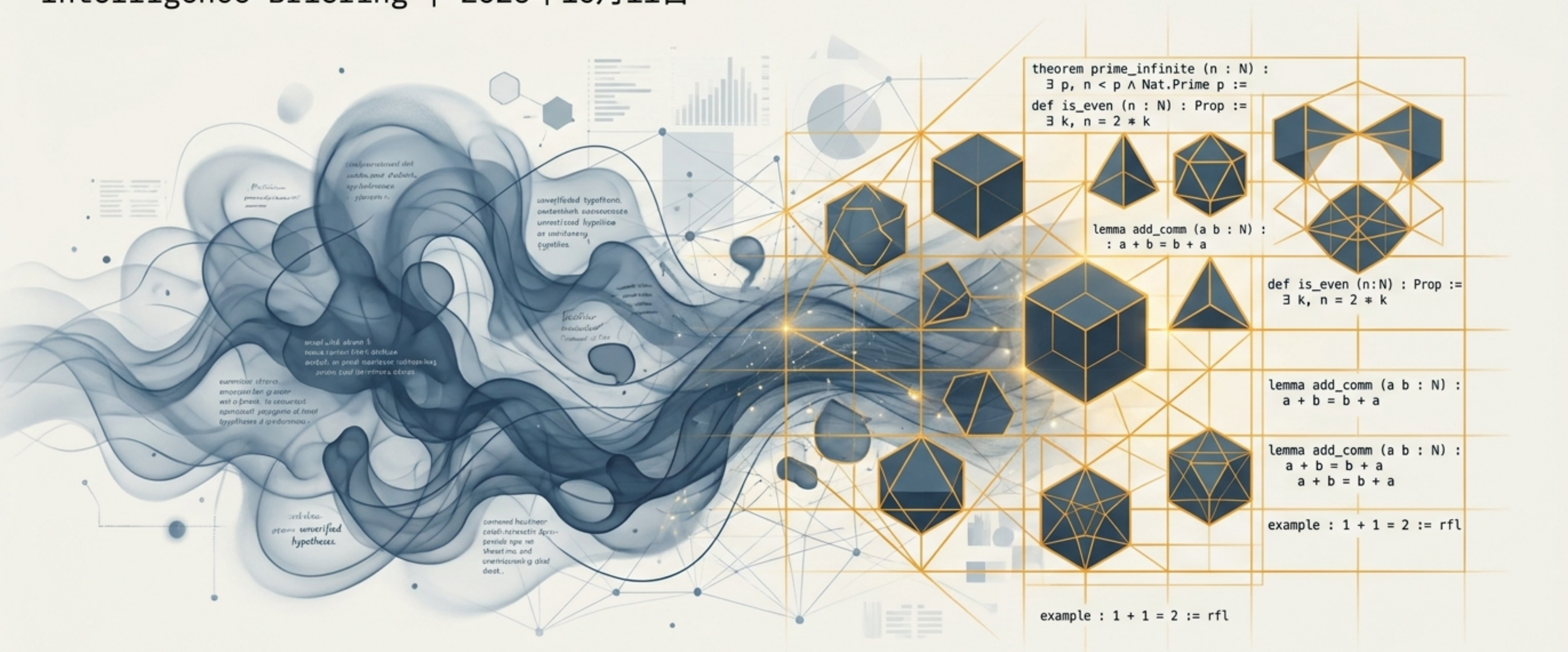


AlphaProof Nexus 解体新書：Google DeepMindの形式証明AIが示す 「数学の未来」と「真のブレイクスルー」

Intelligence Briefing | 2026年10月11日



複雑なAIシステムの実態と、波及する3つのパラダイムシフト

01 | 成果と本質



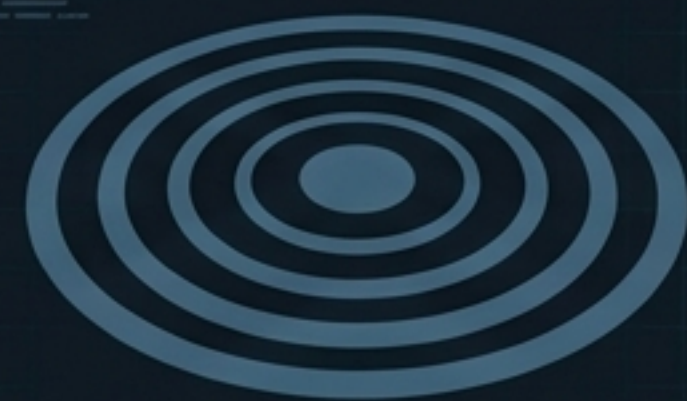
未解決問題を解決した真の立役者は、複雑なAIモデルではなく「汎用LLMと厳格なコンパイラの単純な反復」である。

