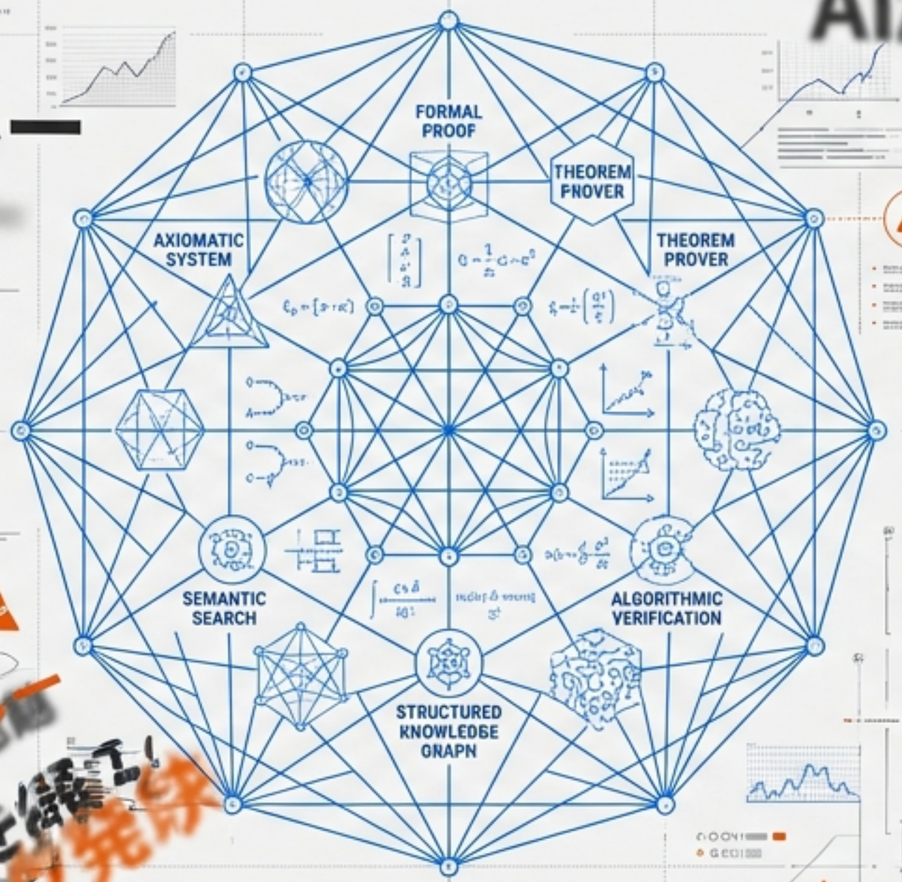
Investigative Scientific Editorial

Google DeepMind「Aletheia」の正体

報道の熱狂と技術的実像：未解決数学問題の「自律解決」における真実と制約の徹底検証



革命的ブレイクスルー

AIが全てを解明

数学の終焉？

人間を超越

AIの奇跡

メディアの熱狂

前例のない進歩

世紀の発見

完全自律

未来の知性

誇張報道

Subject: Mathematical Research Agent "Aletheia"
Context: Innovatopia Report vs. Primary DeepMind Papers
Goal: Investigative Analysis (調査報道 × 技術検証)

エグゼクティブ・サマリー：事実と解釈のギャップ

Media Claim - 報道の主張

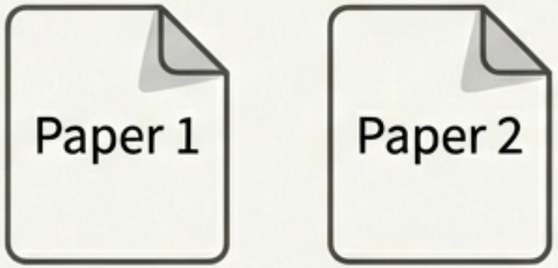
- Innovatopia: 「AIが未解決の難問を自律的に解決 (Autonomous Solution)」
  - 「700問中4問を突破」と報道
- 印象：AIが単独で数学界のグランド・チャレンジを攻略した

