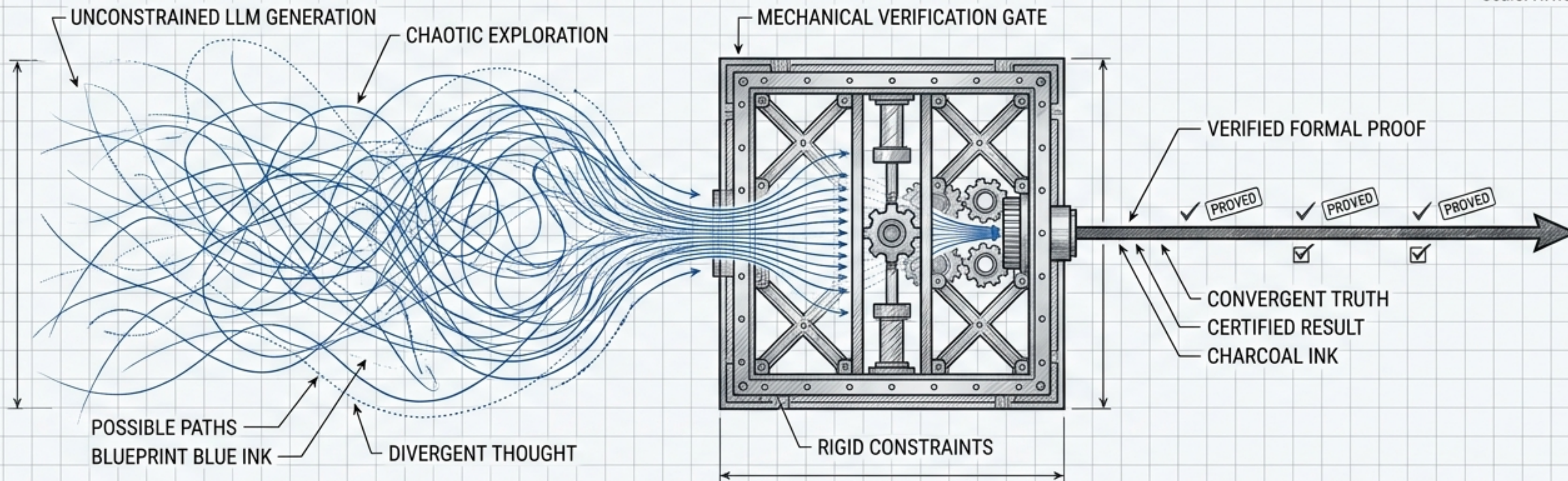


AlphaProof Nexus: AIによる数学的探索の真実と産業への影響



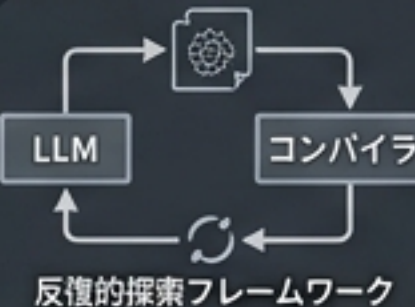
AIが数学者を代替したのではない。これは、LLMの無軌道な生成力を「機械検証のループ」に閉じ込めることで生み出された、新たな知の建築様式である。

PROJECT: AlphaProof Nexus	
DRAWING TITLE: CORE METAPHOR VISUALIZATION	
SHEET NO:	SHEET NO: A-101

メディアの神話

巨大な単一のAIモデルが、
数学の未解決問題を次々と
自律的に解き明かした。

観測された現実



エージェント機構：
単一モデルではなく、LLMと
コンパイラを組み合わせた
反復的探索フレームワーク。

353

Erdős問題：対象 353
件中、解決は 9 件（成功率
約2.55%）。
大半の難問は未解決。

44件

OEIS数列：
492 件の自動形式化のうち、
従来未証明で正しく
形式化されたのは 44 件。

価値は「万能性」ではなく、自動探索から機械検証可能な証明までを
一貫して出力するワークフローが実証された点にある。

1 研究上のブレークスルー性 [高い]

自然言語ではなく、Lean環境での探索・検証を大規模に実証。

2 証明の論理的信頼性 [高い]

sorryなしのLean証明とSafeVerifyによる完全な形式的妥当性。

自律数学者への到達度 [まだ低～中]

Leanが成熟した分野 (組合せ論等) に偏在。新理論の構築は困難。

4 問題設定の信頼性 [人間の審査が必要]

自然言語からLeanへの誤形式化 (仕様エラー) は機械では防げない。

6 教育・産業への影響 [有望だが制度変更が必要]

「証明の完成」だけを評価する従来の仕組みは機能不全に。

5 再現性 [混合]

