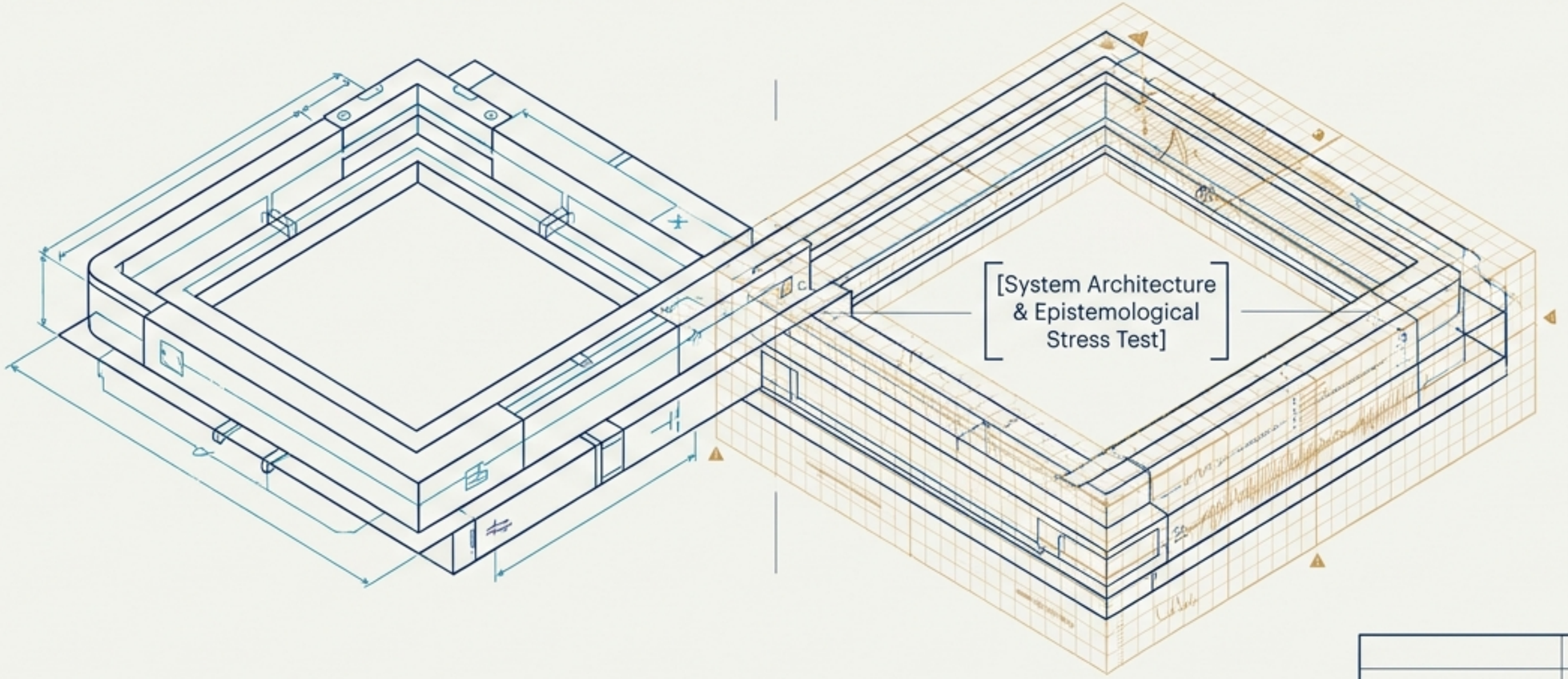
ScientistTwo: 自律型AI研究の到達点と「知識のフロンティア」の錯覚



未解決課題を入力し、完全なSOTAコードと査読済み論文を出力する

致命的欠陥を排除した、自律エージェント統合度の現行最高水準

	The AI Scientist (2024)	ScientistOne (2025)	✨ 知識の複利化
アイデア探索	場当たり的探索	木探索	アンカー付き生成 - 過去の失敗を活用
評価	全件実行 - リソース浪費	全件実行	サブセット先行評価
反論機構	テキスト修正のみ	体系的反論なし	自律的コード実行と 追加実験
再現性・ ハルシネーション	⚠ ⚠ 架空引用・ 無限ループ頻発	CoE導入により 一部改善	✓ 発生率ゼロ 隔離環境監査で 発生率ゼロ

