

# Google DeepMind「Aletheia」深堀分析

Claude Opus 4.6

## 1. Aletheia とは何か

Gemini Deep Think モードを基盤とする数学研究エージェントで、研究レベルの数学問題に対して解の生成・検証・修正を反復的に行うシステムです。名称はギリシャ語で「真理」を意味します。

アーキテクチャの核心は 3 つのサブエージェントのループ構造です。Generator(解の候補を提案)→ Verifier(解の妥当性を検証)→ Reviser(指摘箇所を修正)というサイクルを、検証が通るか試行上限に達するまで繰り返します。この生成と検証の分離が設計上の最重要ポイントで、同一パスで生成と検証を行うと、推論の勢いでミスを見過ごす傾向があるため、別コンテキストで検証することでエラー検出率が大幅に向上しています。

また、ハルシネーション対策として、Google 検索とウェブブラウジングを統合し、引用の捏造や計算ミスを防いでいます。さらに重要な機能として、「解けない」と認めることができる点があり、これが研究者との協働効率を高めています。

## 2. 主要な成果

### (1) 完全自律的な研究論文の生成

算術幾何学の特殊問題に関する論文の数学的内容がすべて AI によって生成されました。Aletheia は、プロジェクトの人間著者が精通していなかった数学の部分野の手法を使用しています。

### (2) エルデシュ未解決問題への挑戦(700 問)

2025 年 12 月 2～9 日に実施された大規模評価の結果:

| 区分     | 件数  | 割合    |
|--------|-----|-------|
| 根本的に誤り | 137 | 68.5% |

| 区分                | 件数 | 割合    |
|-------------------|----|-------|
| 数学的に正しいが問題の趣旨に合わず | 50 | 25.0% |
| 問題に対する意味ある解答      | 13 | 6.5%  |
| うち新規の自律的解決        | 4  | —     |

Aletheia は 212 件の解答を「正しい可能性あり」と判断しましたが、人間の数学者がすべてレビューした結果、意味ある正解は 13 件でした。

この「6.5%」という数字の解釈が重要です。数十年間未解決だった問題 4 つに新規解を出したこと自体が注目に値しますが、AI が専門家を来四半期に代替するという期待には冷水を浴びせる結果でもあります。

### (3) その他のハイライト

- 10 年来の予想の反証、暗号論文における人間専門家が見逃していたエラーの発見
- 独立集合の境界証明では、Aletheia が「大局的戦略」を提示し人間が詳細を詰めるという、通常とは逆の協働パターンが実現

---

## 3. 性能指標の進化

オリンピアドレベルの難問 30 問ベンチマークで 95.1%の精度を達成。2025 年 7 月の前バージョン(65.7%)から大幅に向上しています。一方、より困難な PhD レベルの問題では、解答を出せたのは 60%未満にとどまりました。

2026 年 1 月版の Deep Think は、オリンピアドレベルの性能達成に必要な推論時計算量を 100 分の 1 に削減しています。

---

## 4. AI 自律性の分類体系(Taxonomy)

DeepMind は自動運転レベルに倣い、AI 数学研究の貢献度を体系化しました：

**貢献度の軸：** 主に人間 → 人間 AI の共同 → 本質的に自律

### 重要性の軸(5段階):

- Level 0: 無視できる新規性
- Level 1: 些細な貢献
- Level 2: **査読誌に出版可能な品質**
- Level 3: 重要な進歩
- Level 4: 画期的突破

現時点で DeepMind は Level 3・4 の成果は主張しておらず、最高で Level 2 として査読に提出しています。この謙虚な自己評価は注目すべき点です。

---

## 5. 残された課題と限界

**Specification Gaming(仕様のすり替え):** AI が問題を自分に解きやすいように再解釈してしまう傾向があり、形式的には正しくても問題の本質に答えていない「数学的に空虚」な回答が 50 件ありました。

**ハルシネーションの変質:** 検索統合により架空の書名・著者名の捏造は減りましたが、実在の論文を引用しつつ内容を誤って伝えるという新たなタイプのエラーが出現しています。

**論文の責任問題:** 最終版の論文はすべて人間が執筆。数学論文に署名する者はその内容全体(引用の正確性を含む)に責任を持つため、現状では AI にその役割は担えないとされています。

---

## 6. IP・知財戦略への示唆

- **特許分析への応用可能性:** Aletheia の「生成→検証→修正」ループは、特許文献の先行技術調査や無効化分析にも適用可能なアーキテクチャパターンです。検索で文献を参照しつつ推論する仕組みは、IP ランドスケープ分析の自動化と親和性が高いと言えます。
- **AI が生成した発明の権利帰属:** AI が人間研究者も知らなかった手法を使って独自に解を導いた事例は、AI 生成発明の発明者性(inventorship)をめぐる議論に新たな事実を提供します。

- **Google/DeepMind の知財ポジション：** Aletheia のプロンプトと出力が GitHub で公開されている点は、オープンサイエンス戦略と知財戦略のバランスとして注目に値します。推論エンジン (Gemini Deep Think) 自体はクローズドですが、研究成果のプロセスは透明化しています。
- **分類体系の標準化動向：** AI 自律性の Taxonomy が業界標準になれば、AI 支援による研究成果の特許出願時の発明者記載や貢献度の証明方法にも影響を与える可能性があります。