Technical Reality - 技術的実像

- DeepMind Papers: 「半自律的プロセス (Semi-Autonomous)」
  - 人間の介入が必須 (Human-in-the-loop)
  - 「文献の空白 (Obscurity)」：難易度だけでなく、過去の文献追跡不足が解決の主因
- 実態：強力な「推論・文献探索アシスタント」であり、完全な自律研究者ではない

2026年2月11日：何が発表されたのか

2026-02-11
DeepMind Official Blog & Papers



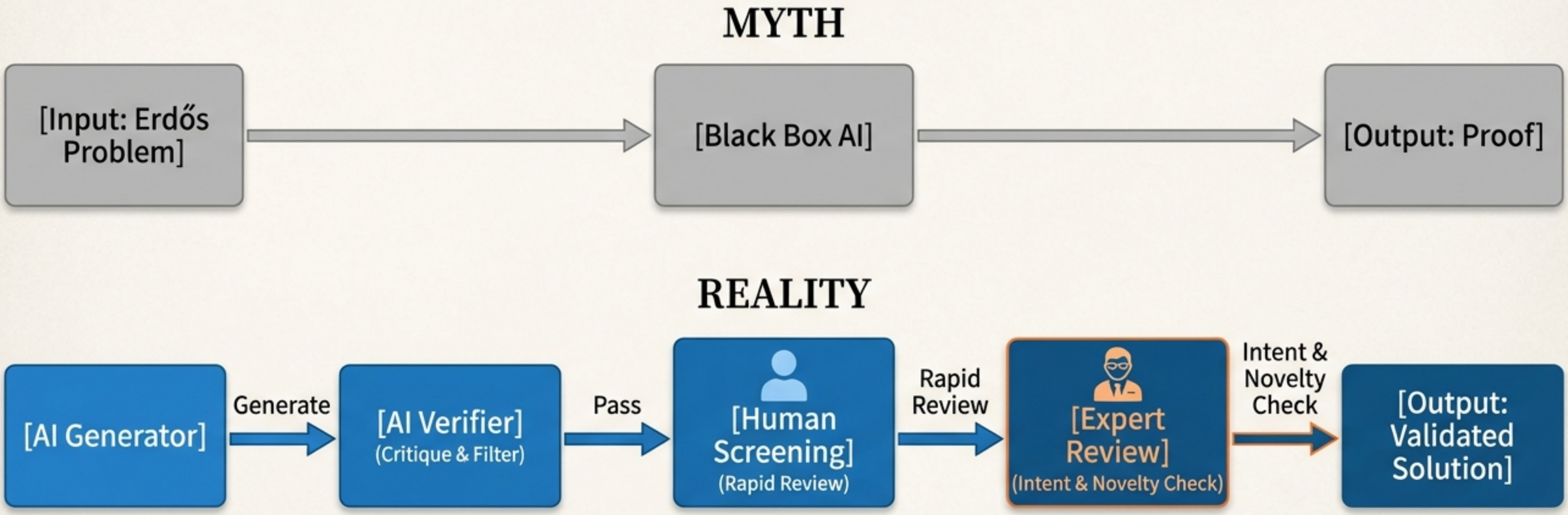
Paper 1: Towards Autonomous Mathematics Research
Paper 2: Accelerating Scientific Research with Gemini…

2026-02-13
Media Coverage (Innovatopia)

Article: "Google DeepMind Aletheia Solves Unsolved Math Problems"