証明成果物 (Leanコード) は完全だが、探索過程・コストは高分散で不透明。



価値は「万能性」ではなく、自動探索から機械検証可能な証明までを一貫して出力するワークフローが実証された点にある。

AlphaProof Nexus (The Orchestrator)

LLM、検証器、専門ツールを統合し、反復的な研究探索ループを回すエージェント・フレームワーク

Gemini 3.1 Pro / Flash

発見と推論を担う
フロンティアLLM
(Generative Engine)

Lean 4 / SafeVerify

証明の論理的妥当性を判定するゲートキーパー
(Verification Engine)

AlphaProof

強化学習とツリー探索を行う専門ツール
(Specialized Prover)

価値は「万能性」ではなく、自動探索から機械検証可能な証明までを一貫して出力するワークフローが実証された点にある。

Layer 4: Agent D (Full Integration)

上記すべての統合と、グローバル・キャッシュによる
既知サブゴールの再利用。

Layer 3: Agent C (Evolutionary Search)

+ 共有Population & Gemini 3.0 Flash
未完成の証明スケッチを複数保持し、Eloスコアで有望なものを選択。

Layer 2: Agent B (Tool Integration)

+ AlphaProof
ツリー探索による専門的な証明候補 (proof / failure) の提供。

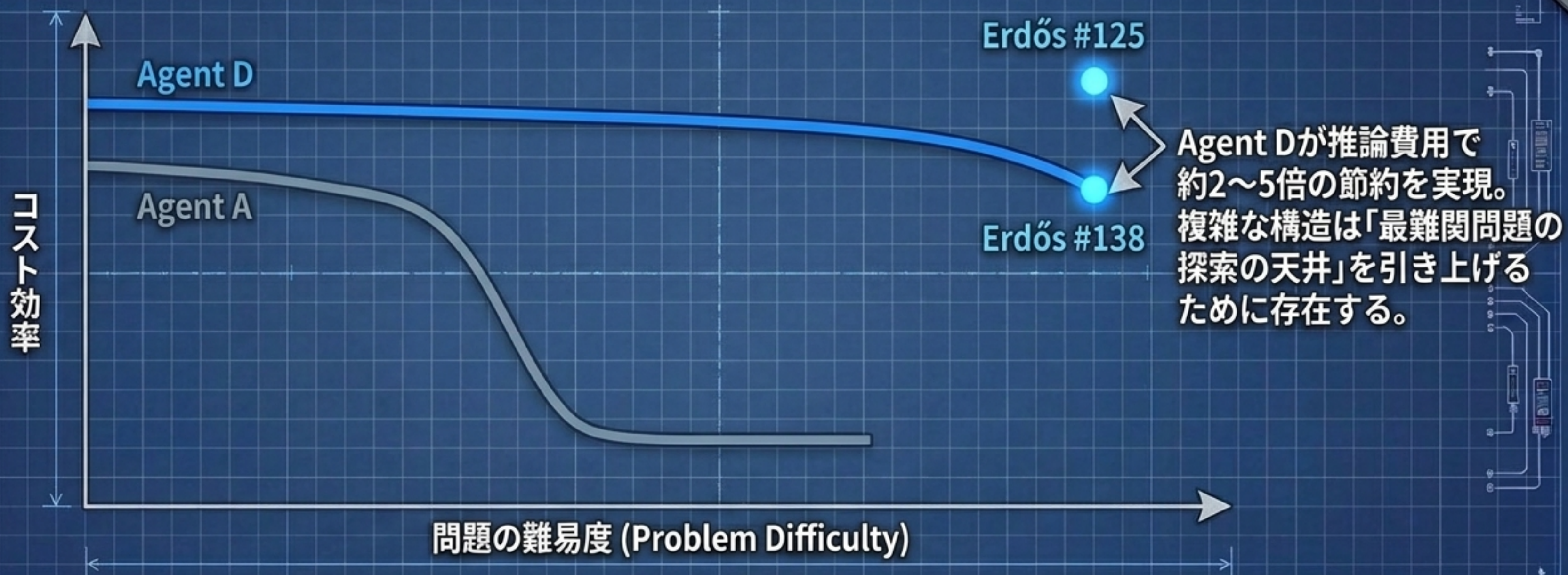
Layer 1: Agent A (The Core Engine)

Gemini 3.1 Pro $\rightleftharpoons$ Lean Compiler
LLMがコードを編集し、コンパイルエラーをフィードバックとして再試行。

