

# Google DeepMind「Aletheia」未解決数学問題自律解決報道の検証と技術分析

## Executive Summary

Innovatopiaの報道は、Google DeepMindが数学研究エージェント「Aletheia」を提示し、Erdős問題データベースの“Open”課題群から複数の問題に有意な貢献をしたという点で、一次資料と整合する。ただし「自律解決」は“AIの出力→AI/人間の検証→人間の補正”を含む半自律プロセスで、解決した問題の多くは難易度より文献追跡不足（obscurity）が主因と明記されている。公開物はプロンプト・出力中心で、モデル本体や学習詳細は未公開で再現性は限定的。 <sup>1</sup>

## 調査範囲と方法

本報告は、(a) DeepMind公式ブログ（2026-02-11）、(b) Aletheia関連プレプリント（主に arXiv）、(c) Erdős Problemsサイト上の各問題定義とステータス、(d) 公開されたGitHubリポジトリ（プロンプト・出力・TeX/PDF）、(e) 主要メディア/準主要メディア報道（日本語・英語）の突合により、Innovatopia記事（2026-02-13）の主張を検証した。 <sup>2</sup>

一次資料内で数値・定義が揺れる箇所（例：v2/v3での「自律解決」数の表現）については、版の記述を優先しつつ、どの表現がどの分類（完全解決/部分解決）を指すかを明示した。 <sup>3</sup>

## 事実確認とエビデンス評価

### 公表の時系列と関係者

DeepMind公式ブログは「2026年2月11日」に、Gemini Deep Thinkを研究問題に適用した2本の論文公表を告知している（数学研究エージェントAletheiaを含む）。 <sup>4</sup>

Innovatopia記事は「2026年2月11日に2本の論文を公開」「Aletheiaで700の“未解決”問題を評価し4つを自律的に解決」と要約しており、主要な骨格（論文の存在、対象がBloomのErdős Problems、4という数）は一次資料（論文・公式ブログ）と一致する。 <sup>5</sup>

観点	一次資料で確認できる内容	Innovatopia記述	評価
公式発表日	2026-02-11に公式ブログ投稿（2本の論文へのリンク付き） <sup>4</sup>	2026-02-11に2論文公開 <sup>6</sup>	概ね一致
論文（数学側）	“Towards Autonomous Mathematics Research”（Aletheia、設計・評価・成果の整理） <sup>7</sup>	同タイトルを明記 <sup>6</sup>	一致
論文（ケーススタディ側）	“Accelerating Scientific Research with Gemini: Case Studies and Common Techniques”（151頁、事例集） <sup>8</sup>	同タイトルを明記 <sup>6</sup>	一致

観点	一次資料で確認できる内容	Innovatopia記述	評価
主要関係者 (論文A)	著者一覧に Demis Hassabis <sup>9</sup> 、Koray Kavukcuoglu <sup>10</sup> 、Quoc V. Le <sup>11</sup> 、Thang Luong <sup>12</sup> ら <sup>13</sup>	個人名の列挙は限定的 (主に組織名中心) <sup>6</sup>	省略はあるが 矛盾は小
外部コミュニティ要素	透明性の記述で Terence Tao <sup>14</sup> の助言に言及 <sup>15</sup>	リスク (無意識の盗作、openが難度でない等) を言及 <sup>6</sup>	方向性一致

## 「未解決数学問題」「自律解決」の実体

AletheiaのErdős問題への適用は、論文上「semi-autonomous (半自律)」として記述され、(i) AIベース自然言語verifierで候補を絞り込み、(ii) 人間の数学者が正当性と新規性を評価する“ハイブリッド評価”として定義されている。 <sup>16</sup>

「4つの自律解決」は、論文中で明確に「Erdős-652, 654, 1040, 1051」を指すと説明されるが、内訳は「完全に (first correct solution として) 解決」2件 (652,1051) + 「複数設問の一部を解決」2件 (654,1040) である。 <sup>17</sup>

さらに著者らは、当該4件は個別には論文級ではなく、既存文献の存在を踏まえると学生演習レベルのものもある、と“過度の興奮を戒める”注意書きを置いている。 <sup>18</sup>

従って、Innovatopiaの「未解決数学問題を自律解決」は“Erdős ProblemsのOpenラベルに対する解答をAIが生成した”という意味では支持される一方、「(数学界が長年行き詰まった) 超難問を単独で突破」あるいは「人間検証を不要にした完全自律」の含意で読むと、一次資料と齟齬が出る。 <sup>19</sup>