探索空間を絞り込む「錨付け」と計算資源を節約する初期スクリーニング

過去の失敗履歴

Limitation Extractor & Verifier

無作為な探索を排除し、ボトルネック解消に直結する仮説のみを生成。

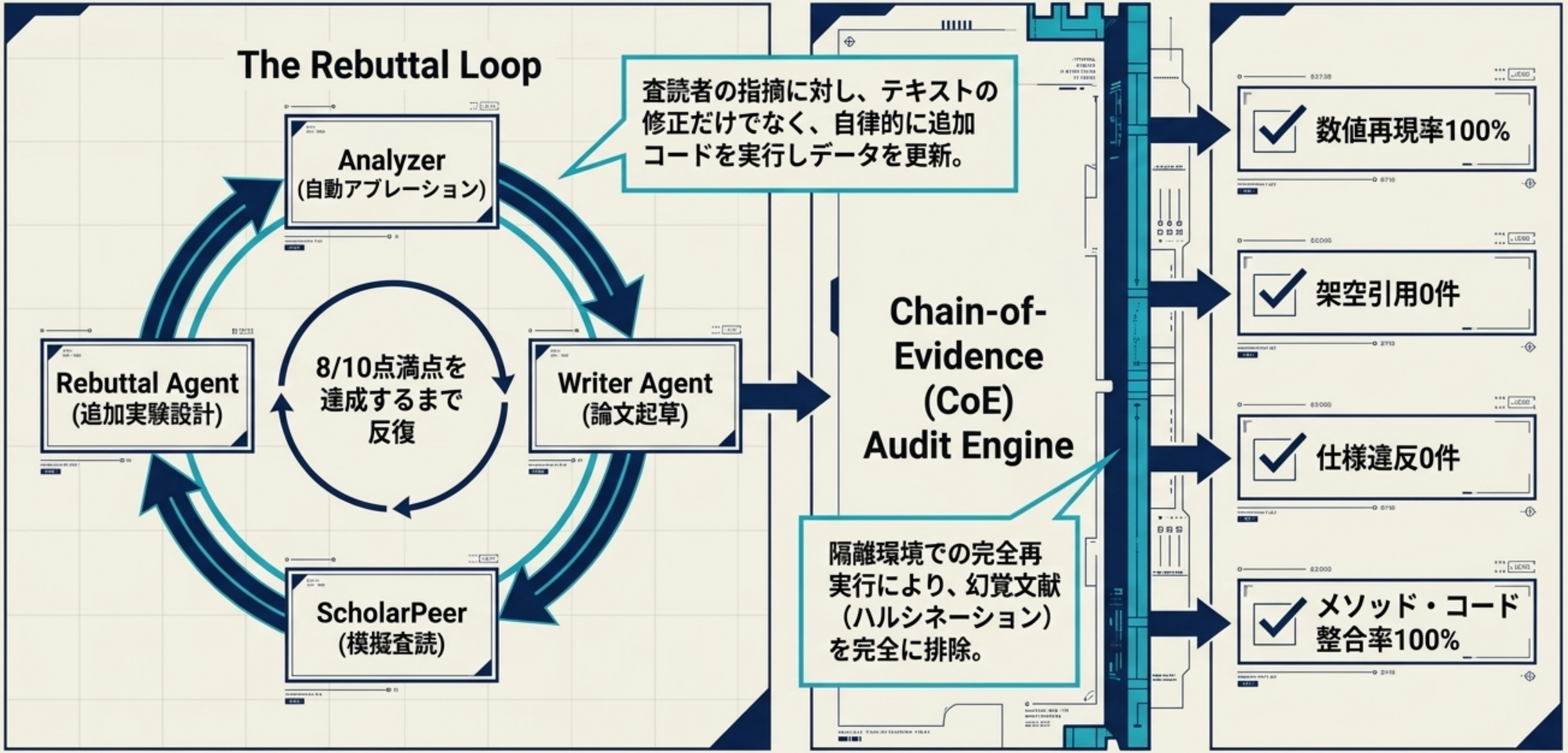
Idea Evolver

Evaluator: サブセット先行評価

縮約データで有望なシグナルを示した仮説のみを本番環境へ移行し、計算資源の浪費を抑止。

フルスケール実験へ昇格

批判的疑念を解消する追加実験ループと、100%の再現性を担保する監査機構