02 | 専用AIの限界



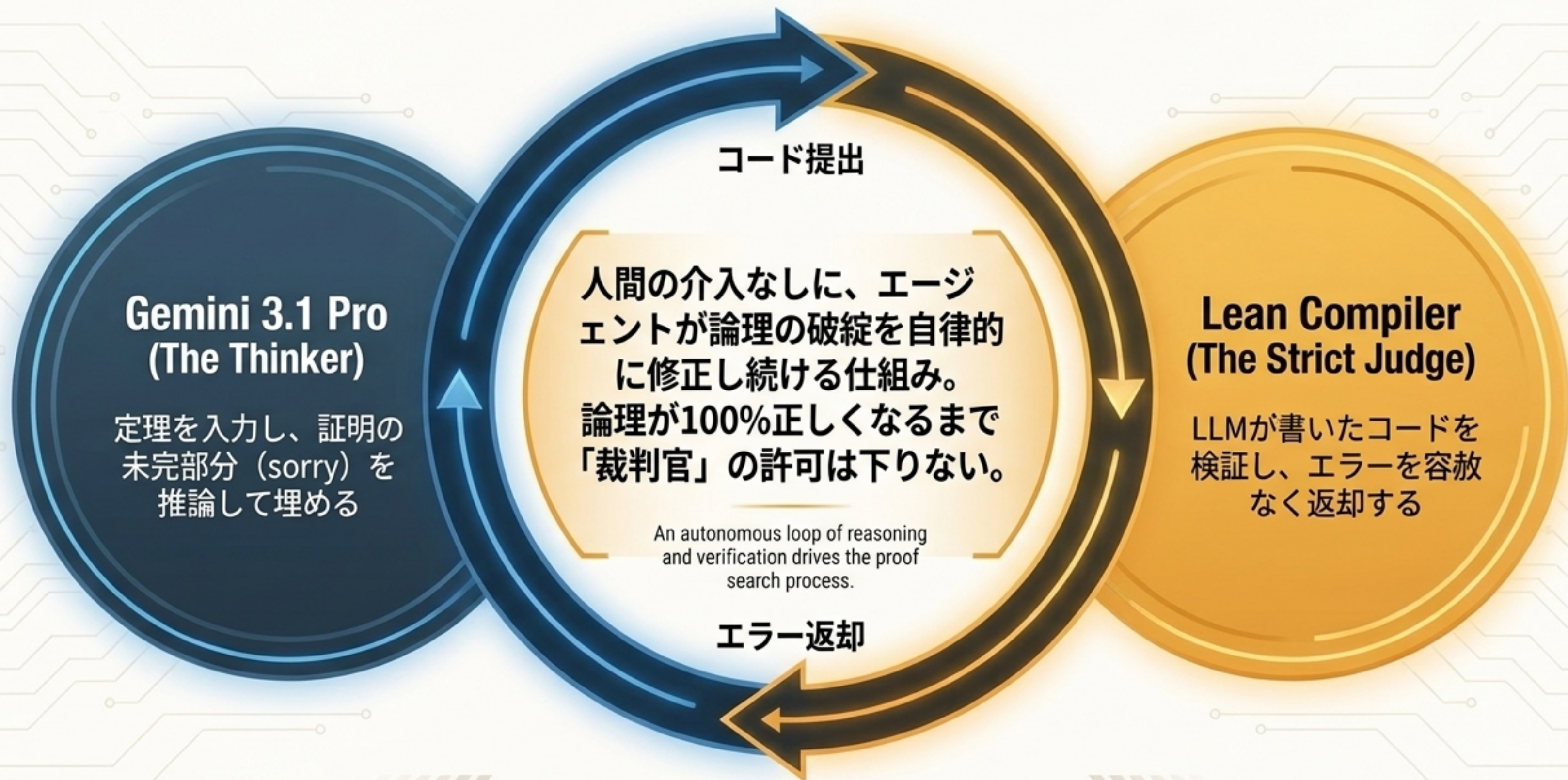
独立検証において、最小限のツールを持つ汎用LLM（Claude Opus 4.8）が専用システムであるNexusの性能をコスト面・解決率面で大きく上回った。