## 解かれた (または部分解決した) 具体的問題の定義と難易度所見

Erdős Problemsサイトの現行記述 (2026-02中旬時点) と論文記述を突合すると、少なくとも以下が「4件」のコアである (※各問題の“Open/Proved”表示はサイト管理者の現時点認識で、文献未探索の可能性を明示している)。 <sup>20</sup>

- Erdős-1051: 増大条件 (liminf) を満たす整数列に対し、ある無限級数が無理数か、という問い。サイト上は「PROVED (LEAN)」で、Aletheiaによる肯定的解答と、その一般化 (別論文) を記載。 <sup>21</sup>
- Erdős-652: 平面点集合の「各点からの距離の種類数」 $R(x_k)$ に関する漸近的下界 ( $\alpha_k \rightarrow \infty$ か) を問う。論文では「文献結果への即時還元」であり、出力内の引用不整合や指数の誤りがあったことも具体的に指摘。 <sup>22</sup>
- Erdős-654: 一般位置条件の強弱が複数版で問われる距離問題。サイトは「最強形の反証 (ただし2直線上の構成で一般位置版には未決着)」としてAletheiaの寄与を記載。 <sup>23</sup>
- Erdős-1040: 複素平面上の集合Fと多項式の劣値集合の測度 $\mu(F)$ が“transfinite diameter (log capacity)”で決まるか、等の問い。サイトは「 $\mu(F)$ はtransfinite diameterだけでは決まらない」反例をAletheiaが与えたとしつつ、全体としてはOPEN表示 (少なくとも副問が残存する可能性)。 <sup>24</sup>

## 公開された証明・コードの有無

結論から言うと、再現性に直結する「モデル本体」「エージェント実装コード」「学習データ・学習手順の詳細」は未公開である。一方で、プロンプトとモデル出力、ならびに一部成果のTeX/PDFがGitHub上に公開されている (Aletheiaディレクトリは“prompts and responses”中心で、少なくとも2件は“not from Aletheia”と注記)。 <sup>25</sup>

## 技術的分析

### アーキテクチャの骨格

論文は、Aletheiaを「Generator（解生成）・Verifier（自然言語検証）・Reviser（修正）」の3サブエージェントで構成し、Verifierが承認するまで反復する枠組みとして定義している。<sup>26</sup>

公式ブログは同様の“generate-verify-revise”ループを図示し、さらに「解けない場合に失敗を明示できる」点が研究協働の効率を上げると述べる。<sup>27</sup>

```
flowchart TD
    P[Research problem / conjecture] --> G[Generator: draft proof / strategy]
    G --> C[Candidate solution]
    C --> V[Verifier: NL critique & consistency checks]
    V -->|accepted| O[Final output + citations]
    V -->|minor fixes| R[Reviser: patch & refine]
    R --> C
    V -->|critically flawed| G
    V -->|timeout / attempt limit| F[Explicit failure / abstain]
```

### 推論・探索・検証プロセスの実装的含意

Aletheia論文は、検証を“中間思考トークンと最終出力の結合”から切り離し、別プロンプト（別呼び出し）で欠陥検出をさせると性能が上がる、という経験則を記述している。<sup>28</sup>

Erdős問題の大規模適用では、(a) 700件→AI verifierで212件を「正しそう」と判定、(b) 人間が素早くスクリーニングして27件へ、(c) ドメイン専門家が精査し「技術的に正しい」63件のうち「意図に整合する」13件を抽出、という多段フィルタになっている。ここで“正しい”が段階によって異なる（技術的正しさ≠問題意図への適合）点が、自然言語数学の評価難を示す。<sup>29</sup>

### 学習・データ・モデルサイズ・ツール利用

モデルサイズ（パラメータ数）や事前学習コーパスの詳細は未公開（不明）である。一方で、Aletheiaが「高度推論版Gemini Deep Think」「推論時計算を増やすスケーリング則」「Google Search/ブラウジング等ツール利用」に依存することは明記されている。<sup>30</sup>

また、別論文（事例集）は、Deep Thinkが「parallel thinking（複数分岐同時探索）」を取り入れ、追加の“novel reinforcement learning techniques”や「高品質解答コーパス」へのアクセス、さらに「長いインタラクティブ検証呼び出し列（人間検証で補強）」を組み合わせる、とより踏み込んで記述する。<sup>31</sup>

