来か」である。

代表サンプルの監査ログ：均質化されたカタログの罠

COMMIT ID	DATE	MESSAGE	DETAILS
Family 007 (Chowla)	2023-10-26	主定理→拡張という定型構造。	所見： <u>全カタログに共通する「Proves...」「Resolves...」の文体は、AIの出力ではなく【OpenAIの編集・要約工程の文体】を検出している可能性が高い。</u>
Family 032 (CM abelian / Hodge)	2023-09-10	9/10~10/4のラベル。 	所見：[例外] 手動編集介入あり。対照群として分けないとStylometry評価に致命的な交絡が生じる。
Family 197 (Kaplansky)	2023-10-05	複数反例の統合。	所見：1つのFamily内に複数の結果が統合されており、「1生成=1結果」と見なせない。編集層の存在を示唆。
Family 362 (Vlasov-Maxwell)	2023-10-12	重大な大域的滑らかさの主張。	所見：結果が重大すぎるため、様式分析ではなく専門家による直接のPDEレビュー(Proof Audit)を最優先すべき。

再現性の4つのスペクトラム：「再現できる/できない」の二元論を破棄する

【高】 文書の再ビルド (Rebuild)

preprints/ 内のソース
ファイルとビルド手
順により、PDFの再構
築は容易。

【中～高】 形式証明 (Machine Verification)

公開済みのLeanライブ
ラリにより、特定の公
理・定義下での証明チ
ェッカー通過は確認可
能（ただし全原稿では
ない）。

【ばらつき大】 人間による追検証 (Human Audit)

未形式化の成果にはす
でに問題の可能性が示
唆されており、分野や
Familyごとに検証の難
易度と結果が激しく分
かれる。

【不可能】 発見プロセスの再実行 (AI Execution)

重み、プロンプト、シ
ード値、失敗ログが非
公開のため、AIがどうそ
の結論に至ったかの再
試行・問題選択率の再
現は完全に不可能。

最大のパラダイムシフト：発見から「検証」へのボトルネック移動

4,000問の並列処理 →
722本の原稿を短期間で一挙生成
(コスト：ChatGPT Pro 3時間/件)

論文ベースの伝統的な
査読システムの限界到達

専門家コミュニティの
「論文を1本ずつ読む」
帯域幅は容易に飽和する

AIによって「定理を発見・記述するコスト」が暴落した結果、科学コミュニティの最大の枯渇資源は「人間の専門家による検証・査読リソース (Verification Bandwidth)」へと移行した。

プレプリント文化の崩壊と、シグナルの希釈

モデルが何百本単位で論文を生成・公開できる時代、「公開された(Published)」という事実と、「読む価値がある/正しい(Verified)」というシグナルの結びつきが完全に消失する。

従来 (Past)

プレプリント



公共通域



プレプリント公開 = 新しい議論の迅速な共有 (専門家が読む価値の担保)

未来 (Future)

論文単位から「クレーム単位」のメタデータ管理へ

[AI-generated]

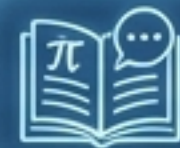
[Human-understood]

[Formally-verified]

[Expert-checked]

[Literature-checked]

生き残るための新たなアーキテクチャ：クレーム・レベルのメタデータ管理



Claim (主張)
自然言語による定理の抽出



Dependency Graph (依存関係)
引用・前提となる既知補題のネットワーク化



Machine-check (機械検証)
Lean等による形式証明ステータス



Human Verification (人間検証)
専門家による意味的等価性と正当性のレビュー



Novelty & Citation (新規性と引用)
先行研究の独立検索とクレジット付与

査読 (Peer Review) を単一のゲートキーパーとするのではなく、
この多層フィルタリング・アーキテクチャへ移行しなければ、数学界はAI出力の海に沈む。

ステークホルダー別 アクション・グリッド (1/2)

	直ちに実施すべき措置 (Short-term)	中期的制度 (Mid/Long-term)
研究者 (Researchers)	722原稿を「証明済み」として 安易に引用しない。 個別version/commitを指定。 先行文献の独立検索。	AI利用時のプロンプト、ツール 使用、人間の検証範囲の研究記化。 AI結果の形式化を独立した研究 成果としてクレジット。
出版社・リポジトリ (Publishers / Repositories)	‘AI-generated’、 ‘Formally-verified’ 等の ステータス・フィールドの 別管理導入。	大量AI投稿を直接査読に流さず、 「機械チェック → 専門家トリアー ジ → フル査読」の多段方式の 制度化。

ステークホルダー別 アクション・グリッド (2/2)

	直ちに実施すべき措置 (Short-term)	中期的制度 (Mid/Long-term)
AI研究所 (AI Labs - e.g., OpenAI)	各原稿に機械可読な Provenance manifest (モデル ID、プロンプト、生データハ ッシュ、失敗回数) を添付。	第三者監査可能なチェックポイン トの供託 (Escrow)。成功例だけ でなく問題選択のプロセスとネガ ティブ・リザルトの公開。
助成機関・政策立案者 (Funders & Policymakers)	Apache等のオープンライセン スと「研究の透明性」を同一視 しない方針の徹底。	独自の「検証インフラ (独立レビュー、 Lean形式化)」への資金提供。プロ プライエタリなAI企業から独立した コミュニティ・リポジトリの恒久支援。

最終監査ダッシュボード：2026年10月リリース評価

公開規模 (Scale)		[非常に大] - 722 manuscripts / 372 families
アーティファクトの透明性 (Artifact Openness)		[高] - PDF, Lean, Apache-2.0
個別の検証可能性 (Individual Verification)		[ばらつき大] - 形式化の有無で激しく分かれる
生成プロセスの透明性 (Generation Transparency)		[低～中] - ラン・レベルのプロヴェナンス不足
End-to-Endの再現性 (Reproducibility)		[低] - 重み・シード値の欠落
制度的リスク (Institutional Risk)		[高] - 検証ボトルネック、クレジット帰属、アクセス格差
科学的確定度 (Scientific Certainty)		[未確定] - コミュニティの評価待ち

今回の大量公開をもたらす最大のレガシーは、
AIが解いた数学の定理そのものではない。

数千の未検証クレームという圧力が、
機械可読な形式証明、バージョン管理されたクレーム・レジストリ、
そして『発見』と同等に『検証』を評価する
次世代の厳格な研究インフラの構築を、
数学界に強制することである。