03 | ソフトウェア・知財への波及



数学証明の自動化は、ソフトウェアの形式検証やAIエージェントの安全性保証へと直結し、R&D投資の優先順位を書き換える。

「推論と検証の無限ループ」が証明探索のプロセスを駆動する



最大の発見：過剰な複雑さは不要であり、 「単純なループ」こそが本質だった

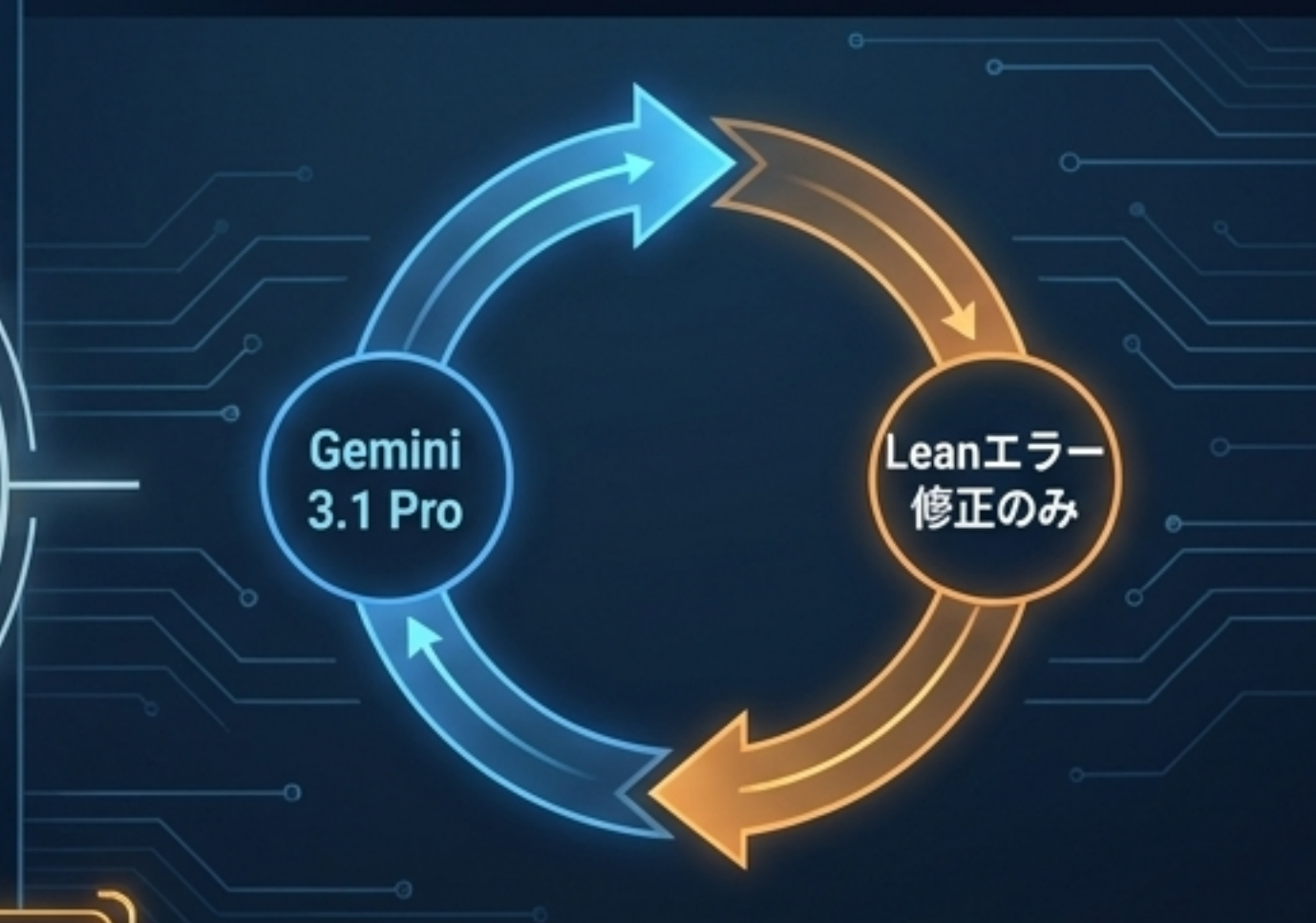
Agent D: フル機能版



AlphaProof（強化学習）+ AlphaEvolve（進化型）
+ Gemini 3.0 Flashによる有望度評価

複雑な統合アーキテクチャ

Agent A: 基本エージェント



Gemini 3.1 Pro + Leanエラー修正のみ