ツール利用の目的は特に「架空引用（spurious citations）の減少」で、単にネット接続するだけでは足りず、“tool useに向けた訓練”を施すと明白な架空引用は減るが、実在論文を誤引用するような“微妙な幻覚”が残る、と述べられる。<sup>32</sup>

### 既存手法との差分

形式証明（Lean等）中心の系では、形式化済みライブラリの範囲に制約される一方、Aletheiaは自然言語で長手順の探索を行い、検索・文献ナビゲーションを組み込む点に重心がある（ただし、その分“文献盗用の疑い”の検査コストも増える）。<sup>33</sup>

Erdősケーススタディでは、最も時間を要した工程が「既存文献に既解がないかの調査」とされ、AI活用が“数学の核心推論”のみならず“注意ボトルネック（膨大な探索・照合）”に効く一方、誤り検出・盗用検査の負担が現時点では無視できないことが示唆される。<sup>34</sup>

## 再現性と検証

### 公開物の範囲と限界

公開の中心は、GitHub <sup>35</sup> 上の「superhuman」リポジトリに置かれた、(i) Aletheiaのプロンプトと出力、(ii) 関連TeX/PDF (Erdős-1051一般化など) である。Aletheiaディレクトリ自体は“prompts and responses”が主で、実行コードが含まれるとは明記されていない。 <sup>25</sup>

従って、第三者が“同一モデル・同一エージェント”を再実行して再現することは現状困難（未公開/不明）である一方、(a) 出力の監査、(b) 証明の形式化（Lean等）による検証、(c) 同一プロンプトを他モデルで再挑戦し“再現しやすさ”を見る、といった限定的な検証は可能である。 <sup>36</sup>

### 独立再現のための要件整理

Erdősケースでは、AI verifier→人手評価というプロセスが前提で、独立研究者が追試するには「(1) 問題定義の正確な固定（意図の同定）」「(2) 文献探索」「(3) 証明の厳密化（必要なら形式化）」「(4) AI出力が文献からの抽出でないことの監査」が重要になる。 <sup>37</sup>

### 再現性チェックリスト

- ・対象問題の定義を一次資料（Erdős Problemsサイトの現行記述、原典引用）と突合し、表現揺れ・欠落条件の有無を確認する。 <sup>38</sup>
- ・Aletheiaの該当プロンプト/出力（GitHub公開分）を取得し、どの版の出力か（ablation出力を含む等）を明記する。 <sup>39</sup>
- ・証明を手で逐次検算し、論文が指摘する“軽微な誤り/定数ミス/誤引用”の修正履歴を残す。 <sup>40</sup>
- ・文献探索を独立に実施し、「既存解の存在」「近接結果の有無」「AIが参照した形跡（subconscious plagiarism疑い）」を評価する。 <sup>41</sup>
- ・可能なら形式化（Lean等）により、自然言語証明の論理ギャップを機械検証する（Erdős-1051はLean検証済み表示）。 <sup>42</sup>
- ・“Open”ステータスの意味を過信しない（サイト自体が文献未把握の可能性を免責として明示）。 <sup>43</sup>
- ・公表数値（例：95.07%/95.1%、条件付き精度、回答率<60%等）を、一次資料の定義（overall/conditional）で再掲する。 <sup>44</sup>
- ・研究成果の“自律性”と“数学的重要性”を切り分けて記述し、誇大な一般化を避ける（著者自身が重要性の過大視を警告）。 <sup>45</sup>

## 歴史的文脈と比較

Aletheiaの位置づけを明確にするには、(i) 形式証明（定理証明支援系）に寄せる系、(ii) プログラム探索＋評価器（外部評価で“正しさ”を担保）系、(iii) 自然言語での長手順推論＋反復検証（エージェント）系、の3潮流で見るのが有用である。 <sup>46</sup>

系統/代表例	アプローチ	検証の主軸	再現性（公開度）	Aletheiaとの差分
DeepMindの形式証明 （例：AlphaProof）	形式言語（Lean）で証明探索、RL的手法を併用	形式検証（proof assistant）	公式ブログで概要公表（詳細は別資料依存） <sup>47</sup>	Aletheiaは自然言語中心＋検索ツールで文献統合 <sup>48</sup>