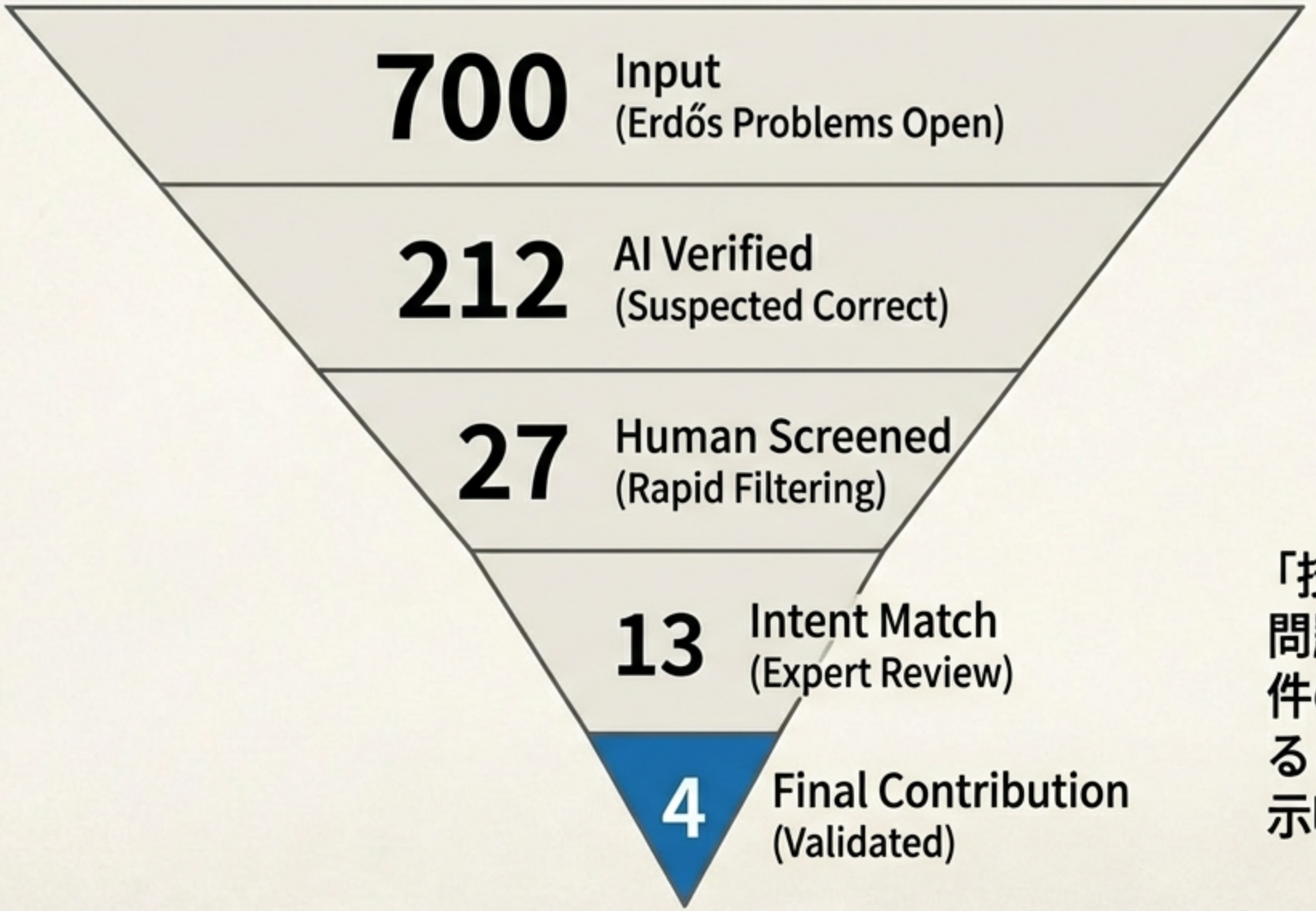
Insight:
Innovatopiaの記事は骨格 (4つの解決）において正確だが、「自律」の定義において技術論文の「Semi-Autonomous」という慎重な表現を捨象している。