Key Insight:
LLMは「発見エンジン」。
論理的妥当性の
「最終判定者」は
常にLeanが担う。

価値は「万能性」ではなく、自動探索から機械検証可能な証明までを
一貫して出力するワークフローが実証された点にある。

SMCの批判: 「単純なAgent Aでも同じ
9問を解けた。進化的探索 (Agent D) は
過剰設計ではないか？」



価値は「万能性」ではなく、自動探索から機械検証可能な証明までを一貫して出力するワークフローが実証された点にある。

Semantic Space (人間の意図)

「Leanが保証しないもの」
「Leanが保証しないもの」

形式化された定理が、人間が
意図した自然言語の数学問題と
完全に同じ意味であること。



実例: Erdős #125での
「density」概念の誤解釈。

Translation / 形式化



Formal Space (Leanコード)

「Leanが保証するもの」
「Leanが保証するもの」

形式的証明が、許可された
公理から論理的に正しく導出
されること (sorry なし)。



AIのハルシネーションは
ここで完全に遮断される。

**Takeaway: 人間による「意味論的レビュー」は、
AI時代において最も不可欠な仕事として残る。**

価値は「万能性」ではなく、自動探索から機械検証可能な証明までを
一貫して出力するワークフローが実証された点にある。

対話型定理証明環境 (ITP)

例: Lean 4, Coq/Rocq,
Isabelle

役割: 人間が操作し、
論理的妥当性をチェック
する「形式基盤」。AIは
自律的に動かない。

ニューラル自動定理証明器 (Neural ATP)

例: GPT-f, AlphaProof
(単体)