系統/代表例	アプローチ	検証の主軸	再現性（公開度）	Aletheiaとの差分
DeepMindの幾何特化（AlphaGeometry）	ニューラル＋記号推論（幾何に特化）	記号推論・証明生成	Nature/公式ブログで詳細 <sup>49</sup>	Aletheiaは領域汎用で研究文献ナビが必須 <sup>50</sup>
LLM＋評価器（FunSearch）	コード空間探索＋自動評価器で“正しさ/スコア”を担保	自動評価（プログラム実行）	Nature/公式ブログ <sup>51</sup>	Aletheiaは自然言語証明が中心で評価が難しい <sup>52</sup>
オープン形式証明基盤（LeanDojo）	Retrieval＋LLMでpremise selection等を支援	形式検証（Lean環境）	オープンソース/ベンチ公開 <sup>53</sup>	Aletheiaはモデル/実装未公開で追試が難 <sup>54</sup>
近年のAI×Erdős（例：#728の形式証明書き起こし）	LLM生成→形式化→人間が記述を公開	形式検証＋人間監査	arXivで書き起こし公開 <sup>55</sup>	Aletheiaは自然言語研究エージェントとして体系化 <sup>56</sup>

## 影響評価と提言

### 学術的・社会的インパクト

数学研究への直接効果は「新しい大定理の自動発見」よりも、(a) 文献探索・照合、(b) 反例生成、(c) 証明の穴探し（adversarial reviewer）、(d) 形式化支援、のような“注意ボトルネック工程”で先に顕在化しやすい、という含意が一次資料に多い。Erdős事例では、最も大変だったのが文献同定と意図解釈だったと明言される。 <sup>57</sup>

教育面では、Aletheiaの「失敗を明示する」設計は、学習者が誤答を強化してしまうリスク（誤証を“それらしく”提示）を抑える方向に働かうが、これはモデル/設定に依存し一般化は慎重であるべき（不明）。 <sup>58</sup>

学術出版・責任の観点では、論文側が「論文の最終責任は人間が負うべき」という立場から、AIを共著者にしない方針（説明責任・引用の適切性を含む）を明示している点が重要である。 <sup>59</sup>

### 倫理的懸念とガバナンス論点

最重要リスクは、(1) 誤証・不完全証明の流通、(2) 文献由来の“無意識の盗作”疑い（subconscious plagiarism）と、その検査負担、(3) 「Open」ラベルの誤解による誇大報道・コミュニティ疲弊、である。これらは当事者論文が明示的に警告しており、Innovatopia記事もリスクに触れているが、見出しや一部表現は“自律”を強めに読ませる可能性がある。 <sup>60</sup>

他方で、事例集論文は、AIを「厳密な査読者」「自動検証＋コード実行のneuro-symbolic loop内の部品」として使う実務的パターンを整理しており、研究倫理上は「どの段階でAIをどう使ったか」の記録（透明性）が中心課題になる。 <sup>61</sup>

この点で、Aletheia論文が“自律性/新規性”の水準を codify し、透明性のための“human-AI interaction cards”概念を提案しているのは、誇大宣伝抑制とコミュニティ監査コスト低減に資する可能性が高い（提案段階）。 <sup>62</sup>

## 今後の研究課題と推奨

技術的に最優先の課題は「自然言語証明の信頼性（誤り率）を、検証コスト込みで下げること」である。Erdősスイープでは“技術的に正しい”63件のうち“意図に合致”が13件、さらに候補段階では誤りも多く、全体最適ではない。<sup>63</sup>

その上で、検証実験としては次が現実的である。第一に、公開されたプロンプト/出力から、第三者が同一の監査プロトコルで「(a)再現可能な正当性、(b)文献独立性、(c)形式化可能性」を採点し、誇大報道が起きにくい評価スキーマを作る。<sup>64</sup>

第二に、Erdős-1051のように形式化が進んだケースを起点に、自然言語→形式証明への変換を含む“二重検証パイプライン”を標準化する（形式化比率の制約は残るが、重要成果から優先的に実装する）。<sup>65</sup>

第三に、政策・出版側では、研究成果のAI関与を「モデルカード」的枠組みの発想で文書化する運用（Aletheia論文のinteraction cards提案とも整合）を、査読・プレプリント段階から要求することが望ましい。<sup>66</sup>

特許・商用化について、一次資料からは具体方針は読み取れず不明である（少なくとも公開資料には明示がない）。これに関しては、研究用途での透明性確保と、商用サービス化時の説明責任（誤証の責任分界、検証保証の範囲）の制度設計が先行課題となる。<sup>67</sup>