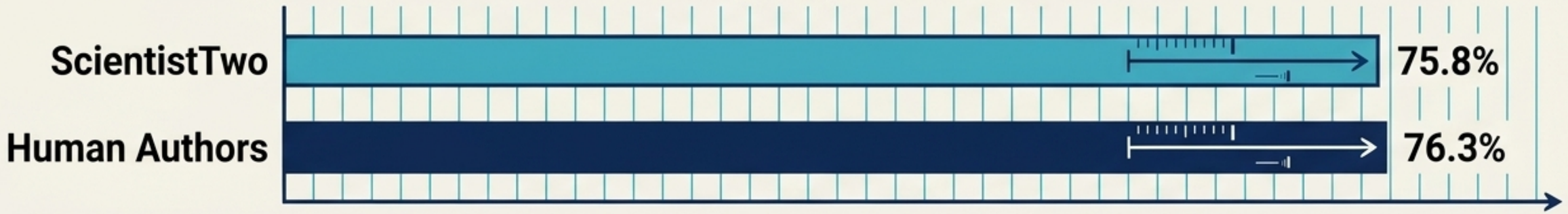
トップティア国際会議の107課題における圧倒的な実証成果

80.4%
86/107課題で人間のSOTA
ベースラインを凌駕

0.0%
1,814件の引用文献中、
架空引用発生ゼロ

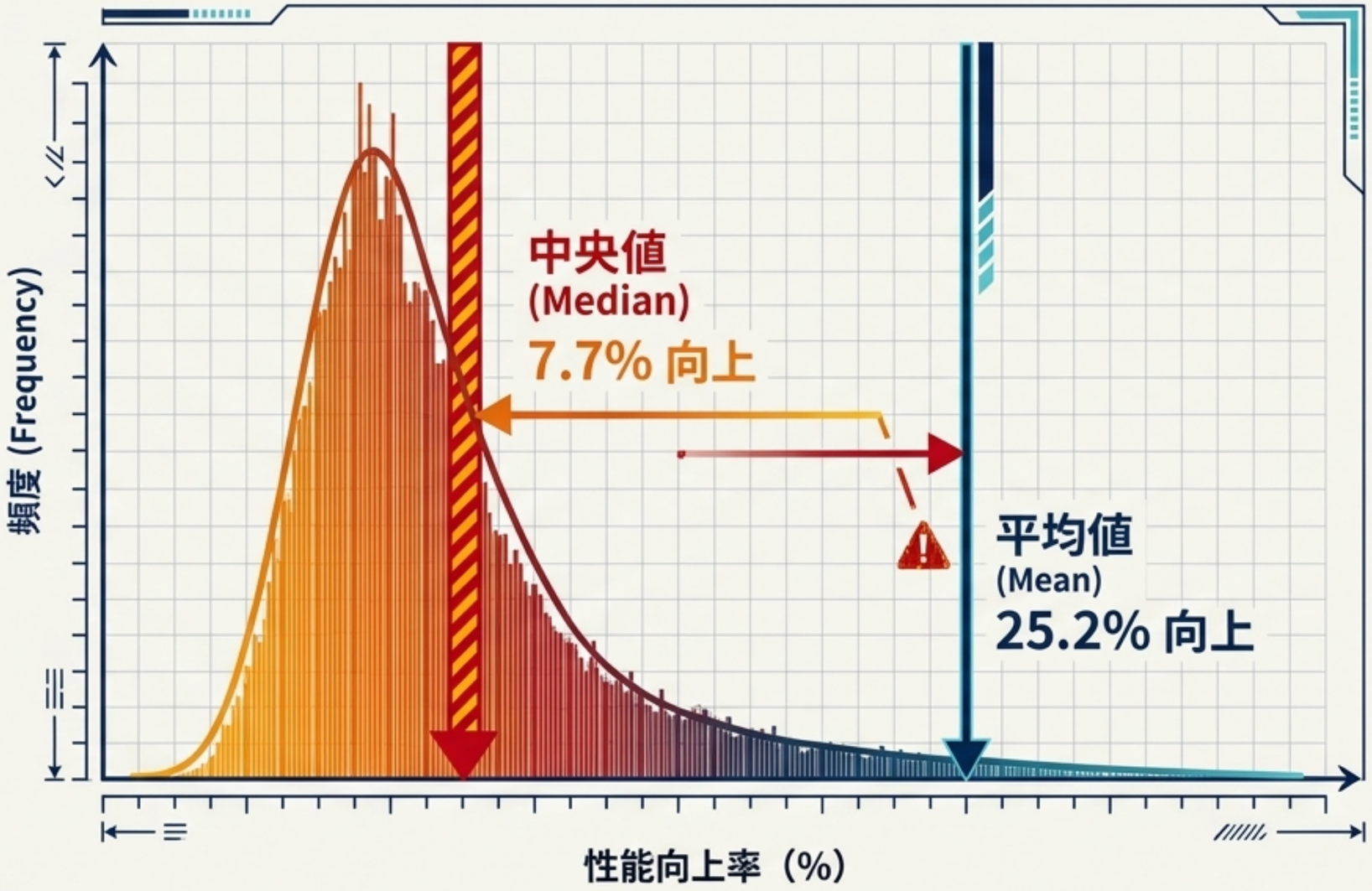
100%
49件のコード監査において
隔離環境で完全再現

Stanford Agentic Reviewer 採択率



LLM、ロボティクス、強化学習から数理最適化まで、広範なサブフィールドで