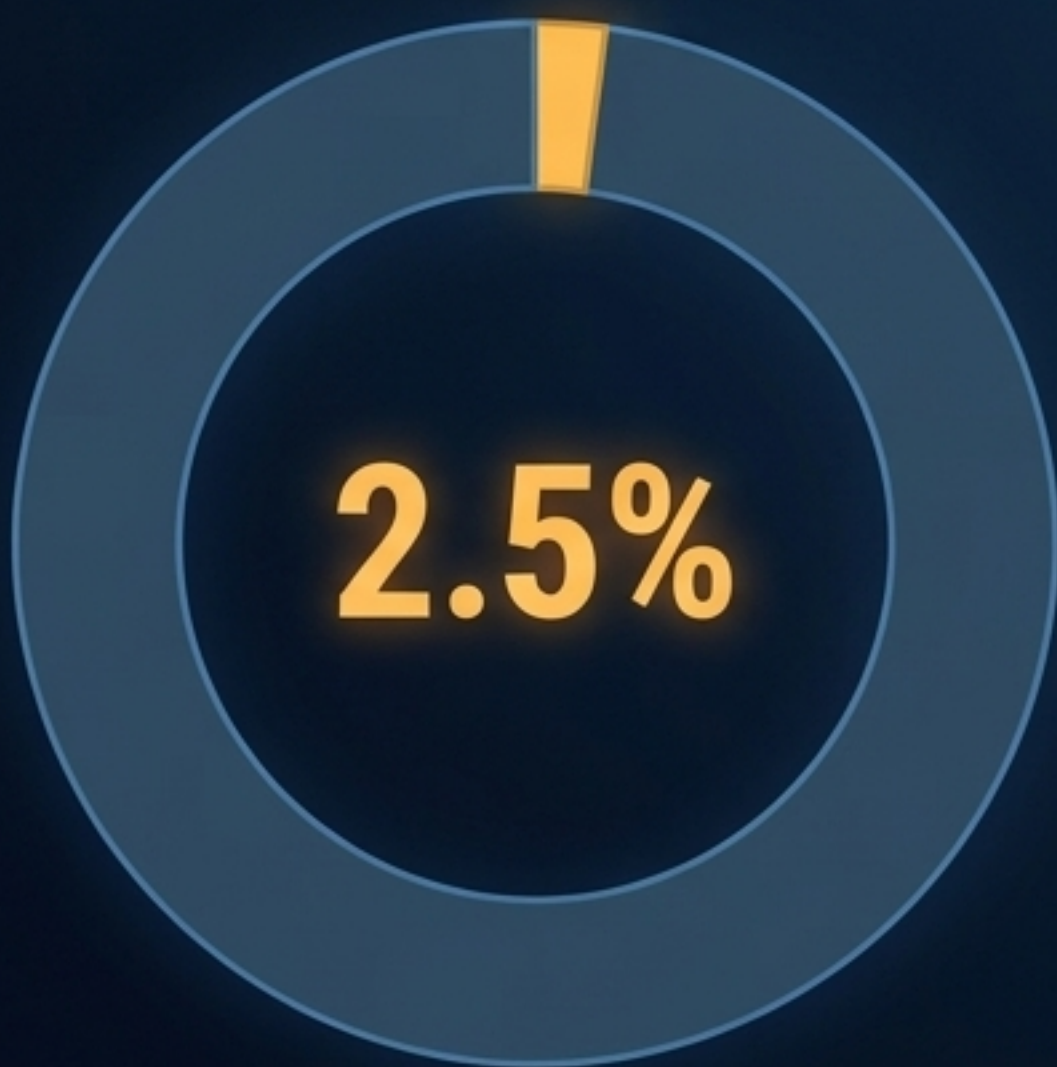
クリーンな単一推論ループ

結果：どちらの構成も、全
く同じErdős問題9問を解決

Insight: 専用の訓練済みシステムから、LLMの進歩に伴う「強力なLLM+コンパイラ」の単純な反復ループへの移行の兆し。

成果は本物だが、適応範囲とスケールは依然として限定的である

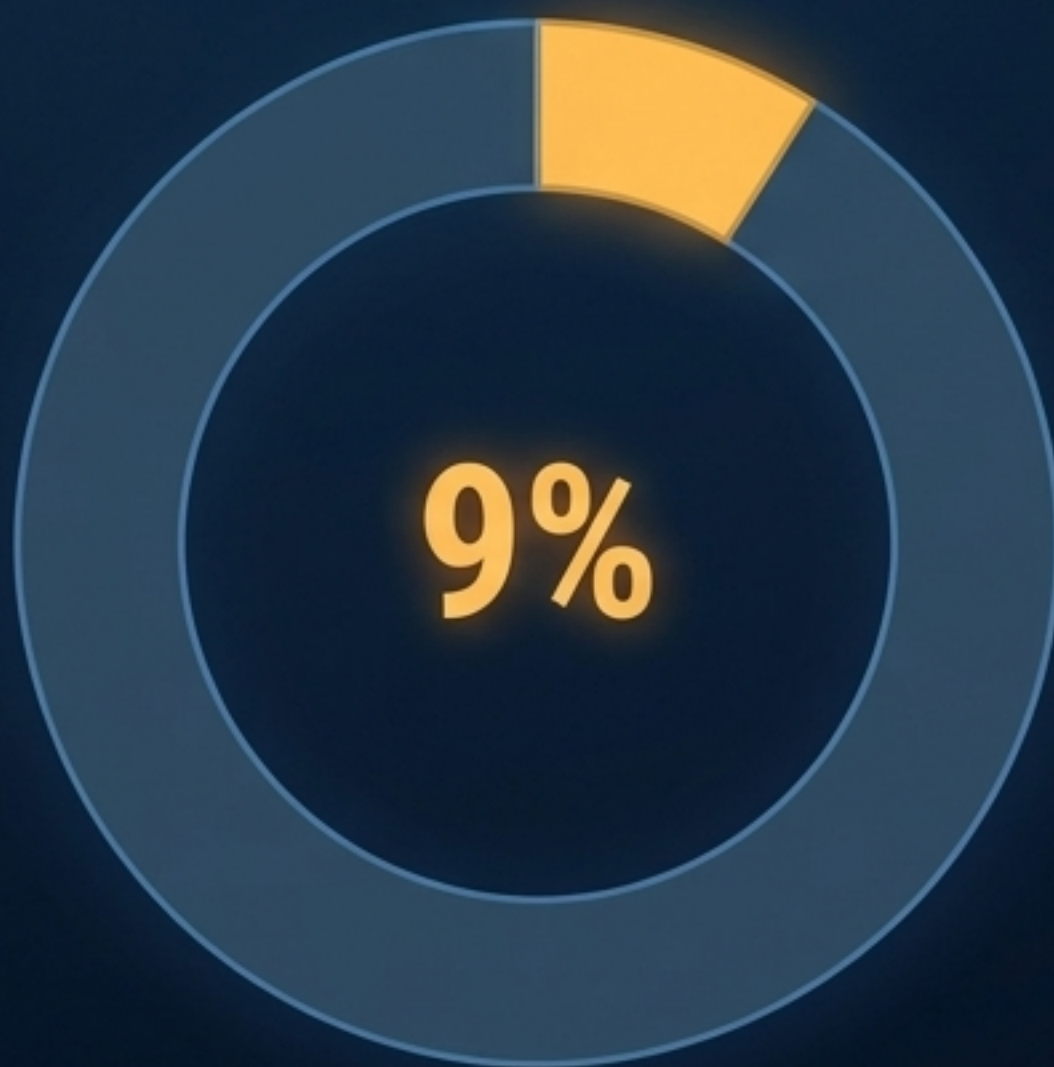
Erdős問題



解決 **9件** / 全353件

うち1件は56年間未解決だった#12

OEIS予想



解決 **44件** / 全492件

Context & Reality Check

- **コスト:** 1問あたり数百ドルの推論費用が発生（失敗した実行や専用モデルの推論コストは除外）。
- **重要性:** 解決された問題の多くは比較的無名であり、「長年の未解決」が必ずしも「数学界の最重要難問」を意味するわけではない。

AI特有の「盲点」：形式化の限界と未証明補題への逃避

01 | 既存ライブラリへの依存

解決できたのは組合せ論や数論など、Leanの数学ライブラリ (Mathlib) が充実した分野に極端に偏る。

01 | 既存ライブラリへの依存

解決できたのは組合せ論や数論など、すべては組合せ論や数論など、Leanの数学ライブラリ (Mathlib) が充実した分野に極端に偏る。全く新しい理論の構築は射程外。