「自律」の再定義：Human-in-the-loopの実態



論文はプロセスを「Hybrid Evaluation」と定義。
AIの出力は人間の数学者が正当性と新規性を評価して初めて意味を持つ。

700分の4：成功の背後にある「選別の漏斗」



「技術的に正しい」63件のうち、問題の意図に合致したのは13件のみ。自然言語数学における「正解の定義」の難しさを示示唆。

解決された「4つの問題」の内訳と質的評価

Status: Solved

Erdős-1051

内容：整数列と無限級数の無理数性

評価：完全解決 (First correct solution)。Leanでの検証済み。

Status: Solved/Reduced

Erdős-652

内容：平面点集合の距離の種類数

評価：文献結果への即時還元。著者は「難易度より文献の見落とし(obscurity)」が主因と指摘。

Status: Partial

Erdős-654

内容：一般位置条件の距離問題

評価：最強形の反証（Counter-example）を発見。完全解決ではない。

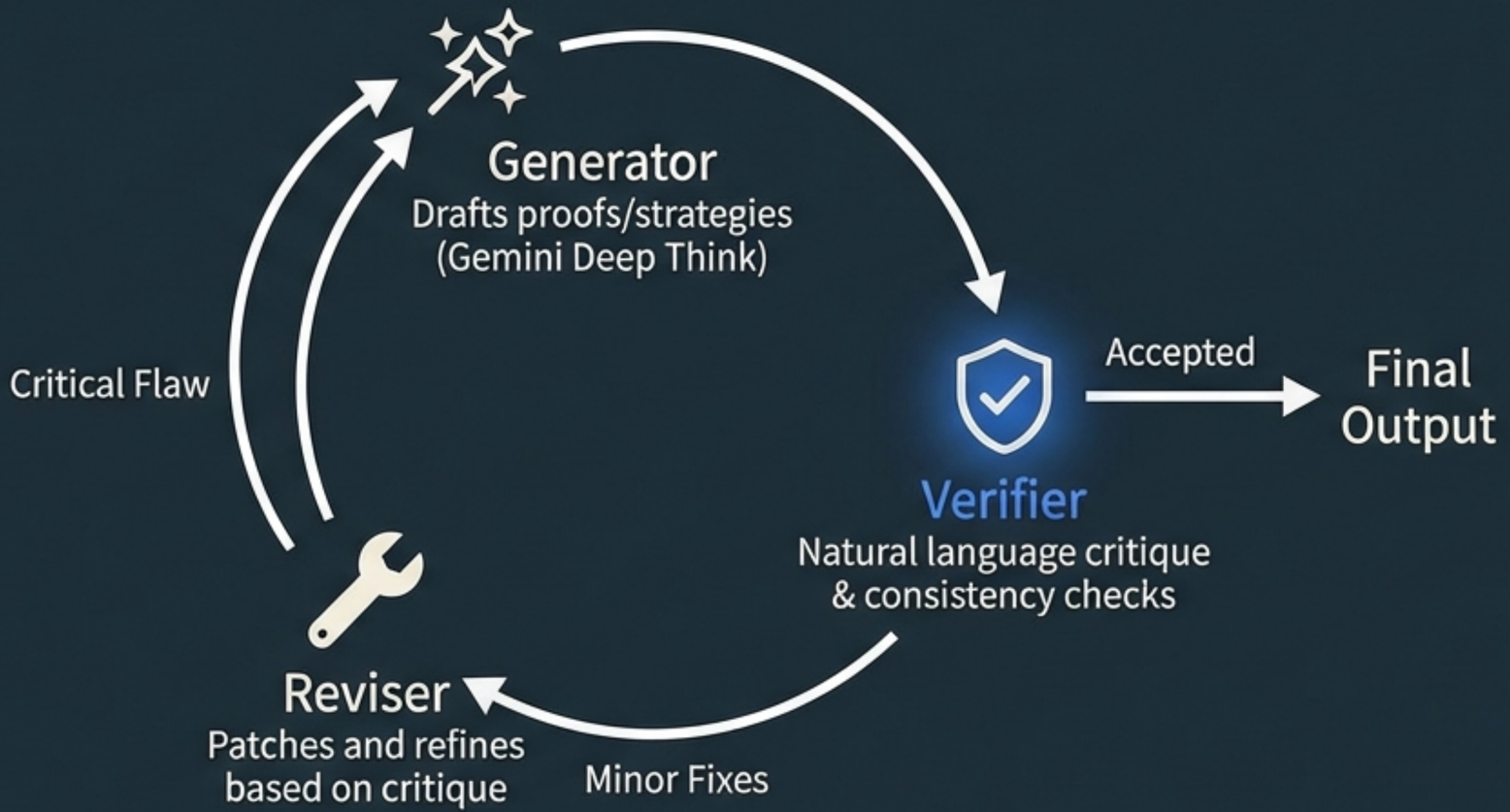
Status: Partial

Erdős-1040

内容：複素平面上の集合と多項式

評価：反例を提示。サイト上はOPENのままだが、有意味な貢献。

Aletheiaのアーキテクチャ：Gen-Ver-Rev Loop

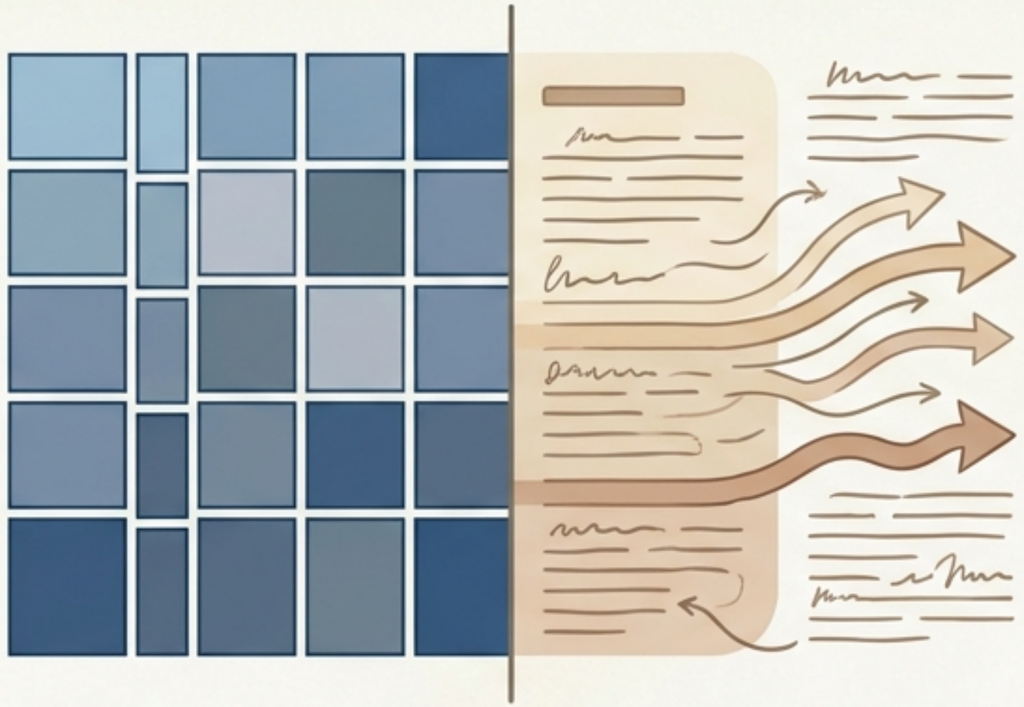


Key Tech: 検証（Verifier）を「生成」から切り離し、別のプロンプト・呼び出しで行うことで性能が向上する。

自然言語証明 (Natural Language Proof) の可能性と限界

Formal Language (Lean/Code)