人間が設計したベースラインを突破。

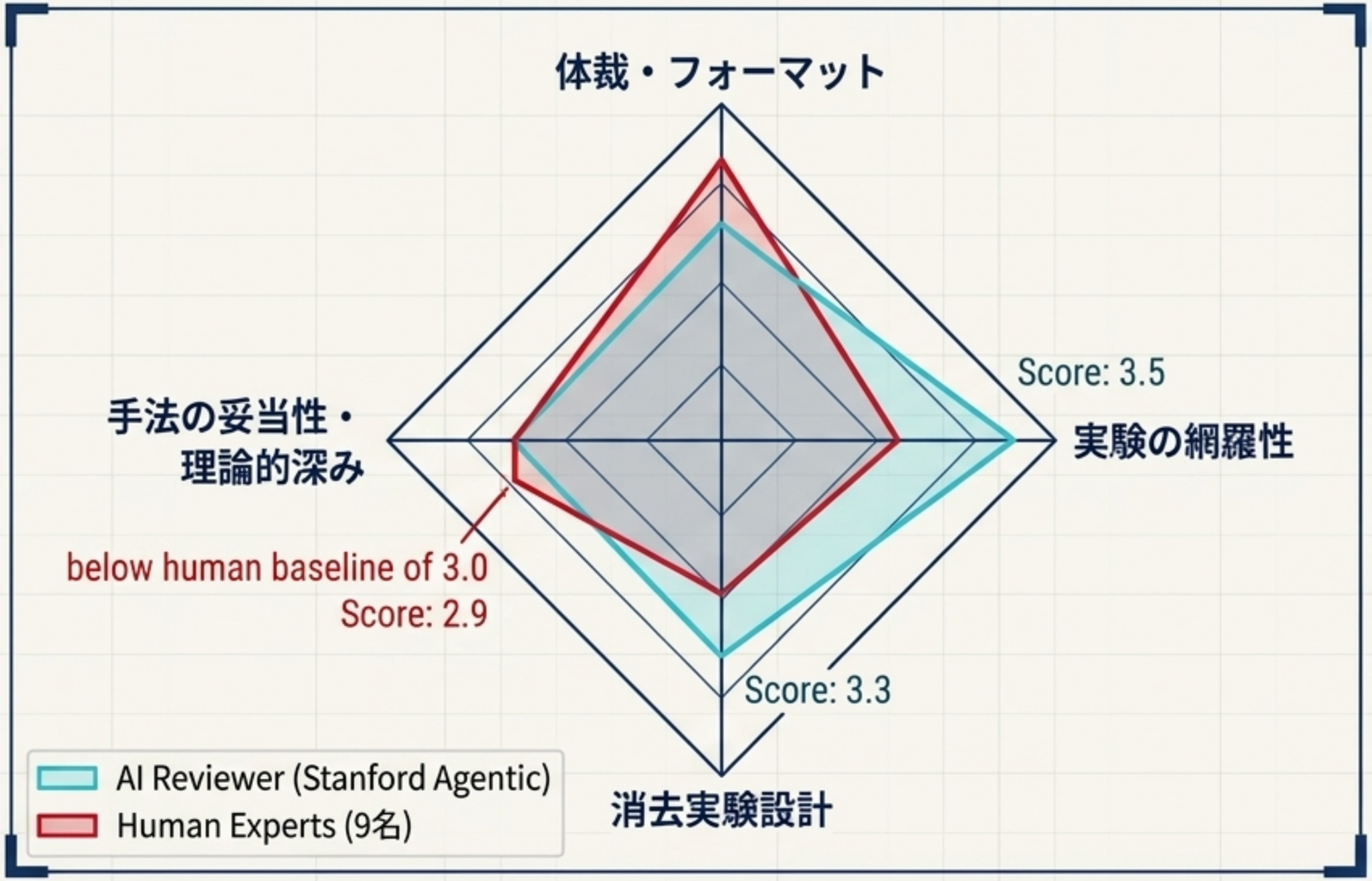
平均25.2%の性能向上が隠す、局所的・漸進的改善の実態



システムが提示する『平均25.2%』の性能向上は、ごく一部の外れ値（極端な成功）によって引き上げられた錯覚に過ぎない。

大多数の成果は1桁台のスコア向上であり、既存パラダイム内のハイパーパラメータ調整やモジュール付加の域を出ていない。

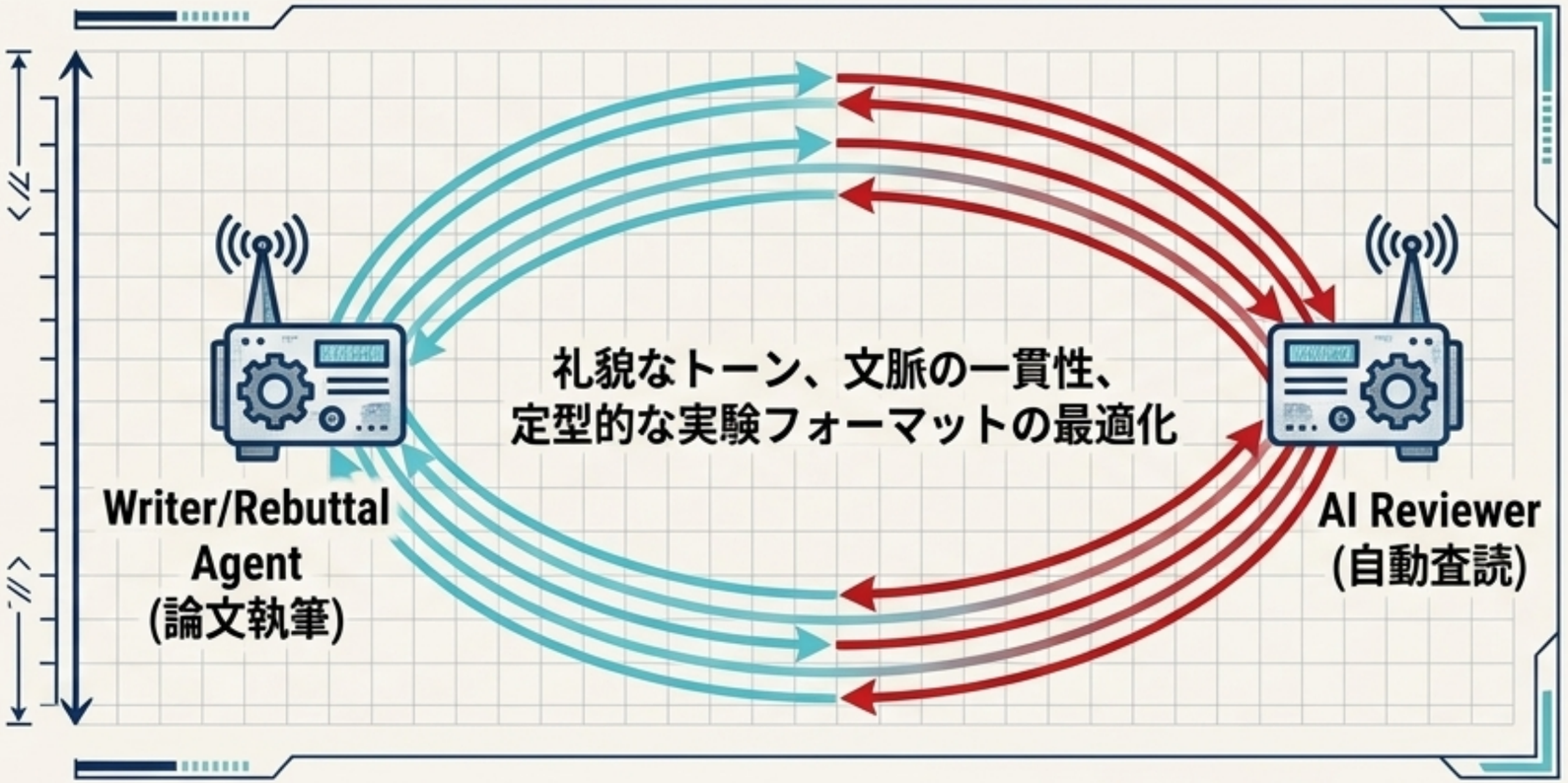
機械と人間の評価非対称性：形式美は凌駕し、理論的深みは劣後する