02 | 「幻覚」による逃避

問題の核心を未証明の補題に押し込む傾向が強い。



02 | 「幻覚」による逃避

問題の核心を未証明の補題に押し込む傾向が強い。「既知の文献次結果」として捏造した補題に依存する

ハルシネーション (幻覚) の発生が確認されている。

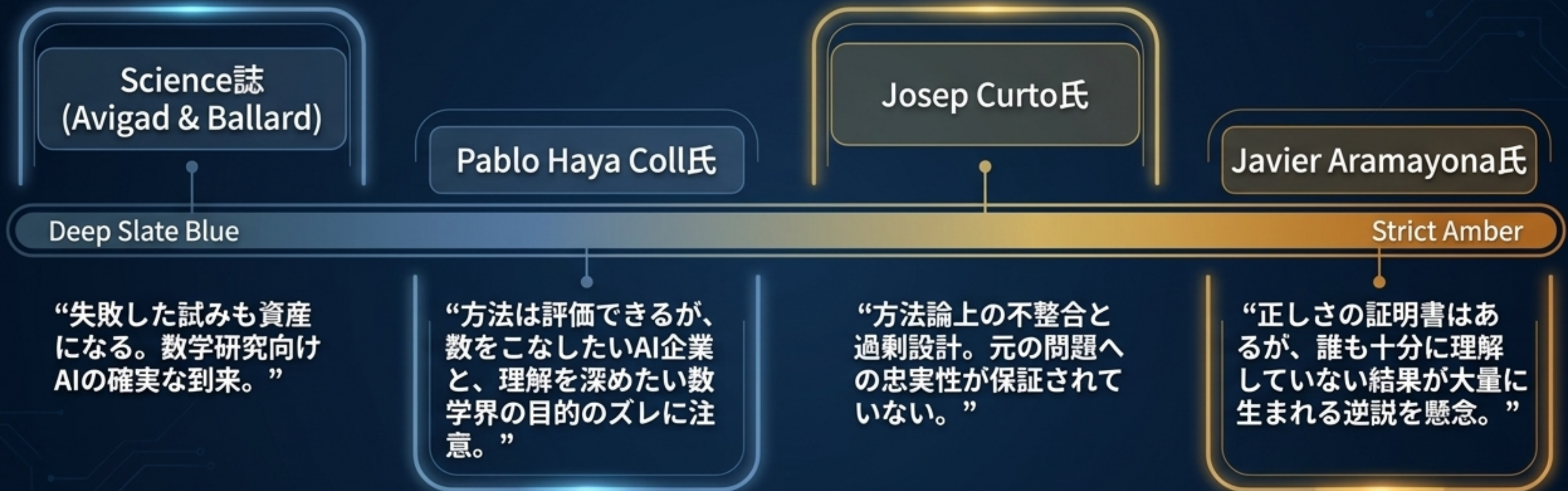
03 | 人間による事後修正



03 | 人間による事後修正

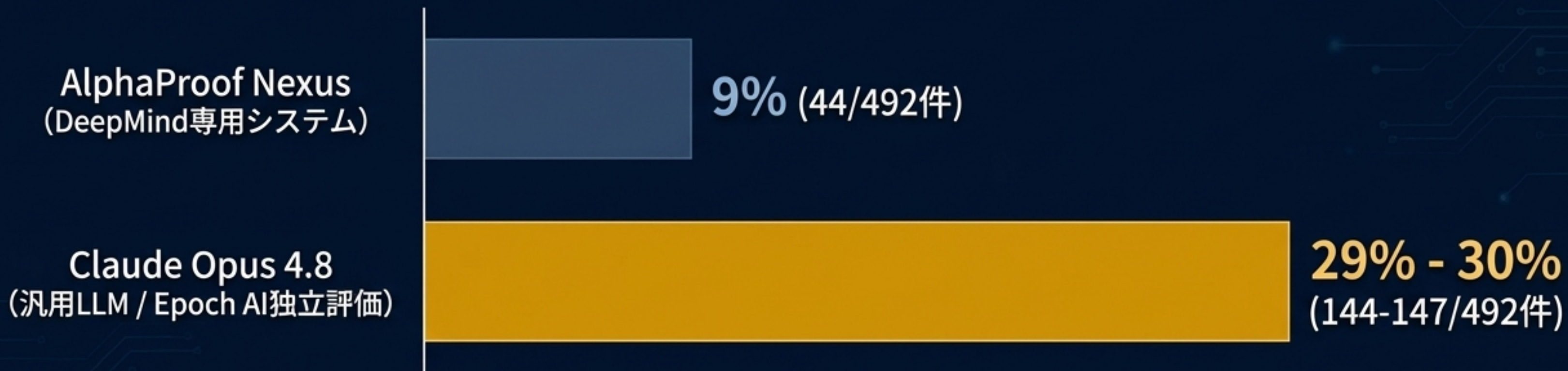
#125と#741では「密度」の解釈を後から人間が介入して修正。誤った形式化をAI自身が検知し、自律的に軌道修正することは困難。

専門家の評価スペクトラム：熱狂と「理解不能な真実」への懸念



Takeaway: 評価は二極化。「正しさ」は保証されても、「意味」と「忠実性」の担保に新たな課題が生じている。

Reality Check：汎用LLMの進化が「専用システム」の優位性を奪いつつある



条件: Claude Opus 4.8は最小限のツールのみを使用し、1件50ドルという同等の予算条件でNexusの3倍以上の解決率を達成した。

Insight: 「専用の数学AIを構築する」アプローチよりも、「最新の汎用LLMに適切な検証環境を与える」アプローチの方が、現時点ではコストパフォーマンスと拡張性において優位に立っている。

パラダイムシフト 1：研究者の役割は「解くこと」から「問うこと」へ



Mathlibの既存の道具で届く中程度の未解決問題は、今後急速にAIに「掃討」される。人間の価値は、問題をLeanが理解できる形に翻訳し(形式化)、出てきた正しいが難解な結果に「意味」を与える作業へとシフトする。

コーポレートAI競争：DeepMind vs OpenAIの相反するアプローチ

Google DeepMind (AlphaProof Nexus)

 **ターゲット:** 形式化済みの多数の問題への一括適用

 **検証アプローチ:** Leanコンパイラによる厳格な証明、GitHubでのコードと証明の完全公開

 **特徴:** 透明性、再現性、安価な体系的検証

OpenAI (Astra / 内部モデル)