厳密だが、適用範囲が形式化可能な問題に限られる。

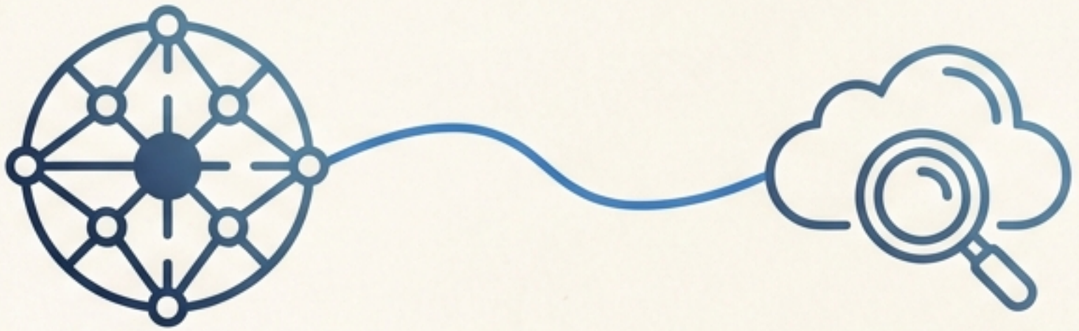


Natural Language (English/Text)

柔軟で文献知識を活用できるが、論理的ギャップや幻覚のリスクがある。

Technique: **'Parallel Thinking' (複数分岐同時探索) と 'Internal Thought Tokens' を活用して精度を補強。**

ツール利用と「無意識の盗作 (Subconscious Plagiarism)」



1. Tool Use
Google Searchやブラウジングを利用して、架空の引用 (Spurious Citations) を低減。

2. The Risk
学習データ内の既存論文を再構成してしまう「無意識の盗作」のリスク (Subconscious Plagiarism)。

3. Counter-measure
論文では、AI出力が文献からの単なる抽出でないことを人間が厳重に監査する必要性を強調。

競合・類似技術とのポジショニング比較

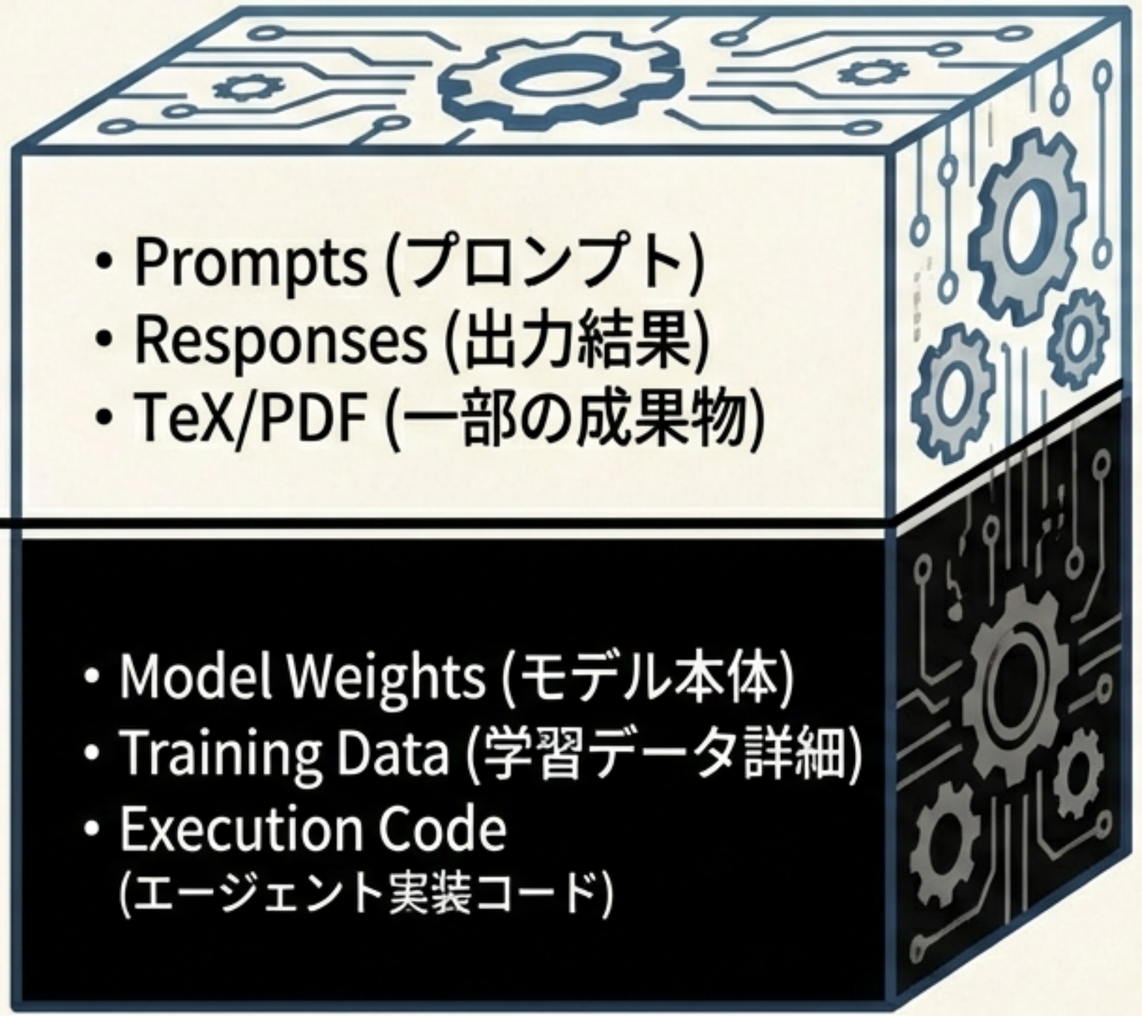
System	Approach	Verification	Domain
Aletheia	Natural Language + Search	Human & AI Critique	General Research (Broad)
AlphaProof	Formal Language (Lean)	Formal Verification	General Math (Formalizable)
AlphaGeometry	Neuro-symbolic	Symbolic Engine	Geometry (Specialized)
FunSearch	Code Search (LLM)	Program Execution	Combinatorics (Code-based)