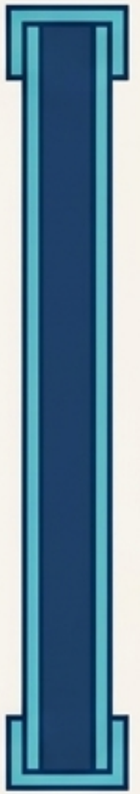


AI査読器は『計算資源を投じた網羅的な実験』と『整然としたフォーマット』に高得点を与える。

一方、人間の専門家はアルゴリズムの動機付けと概念的深さ (Method Soundness) において、人間論文の基準を下回ると判定した。

評価の循環論法と自己言及的な共鳴室（Echo Chamber）



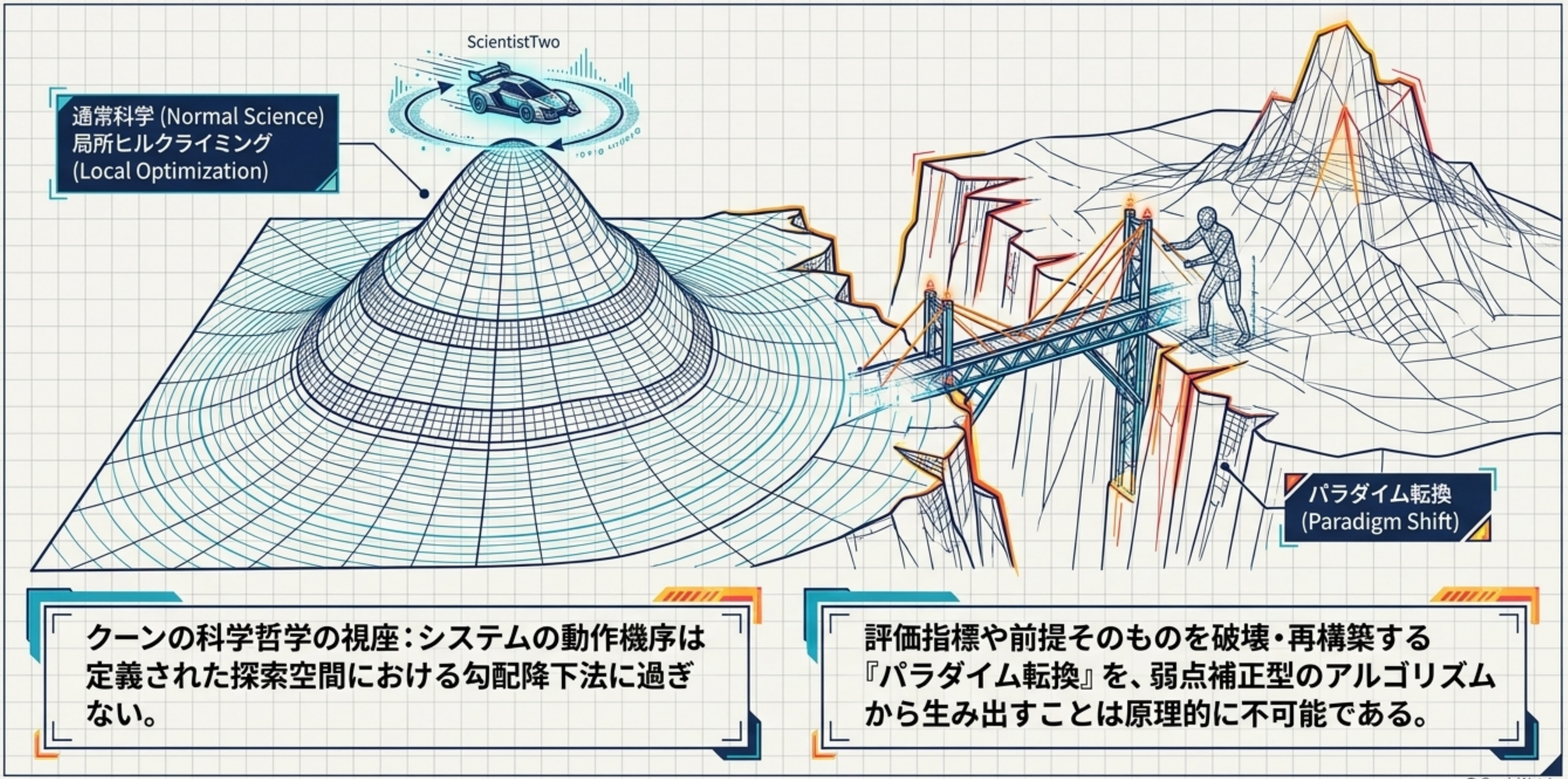


真の科学的洞察
(True Scientific Insight)