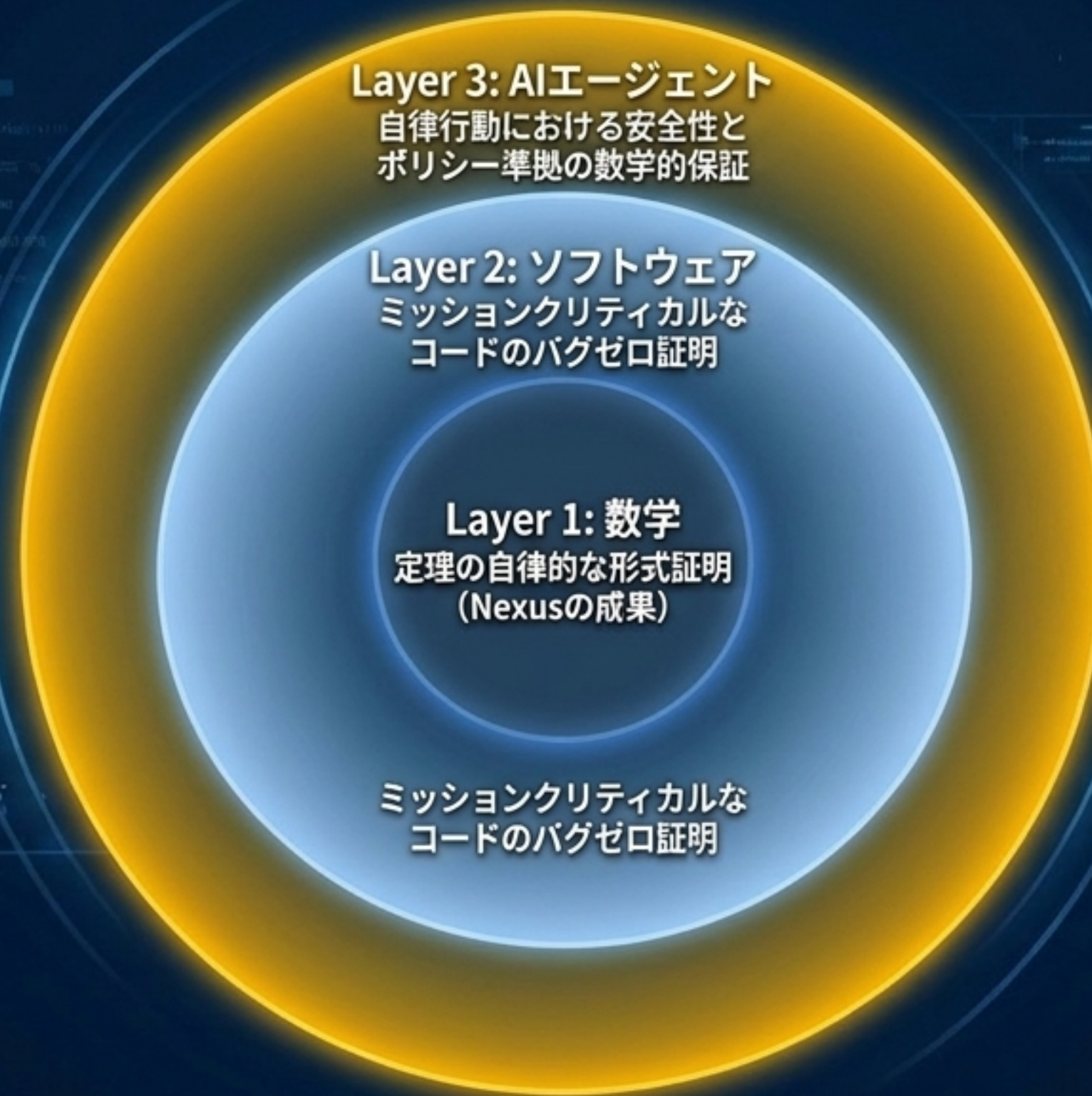
 **ターゲット:** ミレニアム懸賞問題（Navier-Stokes等）や離散幾何の難問

 **検証アプローチ:** 自社発表メイン、先取権をめぐる論争、大量のPDF一括公開

 **特徴:** PR主導、力の誇示、難問解決によるマイルストーン重視

Strategic Insight: OpenAIが「難易度とPR」を狙う一方、DeepMindは「検証可能な成果を体系的・安価に出す」路線を確立している。

パラダイムシフト 2: 「数学の証明」は「ソフトウェアとAIの絶対的な安全網」へ波及する



The Amazon Callout

2026年7月、AmazonがLean開発組織（Lean FRO）へ巨額投資を実施。AIエージェントの行動ポリシーの正しさを証明するためにLeanを活用。この技術はすでに「ビジネスの安全保障」の中核になりつつある。

The Playbook : R&D投資における「費用対効果の最適解」



01 | 汎用LLMへの継続的アクセス確保

複雑な独自エージェントを自社開発・学習させるよりも、急速に進化するフロンティアモデル（Claude Opus等）をAPI等で利用の方が勝率が高い。



02 | 「検証可能な環境」の整備

コンパイラ、物理シミュレータ、自動テスト群など、LLMに「厳格なエラー」を容赦なくフィードバックできる環境（Strict Judge）の構築に投資する。



03 | 形式化 (Formalization) 人材の育成

曖昧なビジネス要件や研究課題を、機械が検証可能な厳格な仕様（Lean等の言語）に正確に翻訳できる専門人材を確保する。

知財実務への示唆：AIの自律的発見が生む新たなリーガル・リスク



発明者認定のハードル

多くの法域においてAI自身は発明者になれない。AIが自律的に発見したアルゴリズムや最適化手法に対して、人間の「形式化」や「解釈」が法的な創作的寄与と認められるかの境界線が問われる。