役割: ニューラルネット
ワークを用いて、局所的な
証明ステップを生成・探索
する「専門ツール」。

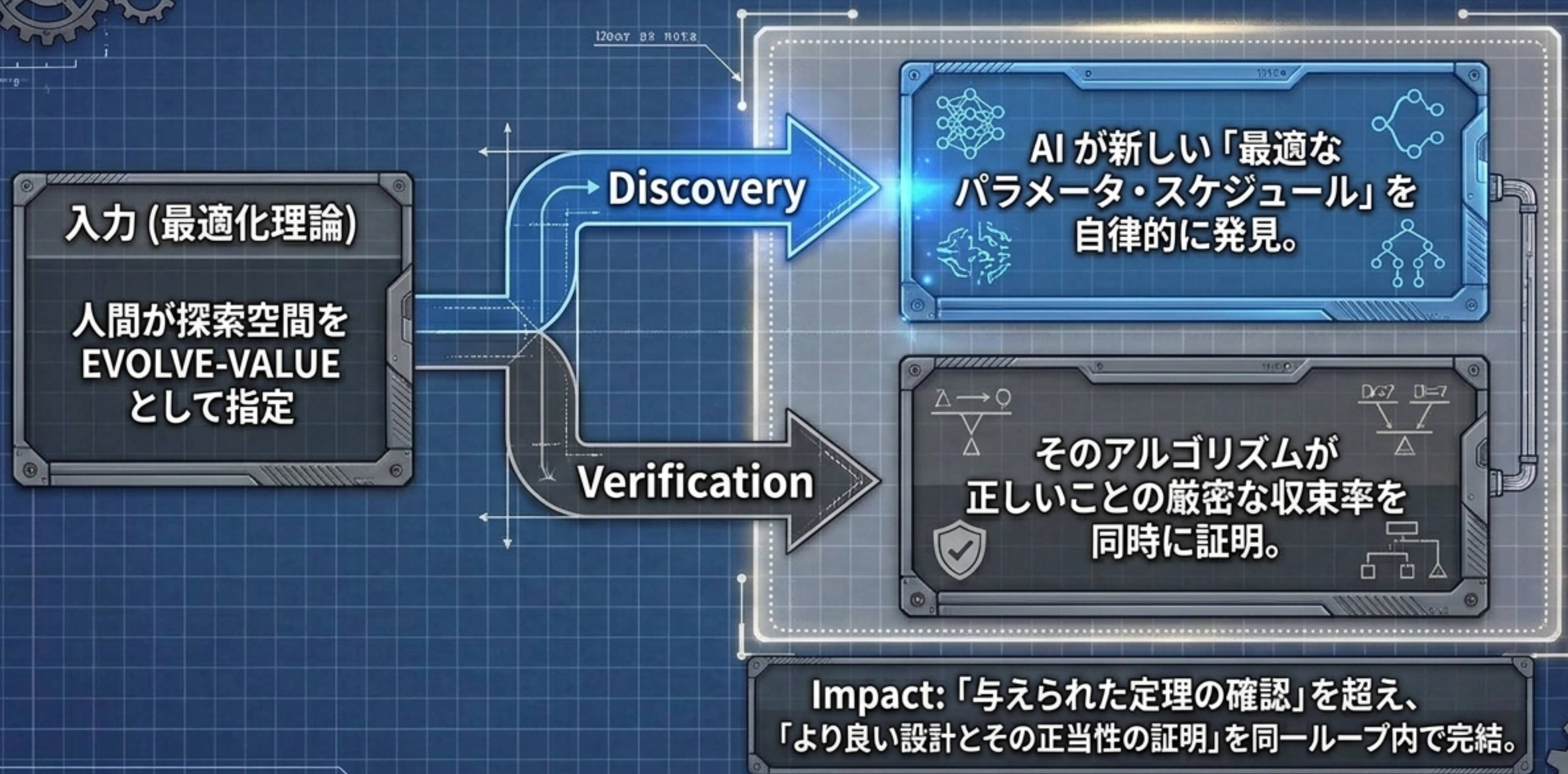
自律的エージェント・ フレームワーク

例: AlphaProof Nexus

役割: 汎用LLMがITPを
反復操作し、専門ATPを
ツールとして呼び出しな
がら、研究レベルの未解
決問題を長時間探索する
自律システム。

価値は「万能性」ではなく、自動探索から機械検証可能な証明までを
一貫して出力するワークフローが実証された点にある。

「証明」から「発見」へ：アルゴリズムの同時探索



DYNAMIC FLOW DIAGRAM

TIME-TO-IMPACT: 研究から実用化へのタイムライン

数学研究・最適化 (Immediate)

アルゴリズム設計と
証明の同時探索。
未解決問題における
補題・反例の並列探索。

ソフトウェア・ハードウェア検証 (Near Future)

検証人材のボトルネック
緩和。仕様証明
(Specification proof) の
自動化コスト削減。

暗号・セキュリティ (Careful)

セキュリティ証明、
プロトコルの不変性の
形式検証への応用。
※Nexusが既存の暗号を
破ったという事実はない。

SOLID LINES = PROVEN IMPACT | DASHED LINES = NEAR FUTURE | DOTTED LINES = SPECULATIVE/CAREFUL

旧来の評価ループ (機能不全)

証明を書いて提出せよ

結果のみを採点

学生は理解していなくても、AIを使えば完全な機械検証済み証明を提出できてしまう。

AI時代の新しい教育・評価ループ

Step 1: 説明 (Explain)

AIが生成した補題の
数学的意味を
口頭で説明させる。

Step 2: 変形 (Modify)

前提条件を変更し、
AIの証明をどう
一般化できるか問う。

Step 3: 意図の検証 (Validate)

故意に仕様エラーを含めた
AI出力を渡し、自然言語と
Leanのズレを発見させる。

**Black Box: 探索過程の再現性
(Opaque)**

Gemini 3.1 Proのパラメータ
数や学習データは秘密。
探索の分散が極めて高く、
膨大な計算コストのため
第三者の再実行は困難。



**Glass Box: 証明の再現性
(100% Transparent)**

Leanの成果物は公開されて
おり、論理プロセスは
完全に透明で、誰でも瞬時に
機械的チェックが可能。

**結論：証明成果物自体は完全に検証可能だが、
それに至る「AIの発見プロセス」の再現性は著しく低いパラドックス。**

数学的労働のアンバンドリング

かつて一人の人間の脳内で行われていたプロセスが、3つの役割に分離される。

Pillar 1:

意味・理解・問題設定
[Human]

原問題の自然言語での
定式化。仕様の策定。
生成された補題の概念抽出
と、新理論への一般化。

Pillar 2:

大規模探索・発見
[AI]

LLMと進化的アルゴリズム
による、並列的かつ高速な
候補・反例・パラメータ
スケジュールの生成。

Pillar 3:

論理の機械検証
[Compiler]

Leanによる、無慈悲で
完璧な論理の正当性検査
(ハルシネーションの
確実な遮断)。

モジュラルの実行統証サミナー

1. 研究者へ

AIを「代筆者」ではなく
「検証可能な探索装置」
として扱うこと。

成功した証明だけでなく、
失敗したスケッチ（未解決
のサブゴール）も保存し、
AIの思考の行き詰まりを
人間の洞察のヒントとする。

2. 研究機関・ 政策立案者へ

Leanやmathlibのような
「形式数学ライブラリ」を、
AI時代の不可欠な
AI時代の不可欠な
公共科学インフラとして
資金援助・育成すること。

3. 社会・メディアへ

AI数学成果の品質を区別す
るため、以下の4つの独立し
たラベル表示を要求する：

- ☒ 人間レビュー済み
(Semantic Valid)
- ☒ 機械検証済み
(Proof Checked)
- ☒ 第三者再検証済み
(Replicated)
- ☒ 自然言語で説明済み
(Explained)