Goodhartの法則: 指標が目標になると、
それは良い指標ではなくなる。

高い採択率は真のブレークスルーではなく、LLM
査読モデルの選好（バイアス）に対する過剰適合
(Overfitting)の結果である。

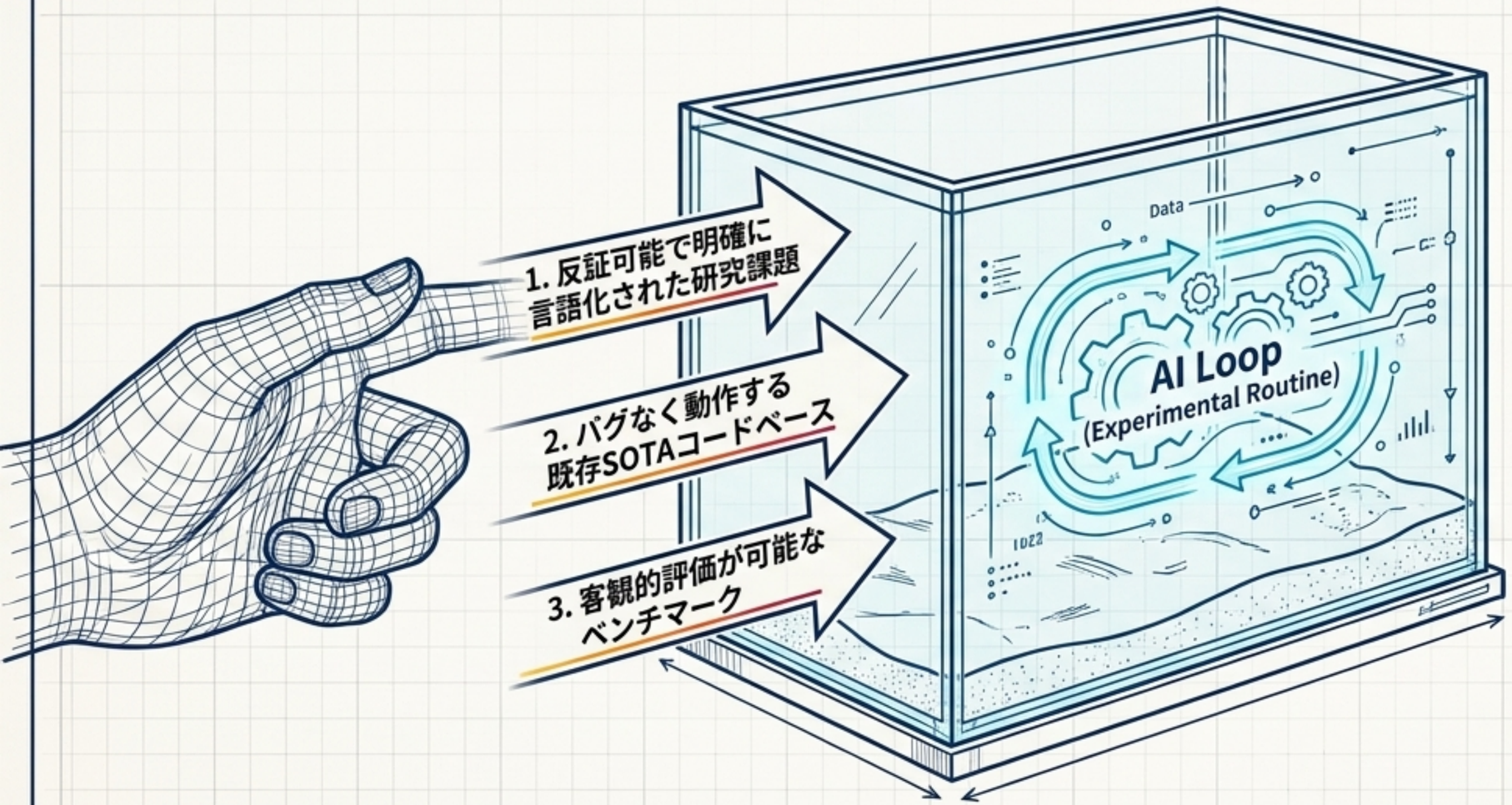
パズル解き解きに閉じ込められた「通常科学」の局所探索



クーンの科学哲学の視座：システムの動作機序は定義された探索空間における勾配降下法に過ぎない。

評価指標や前提そのものを破壊・再構築する『パラダイム転換』を、弱点補正型のアルゴリズムから生み出すことは原理的に不可能である。

砂場（サンドボックス）環境への絶対的依存と課題設定能力の欠落

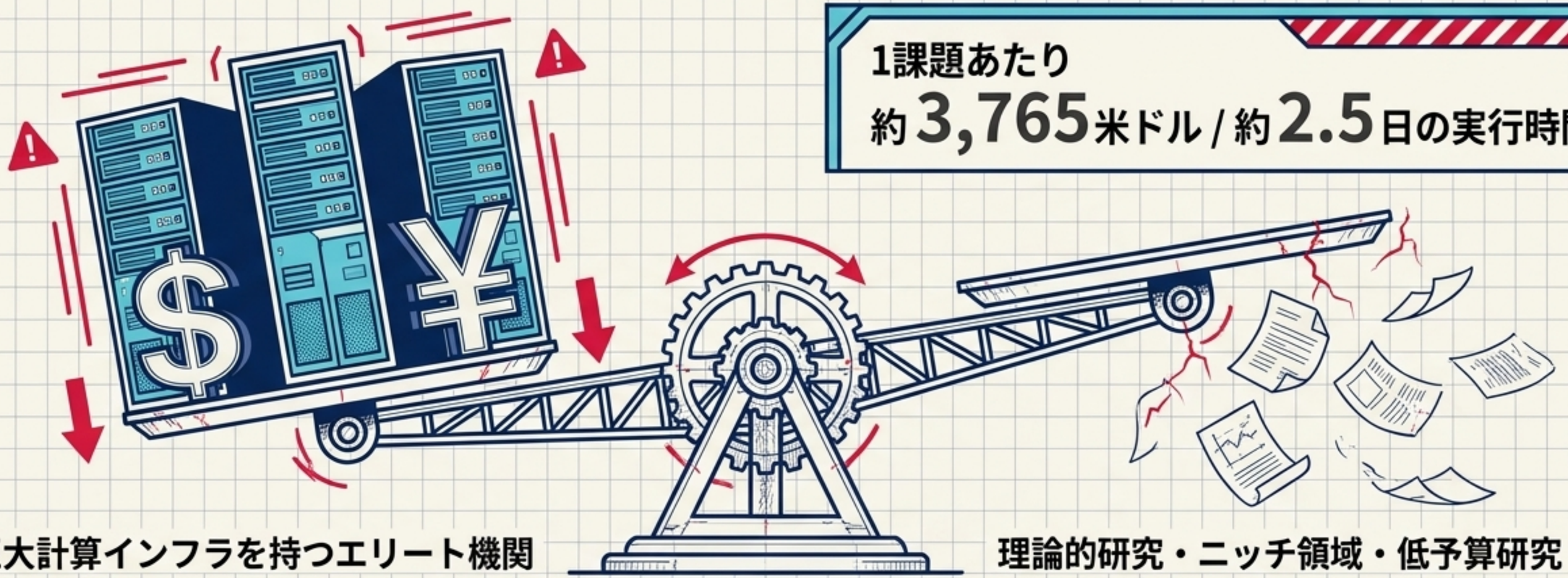