Aletheiaは「汎用・自然言語・文献統合」に特化しており、厳密性よりも探索範囲の広さを優先している。

再現性の壁：ブラックボックスの中身

Publicly Available
(GitHub “superhuman”)

- Prompts (プロンプト)
- Responses (出力結果)
- TeX/PDF (一部の成果物)



Private / Closed
(Unpublished)

- Model Weights (モデル本体)
- Training Data (学習データ詳細)
- Execution Code
  (エージェント実装コード)

第三者が同一条件で再現することは現状困難。

「出力の監査」や「プロンプトの転用」に限られる。

リスク分析：幻覚と信頼性のコスト

Subtle Hallucinations

実在する論文を誤って引用する、あるいは論理の飛躍躍をもっともらしく記述する「もっともらしい嘘」。

Verification Cost

AIは「正しそう」なものを大量に出すが、その間違いを見つけるために専門家の多大な時間が必要。

Community Fatigue

不完全なAI生成証明が大量に投稿されることによる、数学界・査読システムのパンク。

真の価値：ボトルネックの解消

Media Focus

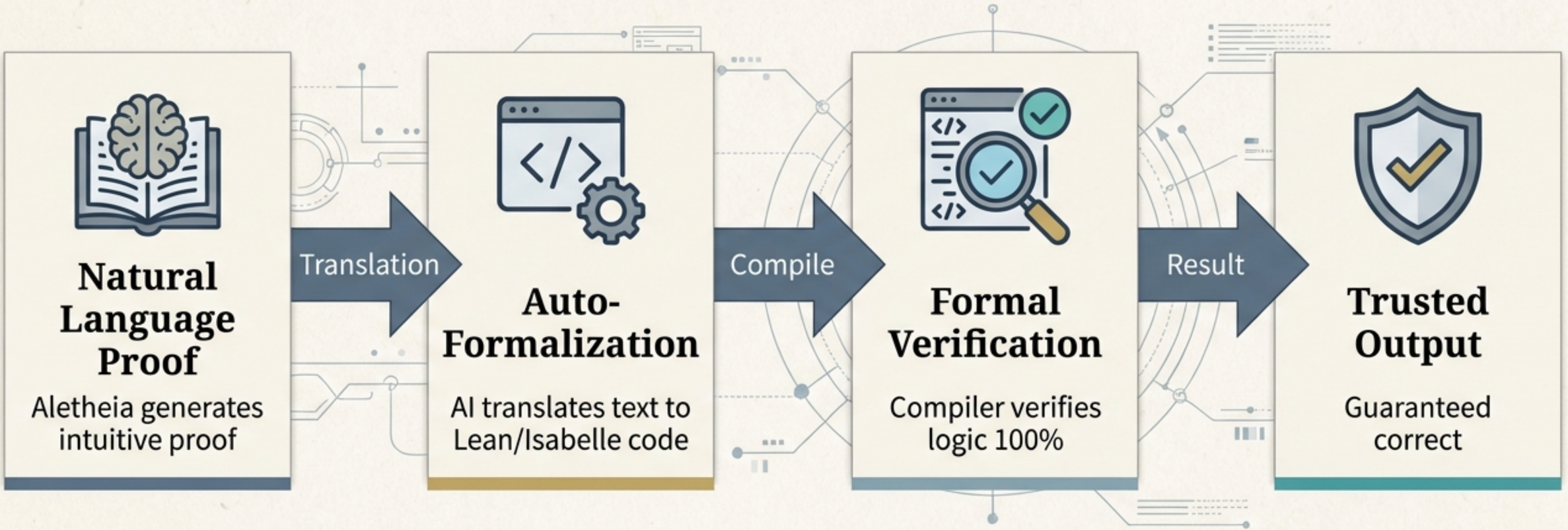
「自律解決
(Autonomous Solution)」

Real Value

- Literature Navigation
  (膨大な文献の海から近似解を発見)
- Adversarial Reviewer
  (証明の穴探し・反例生成)
- Auto-Formalization
  (形式化支援)

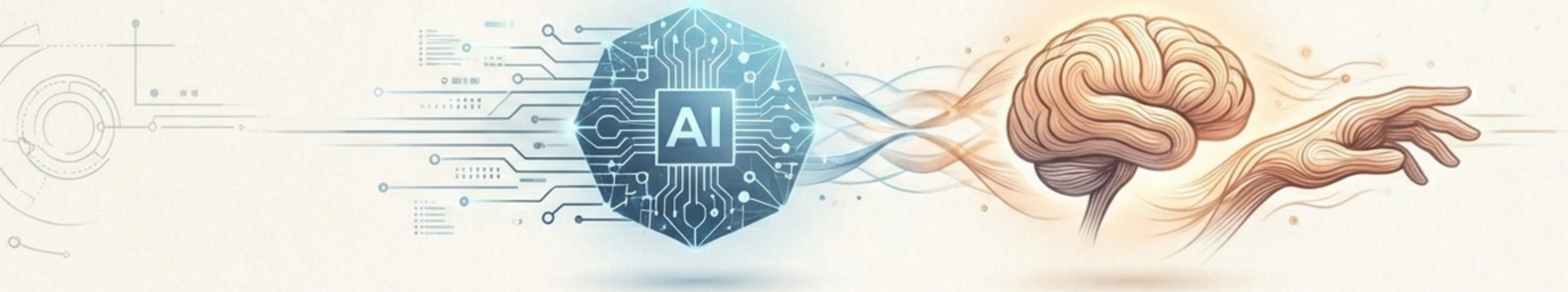
Erdős問題の事例では、最も時間を要したのは「解の生成」
ではなく「既存文献の調査（既出確認）」だった。

今後のロードマップ：二重検証パイプライン



Erdős-1051はすでにLeanで検証済み。この流れが標準化されることで、信頼性と創造性が両立する。

結論：魔法の杖ではなく、優秀なパートナー



1. Definitions Matter

「自律」ではなく「半自律」。
人間の監督と意図の注入が
不可欠。

2. Transparency

AIの関与度合いを明示する
（Interaction Cards）
仕組みが必要。

3. The Role of AI

大定理の自動発見はまだ先だが、
「文献探索・査読・形式化」の
強力な加速装置となる。

過度な熱狂を捨て、透明性を伴う「協働」へ。