---

<sup>1</sup> <sup>2</sup> <sup>5</sup> <sup>6</sup> <sup>9</sup> <sup>10</sup> <sup>19</sup> Google DeepMind「Aletheia」が未解決数学問題を自律解決、AI研究者の時代が到来

<https://innovatopia.jp/tech-social/tech-social-news/80177/>

<sup>3</sup> <sup>12</sup> <sup>15</sup> <sup>16</sup> <sup>17</sup> <sup>18</sup> <sup>29</sup> <sup>33</sup> <sup>34</sup> <sup>36</sup> <sup>37</sup> <sup>39</sup> <sup>40</sup> <sup>41</sup> <sup>45</sup> <sup>52</sup> <sup>57</sup> <sup>60</sup> <sup>63</sup> Semi-Autonomous Mathematics  
Discovery with Gemini: A Case Study on the Erdős Problems

<https://arxiv.org/html/2601.22401v3>

<sup>4</sup> <sup>27</sup> Gemini Deep Think: Redefining the Future of Scientific Research — Google DeepMind  
<https://deepmind.google/blog/accelerating-mathematical-and-scientific-discovery-with-gemini-deep-think/>

<sup>7</sup> <sup>26</sup> <sup>28</sup> <sup>30</sup> <sup>32</sup> <sup>44</sup> <sup>48</sup> <sup>50</sup> <sup>56</sup> <sup>58</sup> Towards Autonomous Mathematics Research  
<https://arxiv.org/html/2602.10177v1>

<sup>8</sup> <sup>14</sup> <sup>31</sup> <sup>35</sup> <sup>61</sup> t.co  
<https://t.co/P7XJ8sMAOz>

<sup>11</sup> <sup>20</sup> <sup>24</sup> <sup>38</sup> <sup>43</sup> Erdős Problem #1040  
<https://www.erdosproblems.com/1040>

<sup>13</sup> Towards Autonomous Mathematics Research  
[https://arxiv.org/abs/2602.10177?utm\\_source=chatgpt.com](https://arxiv.org/abs/2602.10177?utm_source=chatgpt.com)

<sup>21</sup> <sup>42</sup> <sup>65</sup> Erdős Problem #1051  
[https://www.erdosproblems.com/1051?utm\\_source=chatgpt.com](https://www.erdosproblems.com/1051?utm_source=chatgpt.com)

<sup>22</sup> Erdős Problem #652  
[https://www.erdosproblems.com/652?utm\\_source=chatgpt.com](https://www.erdosproblems.com/652?utm_source=chatgpt.com)

<sup>23</sup> Erdős Problem #654  
<https://www.erdosproblems.com/654>

<sup>25</sup> <sup>54</sup> <sup>64</sup> raw.githubusercontent.com  
<https://raw.githubusercontent.com/google-deepmind/superhuman/main/aletheia/README.md>

<sup>46</sup> <sup>47</sup> AI achieves silver-medal standard solving International ...  
[https://deepmind.google/blog/ai-solves-imo-problems-at-silver-medal-level/?utm\\_source=chatgpt.com](https://deepmind.google/blog/ai-solves-imo-problems-at-silver-medal-level/?utm_source=chatgpt.com)

- 49 Solving olympiad geometry without human demonstrations  
[https://www.nature.com/articles/s41586-023-06747-5?utm\\_source=chatgpt.com](https://www.nature.com/articles/s41586-023-06747-5?utm_source=chatgpt.com)
- 51 Mathematical discoveries from program search with large ...  
[https://www.nature.com/articles/s41586-023-06924-6?utm\\_source=chatgpt.com](https://www.nature.com/articles/s41586-023-06924-6?utm_source=chatgpt.com)
- 53 LeanDojo: Theorem Proving with Retrieval-Augmented Language Models  
[https://arxiv.org/abs/2306.15626?utm\\_source=chatgpt.com](https://arxiv.org/abs/2306.15626?utm_source=chatgpt.com)
- 55 Resolution of Erdős Problem #728: a writeup of Aristotle's Lean proof  
[https://arxiv.org/abs/2601.07421?utm\\_source=chatgpt.com](https://arxiv.org/abs/2601.07421?utm_source=chatgpt.com)
- 59 67 Towards Autonomous Mathematics Research  
<https://arxiv.org/html/2602.10177>
- 62 66 Towards Autonomous Mathematics Research  
[https://arxiv.org/html/2602.10177?utm\\_source=chatgpt.com](https://arxiv.org/html/2602.10177?utm_source=chatgpt.com)