科学の歴史において最も創造性を要するのは『未知の現象から、解くべき真の問いを発見すること』である。

問題設定と測定系が完全に整備された後後の『実験ルーティンワーク』を高速代行しているに過ぎない。

計算資本の寡占による研究優先順位の構造的歪み

1課題あたり
約3,765米ドル / 約2.5日の実行時間



巨大計算インフラを持つエリート機関

理論的研究・ニッチ領域・低予算研究

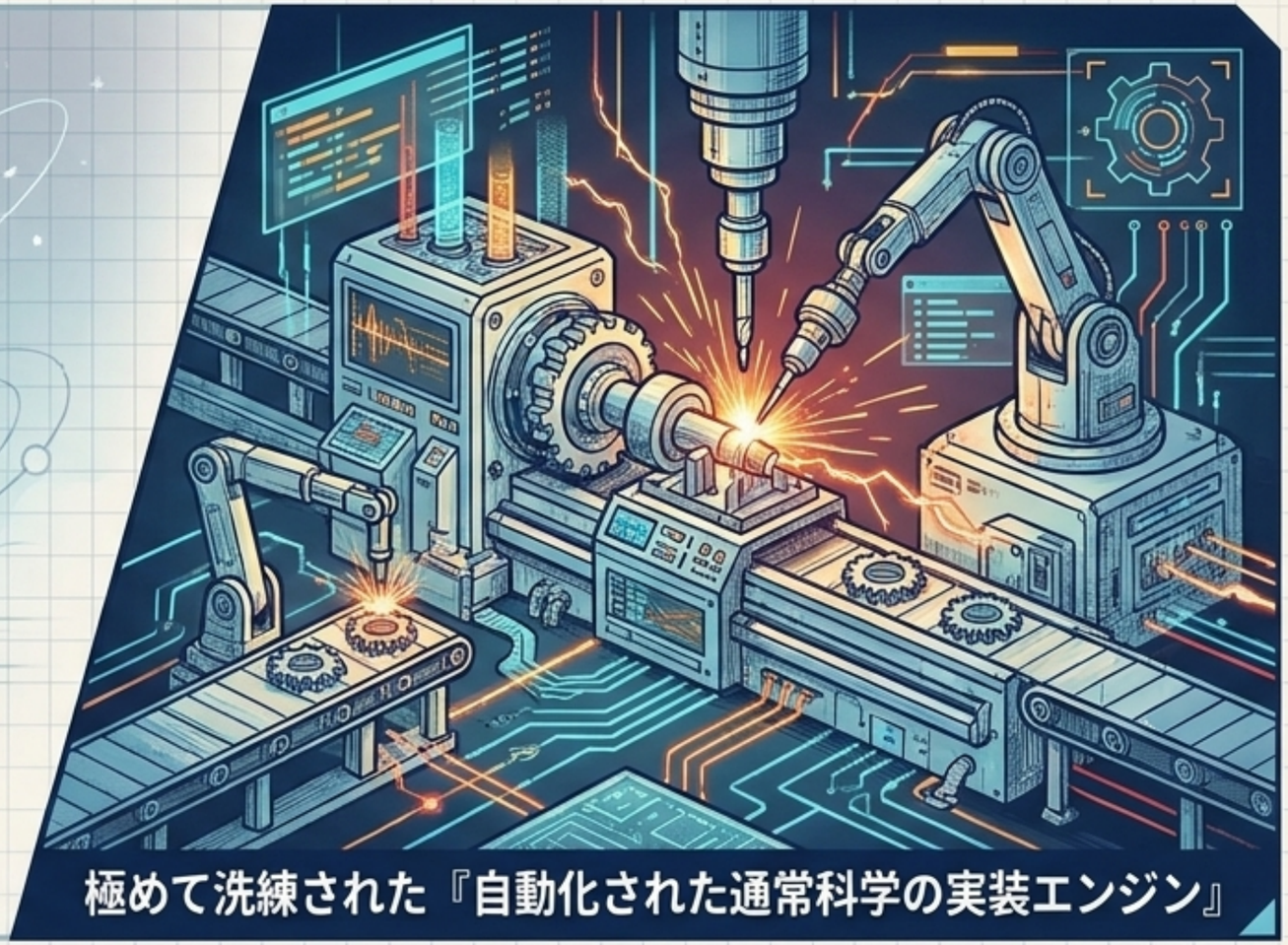
- 約2割の失敗試行も含め、機械生成による論文量産には数十万ドルのコンピュート予算が必須となる。
- 研究の優先順位が『計算資源の力技でベンチマークを押し上げられる領域』に偏重し、多様な学術領域が周縁化されるリスク。