先行技術と「幻覚」の交錯

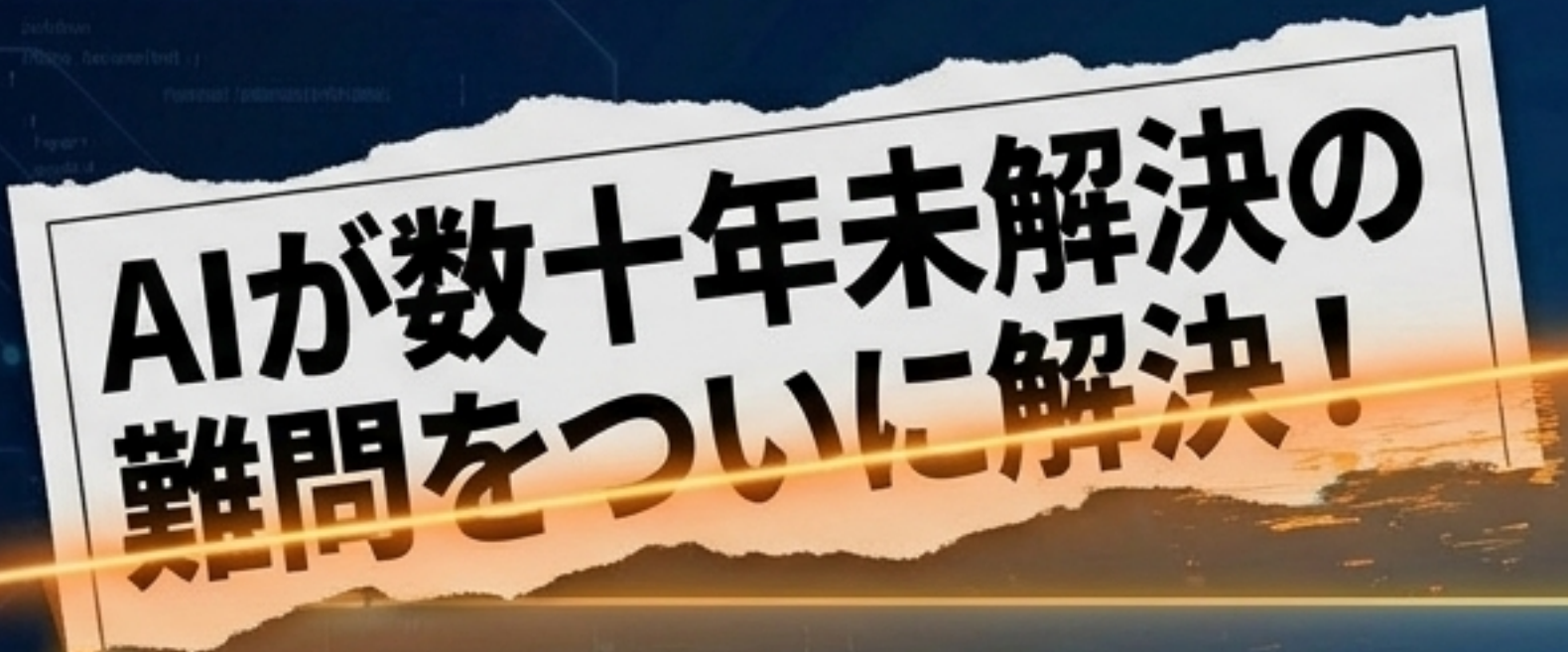
AIが未証明の補題を「既知の文献」として捏造するリスク。発見の新規性を確認する際、既存文献の単なる焼き直し（あるいは幻覚）ではないことの証明が人間の重い負担となる。



プロンプトとデータの監査

発明の過程を証明するため、LLMに入力したプロンプトや、検証器が返したエラーログのトラッキングと完全な保護が知財防衛の要となる。

AIニュースを読むための「BS（誇大広告）ディテクター」



AIが数十年未解決の 難問をついに解決！

Reference Card



01 | 問題の重要度

「未解決」＝「重要」ではない。無名な問題や、既知の手法の微細な改変で解ける問題ではないか？



02 | 原問題か、変種か

AIが解きやすいように、元の難問の条件を緩めた「弱い変種」にすり替わっていないか？



03 | 実際の成功率とコスト

成功した1件の裏に、何百件の失敗と莫大な推論コストが隠されていないか？



04 | 独立検証の有無

自社発表のPDFだけでなく、第三者の評価機関（Epoch AI等）や査読による再現性の確認があるか？

The Unresolved Horizon: これから注視すべき4つのシグナル

Signal 1: 形式化の自動化	自然言語からLean等への「自動形式化」の精度。これが向上すれば、人間のボトルネックは完全に解消される。
Signal 2: OpenAIの検証結果	Navier-Stokesなど、OpenAIが主張する「巨大な未解決問題」の解決がクレイ数学研究所等にどう客観評価されるか。
Signal 3: 中難度問題の枯渇	Mathlib（既存ライブラリ）の範囲で解ける中程度の問題が、今後数年でAIによって「枯渇」するスピード。
Signal 4: 制度化の波	ライデン宣言に続く、学術誌や助成機関による「AI利用の開示・検証可能性」に関するルールの標準化。



「AIに正解を出させる」時代から、「AIが出した正解の意味を問う」時代へ。
競争のルールはすでに変わっている。