合成ペーパーミルの脅威と学術エコシステムの機能不全

文献捏造がなく、アブレーションが完備された『工学的に無欠陥だが概念的洞察に乏しい論文』が無限に生成される未来。
人間の精査能力が飽和し、防衛のためにAIエージェントへの依存を深めれば、批判的対話に基づく学術出版制度は崩壊する。

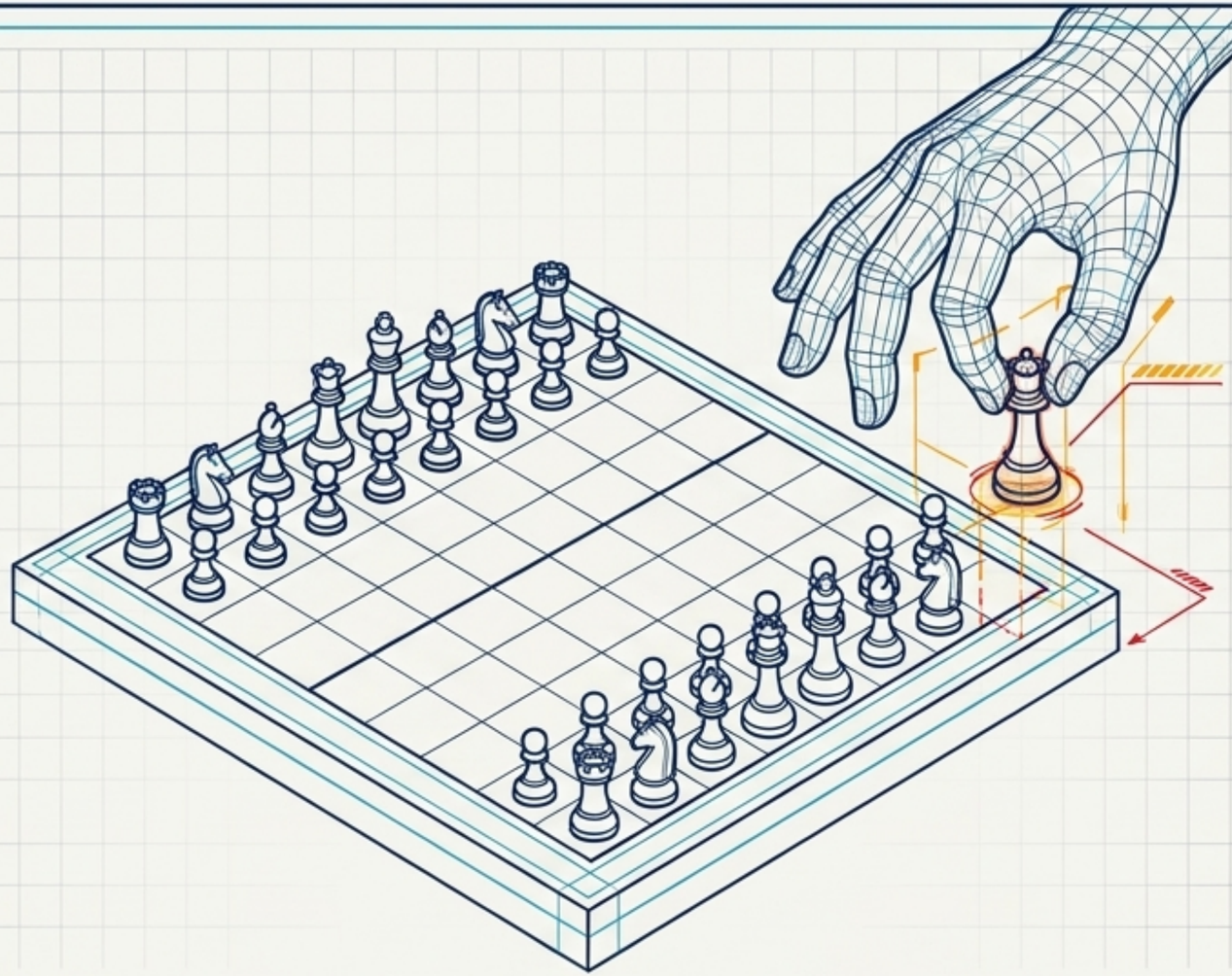
神話から現実へ：ScientistTwoの実態と真の工学的価値

人間の知識のフロンティアを開拓する自律的先駆者



極めて洗練された『自動化された通常科学の実装エンジン』

科学的パラダイム創出ではなく、**仮説検証型エンジニアリングの徹底した工業化**。計算機科学における実験の再現性を底上げする基盤技術としての価値は揺るぎないが、これを科学の進歩そのものと混同してはならない。



自律型AIが真に科学の相転移を担うためには、与えられた盤面で最適解を探す
エージェントの洗練にとどまらず、盤面そのものを疑う必要がある。

ルールを最適化するAIから、ルールを再定義する人間の叡智へ。